



OPEN

Statistical inference using GLEaM model with spatial heterogeneity and correlation between regions

Yixuan Tan^{1,6}, Yuan Zhang^{2,6}, Xiuyuan Cheng^{1✉} & Xiao-Hua Zhou^{3,4,5✉}

A better understanding of 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 previously developed Global Epidemic and Mobility (GLEaM) model, this paper proposes a new 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.

The outbreak of coronavirus disease 2019 (COVID-19) has impacted all aspects of the world significantly for a long. As of 26 Oct 2021, over 243 million confirmed cases of COVID-19 have been reported, including over 4 million deaths¹. Therefore, it is essential to study the spread of COVID-19 for better prediction and prevention of the disease. This paper proposes a new stochastic dynamical model that can describe different spread patterns of COVID-19 in multiple regions. We also develop an algorithm to estimate the corresponding transmission parameters and their posterior distributions. Our model is inspired by the Global Epidemic and Mobility (GLEaM) model proposed in Ref.². GLEaM is a stochastic dynamic model that depicts the spread of epidemics, integrating multiple data layers. The GLEaM model involves 3362 subpopulations in 220 countries obtained from Voronoi tessellation, centered around major airports. These subpopulations are connected by a multi-layered mobility network composed of processes from short-range commuting between nearby subpopulations to international flights. In each subpopulation, the transmission of epidemics is modeled by a variant of an Susceptible-Exposed-Infected-Removed (SEIR) compartmental model³. Please see “[Review of the SEIR model and GLEaM model](#)” for a more detailed review of the SEIR model and GLEaM model.

In the vast majority of GLEaM's applications^{4–11}, the parameters are estimated based on Ref.¹². Reference¹² performed the maximal likelihood analysis of the reproduction number R_0 in the seed region, Mexico. For each value of the reproduction number R_0 , the method generated the distribution of the arrival time of the influenza A(H1N1) in 12 countries produced by 2×10^3 GLEaM simulations. Then, the optimal reproduction number R_0 was chosen by maximizing the likelihood function of arrival time. Reference¹² and subsequent works following its settings^{4–11} 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 Ref.¹² was adopted in Ref.⁶ to estimate the posterior distribution of the reproduction number R_0 of COVID-19, which was assumed to be uniform for all subpopulations at all times. However, this setting is unsuitable 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 world¹³. To model the spread of COVID-19, both spatial and temporal heterogeneity of the transmission parameters are needed, rather than directly modeling the reproduction number R_0 solely as a periodic function of time as in Ref.¹². This is because the social behaviors, containment measures, medical conditions, and other elements that affect the spread of COVID-19 may vary among different countries and over time.

¹Department of Mathematics, Duke University, Durham, USA. ²School of Statistics, Renmin University of China, Beijing, China. ³Center for Statistical Sciences, Peking University, Beijing, China. ⁴Beijing International Center for Mathematical Research, Peking University, Beijing, China. ⁵Department of Biostatistics, School of Public Health, Peking University, Beijing, China. ⁶These authors contributed equally: Yixuan Tan and Yuan Zhang. ✉email: xiuyuan.cheng@duke.edu; azhou@math.pku.edu.cn

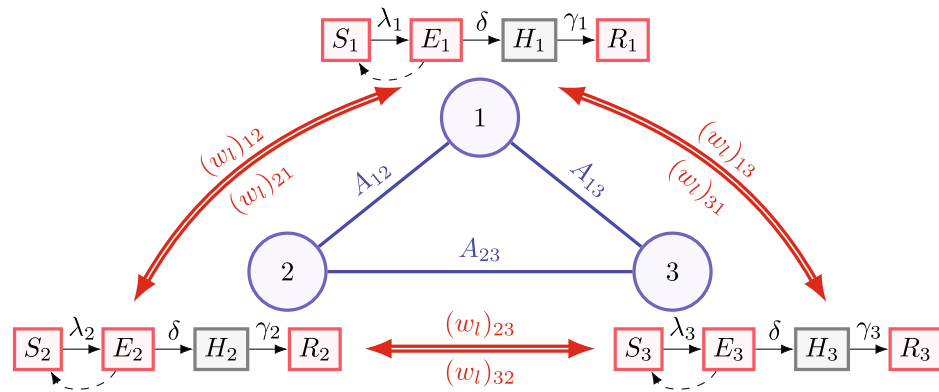


Figure 1. A diagram illustrating the model proposed in this paper. The example includes three regions, marked by circles in blue and indexed by 1, 2, and 3. For $k, j \in \{1, 2, 3\}$ and $k \neq j$, A_{kj} on the edge (k, j) represents the similarity between regions k and j . The square nodes associated with each region denote the compartments for each region, including susceptible, exposed, hospitalized, and removed compartments. The arrows connecting square nodes denote the transition between compartments in each region, and the details can be found in “Model description”. Note that λ_1, λ_2 , and λ_3 , the transmission parameters in the three regions, are allowed to be spatially heterogeneous. The double arrows in red denote the transportation between three regions on the l -th day. For $k, j \in \{1, 2, 3\}$ ($k \neq j$), $(w_l)_{kj}$ represents the total transportation volume from region k to region j on the l -th day.

Recently, Reference¹⁴ improved the inference method in Ref.¹² based on the GLEaM model, by involving spatial heterogeneity. Specifically, Ref.¹⁴ estimated 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. However, international travel was not considered in this study, and GLEaM was applied to each of the aforementioned countries (as an isolated systems) independently. The transmission rates in all subpopulations of these countries were again presumed to be homogeneous. Furthermore, Ref.¹⁴ 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 abovementioned issues, other potential concerns exist in applying GLEaM to model the spread of COVID-19. As mentioned in the last section of Ref.¹⁵, 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-layered mobility networks. Furthermore, the estimate of parameters using GLEaM is based on a large number of simulations to explore the space of parameters, which may potentially take much computational time² when the epidemic parameters to be estimated are spatially heterogeneous. 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 have their own heterogeneity but also be correlated to each other. To the best of our knowledge, 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 mainly focus on estimating only the transmission parameter in the seed region at the very beginning of the outbreak. However, for the current long-lasting spread of the COVID-19 pandemic all over the globe, the spatial and temporal heterogeneity of the transmission parameters is needed to be taken into full consideration.

In this paper, we propose a new stochastic model that incorporates transportation between regions and at the same time enables spatial and temporal heterogeneity of transmission parameters. We model n regions as a graph having n nodes, and the transportation pattern between the regions is encoded as n -by- n matrices. Our graphical model of epidemic dynamics is a general abstract one motivated by and simplified from the GLEaM framework. Figure 1 shows a diagram of the proposed model. In contrast to most applications of GLEaM, which mainly focus on the initial outbreak, our proposed model is able to model the long-lasting spread of epidemics. For the inference of model parameters, we introduce an optimization algorithm that utilizes the correlation between districts. Furthermore, the posterior distribution of parameters is estimated by an Markov Chain Monte Carlo (MCMC) sampling procedure, where we set the initial value of the Markov Chain as the optimal parameter obtained by the optimization algorithm. This approach can potentially accelerate the convergence of MCMC sampling.

In summary, the main contributions of our paper are:

- We propose a new stochastic model to describe the epidemic’s long-lasting spread, allowing spatial and temporal heterogeneity of transmission parameters and transportation between districts.

Notation	Description	Notes
n	Total number of regions	
T	Total number of days	
$S_k(t)$	Susceptible individuals in the k -th region at time t $1 \leq k \leq n, t \in [0, T]$	The same range of k and t also holds below
$E_k(t)$	Exposed individuals in the k -th region at time t	$\{E_k(0)\}_{k=1}^n$ need to be inferred
$H_k(t)$	Hospitalized individuals in the k -th region at time t	$\{H_k(0)\}_{k=1}^n$ need to be inferred
$R_k(t)$	Removed individuals in the k -th region at time t	$R_k(0) = 0, \forall k$
$N_k(t)$	Total individuals in the k -th region at time t	$\sum_{k=1}^n N_k(t)$ is assumed to be constant over time
W_l	Traveling volume matrix on the l -th day	$(W_l)_{ij} = (w_l)_{ij}, l \in \{1, \dots, T\}$
$\tilde{S}_k(t)$	Deterministic counterpart of $S_k(t)$,	Other compartments follow the same
	Determined by (2.4)	Convention of notations, $t \in [0, T]$
$(\tilde{C}_a)_k(t)$	Accumulated confirmed cases determined by (2.4)	$t \in [0, T]$
$(\tilde{R}_a)_k(t)$	Accumulated removed cases determined by (2.4)	$t \in [0, T]$
$(C_a)_k(i)$	Accumulated confirmed cases on the i -th day	$i \in \{1, \dots, T\}$
$(R_a)_k(i)$	Accumulated removed cases on the i -th day	$i \in \{1, \dots, T\}$
$(\Delta \tilde{C}_a)_k(i)$	Newly confirmed cases determined by (2.4) on the i -th day	$i \in \{1, \dots, T\}$
$(\Delta C_a)_k(i)$	Newly confirmed cases on the i -th day	$i \in \{1, \dots, T\}$
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 removed time in the k -th region	$\{\gamma_k\}_{k=1}^n$ need to be inferred
Θ	Set of parameters to be estimated	
W	The matrix characterizing the proximity between regions	
d	Total number of groups that n regions are divided into	
D_m	The m -th group of regions	$1 \leq m \leq d$
μ	Penalty factor of Graph Laplacian regularization	
β	Parameter reducing regularization between inter-group regions	$\beta \in (0, 1)$
A	Affinity matrix constructed by W, μ , and β as in (2.12)	
σ	Parameter in the prior (2.10)	

Table 1. List of notations and parameters.

- Based on the proposed model, we also design an algorithm that first makes inference for the parameters through a two-step procedure and then estimates the posterior distribution efficiently by MCMC sampling with the estimated parameters as the initial points. The parameter inference combines the information of correlation between districts, which is equivalent to imposing graph Laplacian regularization on the transmission parameters.
- We compare the performance of the proposed model with the baseline models on both simulated 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 outperforms the baselines in trajectory prediction.

A strength of the proposed model resides in introducing spatial and temporal heterogeneity of transmission parameters. We compare with more related works and comment on the differences and relations in “[More related works](#)”. Datasets used in this paper are publicly available at Refs.^{16–18}. Our work focuses on the methodology development and we aim at a new stochastic dynamic model that is generally applicable.

We list the default notations and parameters used throughout the paper in Table 1. The rest of the paper is structured in the following way: In “[Methods](#)”, we introduce the stochastic dynamic model and the corresponding inference algorithm. In “[Experimental results for simulated data](#)”, we compare the performance of trajectory prediction and parameter estimation of the models with or without mobility, heterogeneity, and using correlation information in the inference part for the simulated data. Section “[Experimental results on COVID-19 data](#)”

describes the real-world data used in this paper, and presents the results and findings of applying the proposed model to the COVID-19 data in China and Europe. We discuss the limitations and possible extensions in “Discussion”.

Review of the SEIR model and GLEaM model. In this section, we provide a more detailed introduction to the SEIR model and the GLEaM model so as to provide a background of our study and augment the following context.

To depict the evolution of the epidemics, Ref.¹⁹ proposed the celebrated Susceptible-Infected-Removed (SIR) model and characterized the development of the pandemic with a deterministic ordinary differential equation (ODE). There are many extensions of the SIR model, including the Susceptible-Exposed-Infected-Removed (SEIR) model for diseases with a latent period, the Susceptible-Infected-Susceptible (SIS) model for diseases that do not gain immunity after recovery, etc.

These deterministic transmission models are constructed under certain assumptions, including that the population is large, closed, and homogeneous. Due to the random nature of the transmission process, many stochastic dynamic models are developed^{20–22}. Under certain rather generalized conditions, the deterministic models can be seen as the mean-field equations of the corresponding stochastic processes. However, this approximation may not hold when the size of the outbreak has not grown up to the same order of the total population, which is the case in many applications²³. More details can be found in Ref.²⁴ and the references therein.

The Global Epidemic and Mobility (GLEaM) model proposed in Ref.² used a meta-population scheme which balanced between the agent-based stochastic models and the deterministic compartmental models. Specifically, Ref.² adapted a high-resolution population database that divided the surface of the earth with cells of 15 min × 15 min of arc, and then used Voronoi tessellation to assign each cell to one of the major airports around the world. The obtained subdivisions were then called subpopulations.

The stochastic dynamic in the subpopulations was then coupled with two layers of mobility flows apart from the infection dynamic within each subpopulation. The first layer was the worldwide airport network between the airports in the subpopulations, which could be seen as a weighted graph whose edges represented the number of passengers between each pair of airports. This layer was integrated into the model through stochastic transportation between subpopulations. The second layer was the commuting network that connected subpopulations graphically close. This layer was integrated through being used to compute the effective population and infection in each subpopulation. More details can be found in Ref.².

More related works. Several recent works also involved different levels of heterogeneity in their models in various ways. Reference^{25,26} utilized randomness in reproduction numbers to reflect the heterogeneity of the population, using plate model with Bayesian method and heterogeneous well-mixed theory²⁷ with age-of-infection method¹⁹, respectively. References^{28,29} used functional data analysis tools. Specifically, Ref.²⁸ captured two different epidemic patterns in different regions of Italy using the probKMA algorithm Refs.³⁰, and²⁹ revealed different patterns of the epidemic across countries with functional principle component analysis. In addition, Refs.^{31–35} adapted SEIR / Susceptible-Exposed-Infected (SEI) / Susceptible-Infected (SI) compartmental models similar to this paper. Among these works,^{31–33} considered heterogeneity in the aspects of age groups, social links, and vaccination status separately. References^{34,35} bore more similarity with our paper since they also allowed transmission parameters to be spatially heterogeneous and involved transportation between different regions. However, Ref.³⁴ 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 Ref.³⁴ were determined by an SDE whose parameters were to be fitted. Therefore, Ref.³⁴ focused on a different scope from our study. The settings of compartments in Ref.³⁵ were more realistic than the one considered in our paper by considering reporting rates. Nevertheless, compared with the model and inference algorithm described in “Methods”, transmission rates in Ref.³⁵ did not have temporal heterogeneity or correlation with each other. Both Refs.^{34,35} used the 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 might be computationally less efficient than directly applying MCMC according to the posterior distribution with the initial point maximizing the posterior distribution, as implemented in this paper.

Methods

Ethics statement. The medical record data in China and Europe used in this paper are publicly available and can be found on the official websites of the National Health Commission of the People’s Republic of China¹⁶, the Chinese Center for Disease Control and Prevention¹⁷, and European Centre for Disease Prevention and Control¹⁸. The collection of data is performed in compliance with local government regulations. More details about data sources can be found in “Data sources”.

Model description. *Compartmental model over multiple regions.* In GLEaM² and other epidemic models involving transportation^{34,35}, the whole area is usually divided into subdivisions. For example, the GLEaM model divides the total area of 220 countries into over 3300 subpopulations centered around major airports and³⁴ divided Milwaukee County and Dane County in the state of Wisconsin into several regions. In this paper, we consider abstract subdivisions in the whole area, which will be referred to as “regions” hereinafter until further specifications in the later experiment sections. We denote n as the number of regions.

In our model, we use continuous time $t \in [0, T]$, where it is assumed that the evolution of the epidemic lasts within a period of T time units. The unit of time is fixed as one day throughout this paper. Note that when we introduce the transportation model in below the traveling matrix is assumed to be constant within each day,

and the observed data is also collected on a daily basis. Thus we will use notation of discrete time (days) from $1, \dots, T$ hereinafter, however, the evolution dynamic itself is modeled over continuous time.

For each region, we consider the following epidemiology compartments adapted from the SEIR model:

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

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

At time t , we use $N_k(t) = S_k(t) + E_k(t) + H_k(t) + R_k(t)$ to denote the total population in the k -th region. The population $N_k(t)$ is allowed to be time-varying due to the inter-region mobility, especially for the days before the implementation of travel restrictions. However, since the traveling volume is not comparable to the total population in a region, the fluctuation of the total population in a region is not obvious. In this paper, we assume that $N = \sum_{k=1}^n N_k(t)$ keeps constant over time, which means that we consider a closed system, where exported/imported cases are not considered. However, it is worth noting that we do allow the transportation of active virus carriers between regions within our system. We remark in advance that this assumption is reasonable for the real-world data sets considered in this paper. From January to February 2020, strict international travel restrictions were imposed in China. While for data in Europe, from May to August 2020, the local spread of the epidemic has reached a relatively high level, and the imported cases were not comparable to the indigenous cases. 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.

Transportation between regions and the stochastic model. Transportation plays an essential role in the spread of COVID-19. Actually, Refs.^{36,37} indicated that the travel restrictions were remarkably important in mitigating the transmission of COVID-19, especially in the early stage of the pandemic. Recently, as detailed in Ref.³⁸, the Omicron variant had spread to 110 countries and had become dominant in many of them by 22 December 2021, only one month after its first report from South Africa on 24 November 2021. This motivates us also to take transportation into consideration in this paper. In our model, we introduce the transportation between regions via a traveling matrix, which is similar to the notation in the GLEaM model². Specifically, we denote $(w_l)_{kj}$ as the traveling volume from region k to j on the l -th day ($l = 1, \dots, T$). Then the traveling matrix W_l on the l -th day can be written as

$$W_l = \begin{pmatrix} (w_l)_{11} & (w_l)_{12} & \cdots & (w_l)_{1n} \\ (w_l)_{21} & (w_l)_{22} & \cdots & (w_l)_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ (w_l)_{n1} & (w_l)_{n2} & \cdots & (w_l)_{nn} \end{pmatrix}. \tag{2.1}$$

Given the transportation matrix, we describe in below the stochastic model of the dynamic of the compartments over n regions, denoted as $\{(S_k(t), E_k(t), H_k(t), R_k(t)), k = 1, \dots, n\}$, where $t \in [0, T]$ is the continuous time, and the variables $S_k(t)$, $E_k(t)$, $H_k(t)$, and $R_k(t)$ take integer values from 0 to N . The proposed stochastic model is illustrated in Fig. 1.

- Transmission in the k -th region: A case from $E_k(t)$ chooses an individual from $N_k(t)$ randomly at Poisson rate λ_k ($\{\lambda_k\}_{k=1}^n$ are allowed to be spatially heterogeneous), 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 rate γ_k owns spatial heterogeneity due to the uneven distribution of medical resources.
- Transportation between regions: At the end of the l -th day, all the individuals in region k except the ones in $H_k(l)$ have the same probability of traveling from region k to region j , and the total traveling volume from region k to region j is $(w_l)_{kj}$. We assume that there are no transmissions happening during the transportation between regions. If we denote $\{\xi_{kj}^{[S,l]}, \xi_{kj}^{[E,l]}, \xi_{kj}^{[R,l]}\}$ as the number of people transported from $\{S_k(l), E_k(l), R_k(l)\}$ to $\{S_j(l), E_j(l), R_j(l)\}$ at the end of the l -th day, then $\{\xi_{kj}^{[S,l]}, \xi_{kj}^{[E,l]}, \xi_{kj}^{[R,l]}\}$ follows a multinomial distribution. Specifically,

$$P\left(\left\{\xi_{kj}^{[S,l]}, \xi_{kj}^{[E,l]}, \xi_{kj}^{[R,l]}\right\}\right) = \frac{((w_l)_{kj})!}{(\xi_{kj}^{[S,l]})!(\xi_{kj}^{[E,l]})!(\xi_{kj}^{[R,l]})!} \left(\frac{S_k(l)}{(N_a)_k(l)}\right)^{\xi_{kj}^{[S,l]}} \left(\frac{E_k(l)}{(N_a)_k(l)}\right)^{\xi_{kj}^{[E,l]}} \left(\frac{R_k(l)}{(N_a)_k(l)}\right)^{\xi_{kj}^{[R,l]}}, \tag{2.2}$$

with $\xi_{kj}^{[S,l]} + \xi_{kj}^{[E,l]} + \xi_{kj}^{[R,l]} = (w_l)_{kj}$.

As a consequence, $N_k(t)$ is a piece-wise constant function of t which only changes at the end of each day. Specifically, for any time $t \geq 0$,

$$N_k(t) = N_k(0) + \sum_{l=1}^{\lfloor t \rfloor} \sum_{i=1}^n ((w_l)_{ik} - (w_l)_{ki}). \tag{2.3}$$

A more comprehensive stochastic dynamic model has been previously developed in Ref.²⁴. However, the work did not consider transportation between regions, which is a focus of this study.

Differential equation with spatial heterogeneity. Following Refs.^{39,40}, we derive the corresponding mean-field differential Eq. (2.4) of the stochastic dynamic introduced in “[Transportation between regions and the stochastic model](#)”, which is continuous in time, and the compartments $(\tilde{S}_k(t), \tilde{E}_k(t), \tilde{H}_k(t), \tilde{R}_k(t))$ take real values.

$$\begin{cases} \frac{d\tilde{S}_k(t)}{dt} = -\lambda_k \frac{\tilde{S}_k(t)\tilde{E}_k(t)}{\tilde{N}_k(t)} + \sum_{i \neq k} \left[(w_{\lceil t \rceil})_{ik} \frac{\tilde{S}_i(t)}{(\tilde{N}_a)_i(t)} - (w_{\lceil t \rceil})_{ki} \frac{\tilde{S}_k(t)}{(\tilde{N}_a)_k(t)} \right] \\ \frac{d\tilde{E}_k(t)}{dt} = \lambda_k \frac{\tilde{S}_k(t)\tilde{E}_k(t)}{\tilde{N}_k(t)} - \delta \tilde{E}_k(t) + \sum_{i \neq k} \left[(w_{\lceil t \rceil})_{ik} \frac{\tilde{E}_i(t)}{(\tilde{N}_a)_i(t)} - (w_{\lceil t \rceil})_{ki} \frac{\tilde{E}_k(t)}{(\tilde{N}_a)_k(t)} \right] \\ \frac{d\tilde{H}_k(t)}{dt} = \delta \tilde{E}_k(t) - \gamma_k \tilde{H}_k(t) \\ \frac{d\tilde{R}_k(t)}{dt} = \gamma_k \tilde{H}_k(t) + \sum_{i \neq k} \left[(w_{\lceil t \rceil})_{ik} \frac{\tilde{R}_i(t)}{(\tilde{N}_a)_i(t)} - (w_{\lceil t \rceil})_{ki} \frac{\tilde{R}_k(t)}{(\tilde{N}_a)_k(t)} \right] \\ \frac{d\tilde{N}_k(t)}{dt} = \sum_i ((w_{\lceil t \rceil})_{ik} - (w_{\lceil t \rceil})_{ki}) \\ \frac{d(\tilde{C}_a)_k(t)}{dt} = \delta \tilde{E}_k(t) \\ \frac{d(\tilde{R}_a)_k(t)}{dt} = \gamma_k \tilde{H}_k(t) \end{cases} \tag{2.4}$$

The first four equations of (2.4) describe the evolution of the $S_k(t), E_k(t), H_k(t), R_k(t)$ in the deterministic version of our model, which is governed by the transition dynamic explained in “[Transportation between regions and the stochastic model](#)”. The fifth equation characterizes the deterministic total population, which is a piece-wise linear function of time t (since the traveling volume is a piece-wise constant function of t) and coincides with $N_k(t)$ expressed as in (2.3) when t takes integer values. The last two equations depict the evolution of accumulated confirmed and removed cases, denoted by $(\tilde{C}_a)_k(t)$ and $(\tilde{R}_a)_k(t)$ respectively, in the deterministic model. It is worth noting that in the calculation of $(\tilde{C}_a)_k(t)$ and $(\tilde{R}_a)_k(t)$, each case is only accounted for once. In (2.4), $\tilde{S}_k(t)$ is the deterministic counterpart of $S_k(t)$ and the same for $\tilde{E}_k(t), \tilde{H}_k(t), \tilde{R}_k(t), (\tilde{N}_a)_k(t)$.

Furthermore, we assume that the accumulated confirmed and removed cases are available from data on a daily basis, which are denoted as $\{(C_a)_k(i)\}_{i=1}^T$ and $\{(R_a)_k(i)\}_{i=1}^T$, respectively. We also assume that $R_k(0)$ is 0, while $E_k(0)$ and $H_k(0)$ are left to be inferred for each $k = 1, \dots, n$. For inference of parameters, we further denote $(\Delta C_a)_k(i) = (C_a)_k(i) - (C_a)_k(i-1)$ as the deterministic newly confirmed cases on the i -th day determined by (2.4), and $(\Delta C_a)_k(i)$ as the newly confirmed cases computed from data, namely $(\Delta C_a)_k(i) = (C_a)_k(i) - (C_a)_k(i-1), k = 1, \dots, n, i = 2, \dots, T$. The same convention holds for the definitions of $(\Delta R_a)_k(i)$ and $(\Delta R_a)_k(i)$. Note that the data $\{(C_a)_k(i), (R_a)_k(i), (\Delta C_a)_k(i), (\Delta R_a)_k(i)\}_{i=1}^T$ are random in nature.

Note that the model and the inference algorithm described below can be applied to estimate parameters as long as $\{(C_a)_k(i), k = 1, \dots, n\}_{i=1}^T, \{(R_a)_k(i), k = 1, \dots, n\}_{i=1}^T$, and $\{W_l\}_{l=1}^T$ are available. The availability of $\{(C_a)_k(i), k = 1, \dots, n\}_{i=1}^T$ and $\{(R_a)_k(i), k = 1, \dots, n\}_{i=1}^T$ is required in many works that use the SEIR model to estimate transmission rates of the epidemic⁴¹⁻⁴³, and the transportation network is also used in GLEaM² and its applications. However, in contrast to the works based on GLEaM^{4-6,12,15}, we allow parameters to possess both spatial and temporal heterogeneity, and further utilize the correlation between regions in the inference of parameters. We remark that the spatial and temporal heterogeneity is reflected in the fact that the transmission parameters $\{\lambda_k\}$ are allowed to vary in both space and time in our model. The temporal heterogeneity is introduced in more detail for the real-world data in “[Model extension by allowing time-varying parameters](#)”.

Estimation of model parameters. Based on the model described in “[Model description](#)”, the parameters that need to be specified are δ and $\Theta := \{E_k(0), H_k(0), \lambda_k, \gamma_k\}_{k=1}^n$. Using a simplification in Remark 1, we prefix the parameter δ , and estimate the rest in a two-step procedure to be described in this section. As a brief summary,

- Step 1. We first make inference for $\{\gamma_k\}_{k=1}^n$ by maximizing the likelihood of the observed newly removed cases. Details in “[Step 1: Estimate \$\{\gamma_k\}_{k=1}^n\$](#) ”.
- Step 2. After the estimation of $\{\gamma_k\}_{k=1}^n$, $\bar{\Theta} := \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$ are then estimated by maximizing the posterior probability, where we introduce a prior distribution combining the information of correlation between regions. Details in “[Step 2: Estimate \$\bar{\Theta} = \{E_k\(0\), H_k\(0\), \lambda_k\}_{k=1}^n\$](#) ”.

Finally, in the end of Step 2, we introduce an MCMC sampling approach to estimate the marginal posterior distributions of $\bar{\Theta}$. This provides information about the uncertainty of the estimated parameters, like λ_k , which are of scientific interest. We summarize the two-step procedure in this section together in Algorithm 1.

Remark 1 Among the unknown parameters, we prefix the parameter δ , the inverse of the average time for a person from being exposed to hospitalized, to be 0.14 universally in the algorithm. According to⁴⁴, 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^{45,46}. Thus, the mean duration for an individual from being exposed to being hospitalized is 7.2 days, whose inverse value is approximately 0.14.

Step 1: Estimate $\{\gamma_k\}_{k=1}^n$. We first estimate $\{\gamma_k\}_{k=1}^n$ by maximizing the likelihood

$$P\left(\{(R_a)_k(i), (C_a)_k(i), k = 1, \dots, n\}_{i=1}^T \mid \{\gamma_k\}_{k=1}^n\right)$$

over γ_k for each k . We assume that the newly removed 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 removed cases, and thus is observable) the day before. Then, the likelihood of $\{\gamma_k\}_{k=1}^n$ can be written as

$$\begin{aligned} &P\left(\{(R_a)_k(i), (C_a)_k(i), k = 1, \dots, n\}_{i=1}^T \mid \{\gamma_k\}_{k=1}^n\right) \\ &= \prod_{i=1}^T P\left(\{(\Delta R_a)_k(i), k = 1, \dots, n\} \mid \{\gamma_k\}_{k=1}^n, \{(R_a)_k(i-1), (C_a)_k(i-1), k = 1, \dots, n\}\right) \\ &= \prod_{i=1}^T \prod_{k=1}^n \text{Pois}((\Delta R_a)_k(i) \mid \gamma_k((C_a)_k(i-1) - (R_a)_k(i-1))) \\ &= \prod_{k=1}^n \prod_{i=1}^T \frac{(\gamma_k((C_a)_k(i-1) - (R_a)_k(i-1)))^{(\Delta R_a)_k(i)}}{((\Delta R_a)_k(i))!} \exp(-\gamma_k((C_a)_k(i-1) - (R_a)_k(i-1))), \end{aligned}$$

where $\text{Pois}(k \mid \beta)$ ($k \in \mathbb{N}, \beta > 0$) denotes the probability that k occurrences are observed for a discrete random variable X having a Poisson distribution with mean β .

Then, we estimate $\gamma_k^* = \arg \max_{\gamma_k} \prod_{i=1}^T \text{Pois}((\Delta R_a)_k(i) \mid ((C_a)_k(i-1) - (R_a)_k(i-1))\gamma_k)$ for each k separately.

Step 2: Estimate $\bar{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$. Next, we estimate the remaining parameters $\bar{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$, by finding $\bar{\Theta}$ that achieves maximum a posteriori probability (MAP).

Posterior distribution of $\bar{\Theta}$ and MAE estimate. We denote the posterior distribution of $\bar{\Theta}$ given data $\{(\Delta C_a)_k(i), k = 1, \dots, n\}_{i=1}^T$ as $\pi(\bar{\Theta})$. Then by Bayesian formula,

$$\pi(\bar{\Theta}) = P\left(\bar{\Theta} \mid \{(\Delta C_a)_k(i)\}_{k,i}\right) = \frac{1}{Z} P\left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta}\right) P(\bar{\Theta}), \tag{2.5}$$

where $P(\bar{\Theta})$ is the prior distribution of $\bar{\Theta}$ to be determined and $Z = P(\{(\Delta C_a)_k(i)\}_{k,i})$ is a constant irrelevant to $\bar{\Theta}$. We further denote

$$V(\bar{\Theta}) = -\log\left(P\left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta}\right)\right) - \log(P(\bar{\Theta})), \tag{2.6}$$

then

$$\pi(\bar{\Theta}) = \frac{1}{Z} \exp(-V(\bar{\Theta})). \tag{2.7}$$

To fit the realistic evolution of the epidemic more precisely, $\bar{\Theta}$ is estimated as

$$\bar{\Theta}^* = \arg \max \pi(\bar{\Theta}) = \arg \min V(\bar{\Theta}), \tag{2.8}$$

with reasonable prior distribution $P(\bar{\Theta})$. Then, MCMC sampling scheme starting from $\bar{\Theta}^*$ is applied to get the posterior distribution for $\bar{\Theta}$. This process might possess higher computational efficiency than choosing the initial point for MCMC randomly or empirically.

Next, we specify the formulas for the likelihood function $P\left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta}\right)$ and the prior distribution $P(\bar{\Theta})$.

Likelihood function of $\bar{\Theta}$. Notice that the ODE system (2.4) is the mean-field version of our stochastic model, and $\{(\Delta C_a)_k(i)\}_{k,i}$ are determined by the parameters $\bar{\Theta}$ (δ and $\{\gamma_k\}_{k=1}^n$ are treated as given), thus by the Markov property, $\{(\Delta C_a)_k(i)\}$ are all independent for $k = 1, \dots, n, i = 1, \dots, T$ conditioned on the parameters $\bar{\Theta}$. Furthermore we suppose that $(\Delta C_a)_k(i) \sim \text{Pois}((\Delta C_a)_k(i))$. Thus, the likelihood of $\bar{\Theta}$ can be written as

$$P(\{(\Delta C_a)_{k(i)}\}_{k,i}|\bar{\Theta}) = \prod_{i=1}^T P\left(\{(\Delta C_a)_{k(i)}\}_{k=1}^n \middle| \bar{\Theta}\right) = \prod_{i=1}^T \prod_{k=1}^n \frac{((\widetilde{\Delta C_a})_{k(i)})^{(\Delta C_a)_{k(i)}}}{((\Delta C_a)_{k(i)})!} e^{-((\widetilde{\Delta C_a})_{k(i)})}. \tag{2.9}$$

Choice of prior distribution of $\bar{\Theta}$. The remaining problem is to choose the prior distribution $p(\bar{\Theta})$. Presuming that the transmission rates in the regions owning more similarities are closer, $p(\bar{\Theta})$ 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 $\lambda = (\lambda_1, \dots, \lambda_n)^T \in \mathbb{R}^n$,

$$P(\bar{\Theta}) = \frac{1}{C_{\sigma,A}} \exp(-\lambda^T(D - A)\lambda - \sigma \|\lambda\|_2^2), \tag{2.10}$$

where $D = \text{diag}\{d_1, \dots, d_n\}$ is the degree matrix of A with $d_i = \sum_{j=1}^n A_{ij}$, $C_{\sigma,A} = \int_{\mathbb{R}^n} \exp(-\lambda^T(D - A)\lambda - \sigma \|\lambda\|_2^2) d\lambda$ is a constant depending on σ and A . Here, a small σ is chosen for $p(\bar{\Theta})$ to be a probability measure without imposing much restriction on λ . Then, by (2.6) and (2.10)

$$\begin{aligned} V(\bar{\Theta}) &= -\log\left(P\left(\{(\Delta C_a)_{k(i)}\}_{k,i} \middle| \bar{\Theta}\right)\right) + \lambda^T(D - A)\lambda + \sigma \|\lambda\|_2^2 + \log C_{\sigma,A} \\ &= -\log\left(P\left(\{(\Delta C_a)_{k(i)}\}_{k,i} \middle| \bar{\Theta}\right)\right) + \frac{1}{2} \sum_{i,j} a_{ij}(\lambda_i - \lambda_j)^2 + \sigma \sum_i \lambda_i^2 + \log C_{\sigma,A}. \end{aligned} \tag{2.11}$$

The parameter estimation procedure could be extended to the case when $\{\lambda_k\}$ are time-varying by modifying (2.13), which we will introduce in more detail in “**Model extension by allowing time-varying parameters**” for the real-world data in China and Europe.

Construction of affinity matrix A appearing in prior distribution (2.10). Now, we specify the construction of affinity matrix A in (2.10) that reflects the similarity between regions. A is constructed from affinity matrix W by further addressing the correlation between regions with more similarities. To treat different data sets and W with a unified approach, we assume that $\max_{i,j} W_{ij} = 1$ (W can be re-scaled entry-wise if necessary).

For a given $W = (W_{ij})_{i,j \in \{1, \dots, n\}}$ whose choice is detailed later, the next step of attaining A is to divide the n regions into d groups ($D_1, \dots, D_d, \cup_{m=1}^d D_m = \{1, \dots, n\}$, and $\forall i \neq j, D_i \cap D_j = \emptyset$) where the regions in the same groups have more similarities. Then, for given $\beta \in (0, 1)$ and a given penalty factor $\mu > 0$, a_{ij} is constructed as follows:

$$a_{ij} = \mu \begin{cases} W_{ij}, & \text{if regions } i \text{ and } j \text{ are in the same group,} \\ \beta W_{ij}, & \text{otherwise.} \end{cases} \tag{2.12}$$

We remark that β in (2.12) is taken to be 0.1 for all the experiments in this paper. By constructing A as in (2.12), correlations for regions in the same groups are further addressed, whose transmission parameters are imposed with stronger restrictions.

Now we specify the choice of W for data sets that will be analyzed later in this paper. For simulated data and real-world data in China, in which cases the transportation data are available, we construct W from the traveling volume matrices $\{W_i\}_{i=1}^T$. Specifically, $W := \bar{W} / \max_{i,j} \bar{W}_{ij}$, where $\bar{W} := \frac{1}{2}((\frac{1}{T} \sum_{l=1}^T W_l) + (\frac{1}{T} \sum_{l=1}^T W_l)^T)$. 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. 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.

Specified formula for MAP estimate of $\bar{\Theta}$. From the MAP estimate (2.8), definition of V (2.11) given prior distribution (2.10), and the definition of a_{ij} in (2.12), the inference of $\bar{\Theta}$ can be equivalently written as follows

$$\begin{aligned} \bar{\Theta}^* &= \arg \min V(\bar{\Theta}) \\ &= \arg \min \left(-\log\left(P\left(\{(\Delta C_a)_{k(i)}\}_{k,i} \middle| \bar{\Theta}\right)\right) + \frac{1}{2} \sum_{i,j} a_{ij}(\lambda_i - \lambda_j)^2 + \sigma \sum_i \lambda_i^2 \right) \\ &= \arg \min \left(-\log\left(P\left(\{(\Delta C_a)_{k(i)}\}_{k,i} \middle| \bar{\Theta}\right)\right) + \frac{\mu}{2} \left(\sum_m \sum_{i,j \in D_m} W_{ij}(\lambda_i - \lambda_j)^2 \right. \right. \\ &\quad \left. \left. + \beta \sum_{m_1 < m_2} \sum_{i \in D_{m_1}, j \in D_{m_2}} W_{ij}(\lambda_i - \lambda_j)^2 \right) + \sigma \sum_i \lambda_i^2 \right), \end{aligned} \tag{2.13}$$

where $\left(P\left(\{(\Delta C_a)_{k(i)}\}_{k,i} \middle| \bar{\Theta}\right)\right)$ is given in (2.9).

It can be seen that by choosing the prior distribution as in (2.10), a l_2 regularization term is imposed for better generalization.

Estimation of the marginal posterior distribution of $\bar{\Theta}$. Finally, after choosing $P(\bar{\Theta})$ determined by A and σ , the optimization process $\bar{\Theta}^* = \arg \min V(\bar{\Theta})$ as in (2.13) is accomplished by a BFGS algorithm^{47–50}. To obtain the posterior distribution of $\bar{\Theta}$, we use classical MCMC sampling scheme starting from $\bar{\Theta}^*$ solved by the optimization.

Algorithm 1 GLEaM model with Graph Laplacian Prior Distribution

Input: Data of total confirmed cases and total removed cases $\{(C_a)_k(i)\}_{k,i}, \{(\widehat{R}_a)_k(i)\}_{k,i}$, traveling volume matrices $\{W_t\}_{t=1}^T$, graph Laplacian penalty factor μ , parameter β , prior parameter σ and affinity matrix W , division of n provinces D_1, \dots, D_d , number of MCMC iterations N_{ite}

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

- 1: Compute newly confirmed cases and newly removed cases, $(\Delta C_a)_k(i) \leftarrow (C_a)_k(i) - (C_a)_k(i-1)$, $(\Delta R_a)_k(i) \leftarrow (R_a)_k(i) - (R_a)_k(i-1)$
 - 2: Compute $\gamma_k^* \leftarrow \arg \max_{\gamma_k} \prod_{i=1}^T \text{Pois}((\Delta R_a)_k(i) | ((C_a)_k(i-1) - (R_a)_k(i-1))\gamma_k)$. (Step 1, c.f. Section 2.3.1)
 - 3: Construct affinity matrix A with W , μ , β , and divisions D_1, \dots, D_d as in (2.12). Compute the degree matrix D , $D_{ii} \leftarrow \sum_j a_{ij}$
 - 4: Compute the inferred parameters $\bar{\Theta}^* \leftarrow \arg \min_{\bar{\Theta}} V(\bar{\Theta}) = \arg \min_{\bar{\Theta}} \left(-\log \left(P \left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta} \right) \right) + \lambda^T (D - A)\lambda + \sigma \|\lambda\|_2^2 \right)$. (Step 2, c.f. Section 2.3.2).
 $\triangleright V(\bar{\Theta})$ should be changed accordingly if λ is allowed to be time-varying.
 - 5: Let $\Theta^* \leftarrow \{ \{\gamma_k^*\}_{k=1}^n, \bar{\Theta}^* \}$, and compute the predicted deterministic trajectory $\hat{y}_{1:T}$ by applying explicit Euler's method to (2.4) with Θ^*
 - 6: Run MCMC starting from $\bar{\Theta}^*$ for N_{ite} iterations
 - 7: Compute the estimated marginal posterior distribution of $\theta \in \bar{\Theta}$ by $\hat{p}(\theta) \propto \{(\Delta C_a)_k(i)\}_{k=1, \dots, n, i=1, \dots, T} \leftarrow \sum_{l=1}^{n_{bin}} \frac{c_l}{N} \mathbf{1}_{\theta \in H_l}$, 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
-

Prediction of the epidemic trajectories with the estimated parameters. Once the parameters $\Theta^* = \{ \{\gamma_k^*\}_{k=1}^n, \bar{\Theta}^* \}$ are estimated from the optimizations $\gamma_k^* = \arg \max_{\gamma_k} \prod_{i=1}^T \text{Pois} \left((\Delta R_a)_k(i) | ((C_a)_k(i-1) - (R_a)_k(i-1))\gamma_k \right)$ and $\bar{\Theta}^* = \arg \min V(\bar{\Theta})$ as described in “[Estimation of model parameters](#)”, 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 $\bar{\Theta}$ instead of using Θ^* alone, which also takes the randomness from $\bar{\Theta}$ into account. Particularly, this could be achieved by sampling $\bar{\Theta}$ from MCMC and then simulating trajectories with the sampled $\{ \{\gamma_k^*\}_{k=1}^n, \bar{\Theta} \}$. Additionally, deterministic trajectories determined by (2.4) could also be computed by explicit Euler's method.

Experimental results for simulated data

Two specific cases are considered for simulated data. We first remark that the regions in “[Methods](#)” are called as provinces in this section. Section “[Four provinces 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. Section “[Thirty provinces case](#)” considers thirty provinces randomly separated into three groups, with the other settings similar to the previous case. Section “[More details of experimental settings and sensitivity analysis](#)” includes more details of the experimental settings and sensitivity analysis.

More details of experimental settings and sensitivity analysis. *Experimental settings.* The results in “[Experimental results for simulated data](#)” are for 100 replicas. In each replica, three random trajectories are sampled independently according to the stochastic model with prefixed parameters, part of which are treated as the ground truth training, validation, and testing trajectory, respectively (see more details in Sect. A.1.1 of Supplementary Information).

For each model, we first fit the parameters using the training trajectory and then predict the testing trajectory using the estimated parameters. Note that we detail the choice of hyper-parameters for the model proposed in Sect. A.2 of Supplementary Information. In particular, the penalty factor $\mu > 0$ is chosen by cross-validation and chosen as the value minimizing the validation error, since for the simulated data, the validation error are identically distributed as the testing error. 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

Model	Migration	Heterogeneity	Prior of Θ
1	×	×	Uniform prior
2	✓	×	Uniform prior
3	×	✓	Uniform prior
4	✓	✓	Uniform prior
5	✓	✓	Graph Laplacian prior

Table 2. Models to be compared when the transportation data are available. Model 5 is the proposed model in this paper, and Models 1–4 are baseline models with different settings.

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, validation, and testing errors can be found in Sect. B of Supplementary Information.

Sensitivity analysis of σ . Note that parameter inference with the proposed model involves the parameter σ , as shown in (2.13). Therefore, the sensitivity analysis for the parameter σ in (2.13) is performed for the four provinces case. Specifically, the results for σ varying from 10^{-6} to 10^0 are presented and compared. Details can be found in “Results of parameter estimation” and “Further model evaluation”. Similar results are obtained for other data sets, and details are omitted.

Mismatched partition of regions. We also note that the graph Laplacian penalty of the proposed model depends on the partitioning the regions into several groups, as described in “Estimation of model parameters”. Since the graph knowledge is usually not fully known, it is a question whether our methods can still perform well without accurate prior knowledge. For the thirty provinces case, we report the results of the proposed model with a mismatch between the partition of the regions and the ground truth division, the details of which can be found in “Thirty provinces case”.

Four provinces case. *Data description.* In this simulated study, we let $n = 4$, $T = 20$, and set the threshold T_{th} separating training and testing data to be 10 (more detailed can be seen in Supplementary Information S1). 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 $l \in \{1, \dots, T\}$, $i, j \in \{1, 2, 3, 4\}$ and $i \neq j$, $(W_l)_{ij} = 5 \times 10^3$.
- $\lambda_1 = 0.5$, $\lambda_2 = 0.47$, $\lambda_3 = 0.4$, $\lambda_4 = 0.37$, $\delta = \gamma_1 = \dots = \gamma_4 = 0.14$.

The four provinces are divided into two groups, with the first group consisting of Provinces 1 and 2 and the second group consisting of Provinces 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.

Models to compare. **The proposed model and other four baseline models.** 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. The model with uniform prior distribution, without heterogeneity or migration.
2. The model with uniform prior distribution, without heterogeneity but with migration.
3. The model with uniform prior distribution, with heterogeneity but without migration.
4. The model with uniform prior distribution, with both heterogeneity and migration.
5. The model with prior distribution based on graph Laplacian, with both heterogeneity and migration.

For better illustration and comparison between the models in the experiment results, the Models 1–5 are summarized in Table 2 below.

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.

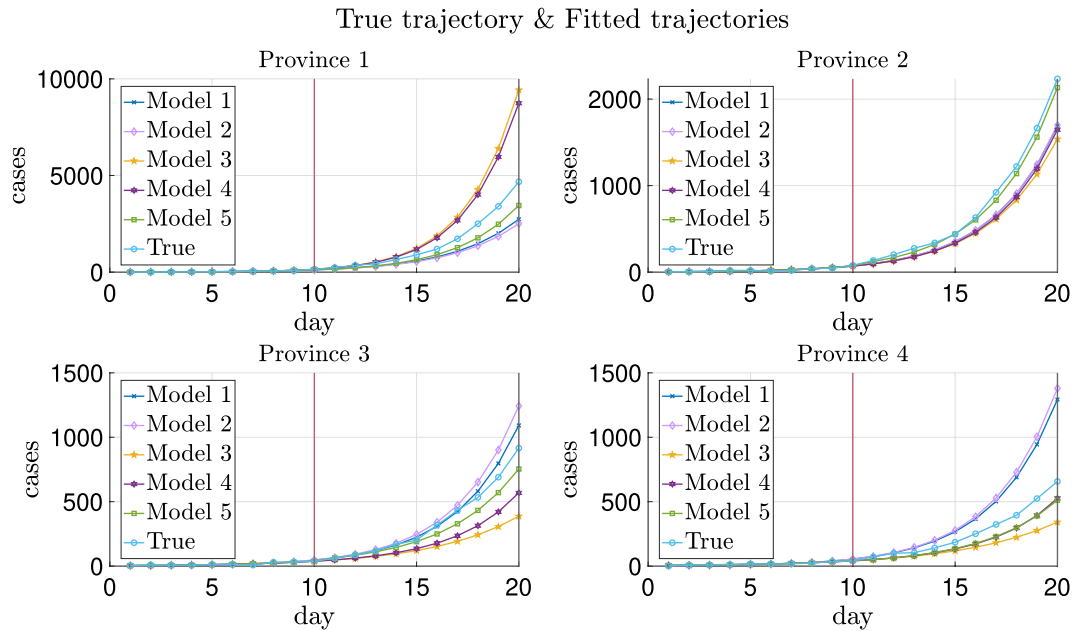


Figure 2. True and fitted trajectories for the simulated data with four provinces. The vertical lines show the threshold of training-testing split. In Model 5, $\mu = 10^{2.7}$ is chosen, at which the averaged validation errors over 100 replicas are minimized as shown in Supplementary Figs. S2 and S3.

Parameter inference of the five models and sensitivity of σ . For Model 5, the proposed model, the parameters $\Theta = \{\{\gamma_k\}_{k=1}^n, \bar{\Theta}\}$ are estimated following the two-step procedure described in “[Estimation of model parameters](#)”, where $\bar{\Theta} = \{E_k(0), H_k(0), \lambda_k\}_{k=1}^n$. For the estimation of $\bar{\Theta}$, following the general formula (2.13) in “[Estimation of model parameters](#)”, the specific formula of $\bar{\Theta}^*$ for the four provinces case is as follows,

$$\bar{\Theta}^* = \arg \min(-\log p(y_{1:T_{th}}|\bar{\Theta})) + \mu \left((\lambda_1 - \lambda_2)^2 + (\lambda_3 - \lambda_4)^2 + \beta \sum_{(i,j) \neq (1,2) \text{ or } (3,4)} (\lambda_i - \lambda_j)^2 \right) + \sigma \|\lambda\|_2^2, \tag{3.1}$$

where β is taken to be 0.1. For Model 5, we conduct the sensitivity analysis for parameter σ in (3.1) and present results for $\sigma = 10^{-6}, 10^{-3}$ and 10^0 respectively in “[Results of parameter estimation](#)” and “[Further model evaluation](#)”.

We remark that in Models 1–4, the estimation of $\Theta = \{\{\gamma_k\}_{k=1}^n, \bar{\Theta}\}$ still follows a similar two-step procedure as in Model 5, and the first step of obtaining $\gamma_k^* = \arg \max_{\gamma_k} \prod_{i=1}^T \text{Pois}((\Delta R_a)_k(i) | ((C_a)_k(i-1) - (R_a)_k(i-1))\gamma_k)$ remains formally the same. The difference lies in the optimization object of $\bar{\Theta}$. 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 ODE system (2.4). For models without transportation between regions, terms involving W_t disappear in (2.4). 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 Ref.¹² is similar to Models 1 and 2, since they all assume a spatially homogeneous transmission parameter. However, Ref.¹² assumed that the epidemic was seeded from one seed region while Models 1 and 2 do not make such assumption. Moreover, Ref.¹² 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.

Results of trajectory prediction. First, we remark that in Model 5, μ is chosen to be the minimizer of the averaged validation errors over 100 replicas over a range of values of μ . The weighted (simply averaged) validation errors, $\text{MAE}_{(w)}^{[Val]}$ and $\text{MSE}_{(w)}^{[Val]}$ ($\text{MAE}_{(s)}^{[Val]}$ and $\text{MSE}_{(s)}^{[Val]}$), are defined as in Sect. B of Supplementary Information. We remark that the superscript $^{[Val]}$ refers to when the error is computed on validation data, and the subscripts $_{(w)}$ and $_{(s)}$ denote that the errors are the weighted and simple average of relative errors over time respectively.

The averaged weighted validation errors $\text{MAE}_{(w)}^{[Val]}$ and $\text{MSE}_{(w)}^{[Val]}$ over replicas are shown in Supplementary Fig. S2, and the simply averaged counterparts are shown in Supplementary Fig. S3. For parameter inference using Model 5, μ is chosen to be $10^{2.7}$, at which all the averaged validation errors ($\text{MAE}_{(w)}^{[Val]}$, $\text{MSE}_{(w)}^{[Val]}$, $\text{MAE}_{(s)}^{[Val]}$ and $\text{MSE}_{(s)}^{[Val]}$) over 100 replicas are minimized, as can be seen from Supplementary Figs. S2 and S3.

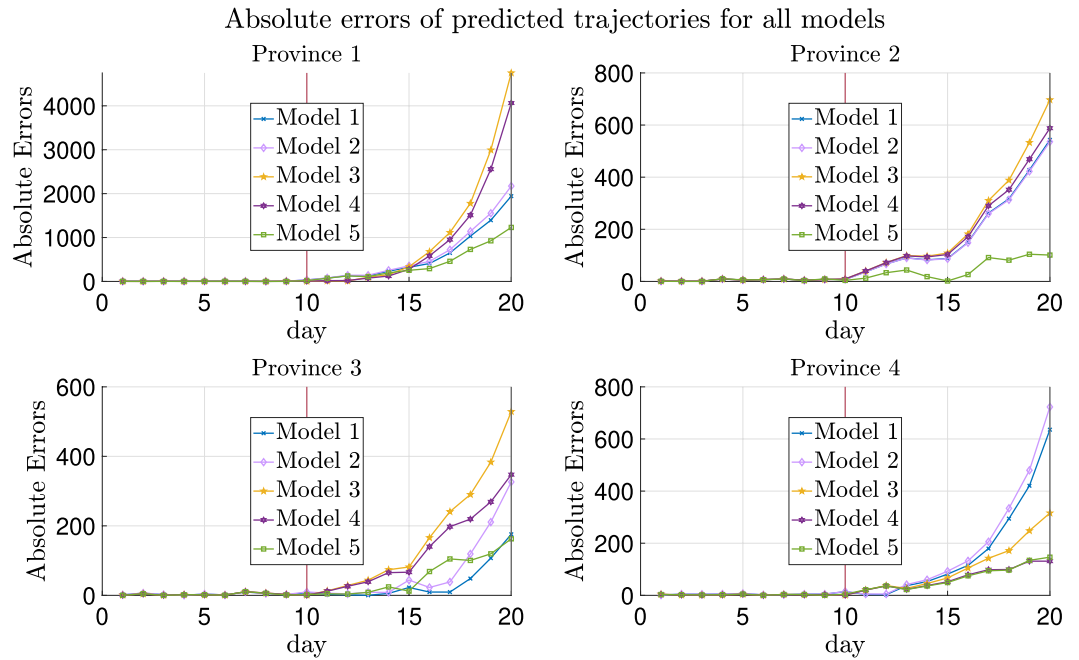


Figure 3. Absolute errors of fitted trajectories of Models 1–5 for the simulated data with four provinces. The vertical lines show the threshold of training-testing split.

Model	M	H	GL prior	λ_1	λ_2	λ_3	λ_4
1	×	×	×	0.4458 ± 0.0192			
2	✓	×	×	0.4459 ± 0.0192			
3	×	✓	×	0.4927 ± 0.0331	0.4620 ± 0.0370	0.3988 ± 0.0389	0.3705 ± 0.0349
4	✓	✓	×	0.4980 ± 0.0337	0.4643 ± 0.0376	0.3922 ± 0.0407	0.3605 ± 0.0366
5	✓	✓	✓ ($\sigma = 10^{-6}$)	0.4805 ± 0.0254	0.4654 ± 0.0260	0.3979 ± 0.0256	0.3879 ± 0.0231
	✓	✓	✓ ($\sigma = 10^{-3}$)	0.4805 ± 0.0254	0.4654 ± 0.0260	0.3979 ± 0.0256	0.3879 ± 0.0231
	✓	✓	✓ ($\sigma = 10^0$)	0.4800 ± 0.0254	0.4648 ± 0.0260	0.3972 ± 0.0256	0.3871 ± 0.0230

Table 3. Estimated λ_i with standard deviation using Models 1–5 for simulated data with four provinces. Recall that the ground truth is $\lambda_1 = 0.5, \lambda_2 = 0.47, \lambda_3 = 0.4, \lambda_4 = 0.37$. Models in Column 1 are detailed in Table 2. Columns 2–4 indicate whether the model permits migration between provinces (for “M”), the heterogeneity of parameters (for “H”) and the Graph Laplacian prior (for “GL 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, for parameter inference using Model 5 $\mu = 10^{2.7}$, ($\sigma = 10^{-6}, 10^{-3}, 10^0$ respectively), at which the averaged relative validation errors over 100 replicas are minimized as shown in Supplementary Figs. S2 and S3.

The trajectories of a typical realization are plotted in Fig. 2 and the absolute errors of the fitted trajectories are shown in Fig. 3. As can be seen in these two figures, heterogeneity helps improve the prediction of testing data more than transportation, while introducing migration without heterogeneity of parameters worsens the estimate as can also be noticed from Table 4. More explanations can be found in “Further model evaluation”.

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

As shown in Figs. 2 and 3, Models 1 and 2 have slightly better generalization accuracy than Model 5 for Province 3. On the one hand, for data in this replica, the estimated λ_3 is 0.3860 using Model 5 and 0.4640 using Models 1 or 2 (recalling that the ground truth λ_3 is 0.4). On the other hand, due to the randomness of the generated testing data, the sampled newly confirmed cases are much more than the deterministic ones in Province 3 obtained by running (2.4) with the ground truth parameters. Hence, although all these estimates of λ_3 are biased from the ground truth 0.4, estimates using Models 1 and 2, which are biased up, lead to less absolute errors.

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 3. It can be observed that models allowing heterogeneity estimate parameters more accurately, and Model 5 that integrates the correlation leads to slightly better estimate for λ_2 . We can see that compared to Model 4, the estimates of smaller λ_k 's (such as $\lambda_2, \lambda_3, \lambda_4$) become larger, and the

Model	M	H	GL prior	$MAE_{(w)}^{[Tr]}$	$MAE_{(w)}^{[Te]}$	$MSE_{(w)}^{[Tr]}$	$MSE_{(w)}^{[Te]}$
1	×	×	×	0.1728 ± 0.0273	0.4700 ± 0.1879	0.2379 ± 0.0375	0.4801 ± 0.1887
2	✓	×	×	0.1828 ± 0.0296	0.5432 ± 0.2048	0.2487 ± 0.0390	0.5551 ± 0.2067
3	×	✓	×	0.1417 ± 0.0232	0.4095 ± 0.1692	0.1997 ± 0.0348	0.4173 ± 0.1683
4	✓	✓	×	0.1418 ± 0.0229	0.3834 ± 0.1672	0.1997 ± 0.0343	0.3900 ± 0.1665
5	✓	✓	✓ ($\sigma = 10^{-6}$)	0.1481 ± 0.0230	0.3300 ± 0.1519	0.2069 ± 0.0355	0.3365 ± 0.1513
	✓	✓	✓ ($\sigma = 10^{-3}$)	0.1481 ± 0.0230	0.3300 ± 0.1519	0.2069 ± 0.0355	0.3365 ± 0.1513
	✓	✓	✓ ($\sigma = 10^0$)	0.1482 ± 0.0230	0.3283 ± 0.1498	0.2071 ± 0.0356	0.3347 ± 0.1490

Table 4. Training and testing errors with standard deviation of Models 1–5 for simulated data with four provinces. The formulas of errors are detailed in Sect. B (Eq. (S5)) of Supplementary Information. Remarks for Columns 1–4 and the choice of $\mu = 10^{2.7}$ in Model 5 are the same as those in Table 3. $MAE_{(w)}^{[Tr]}$, $MAE_{(w)}^{[Te]}$, $MSE_{(w)}^{[Tr]}$ and $MSE_{(w)}^{[Te]}$ are computed for each of the 100 replicas, then the mean and standard deviation are presented in the table above.

estimate of λ_1 which has the largest value becomes smaller, since the graph Laplacian penalty tends to make $\{\lambda_k\}_{k=1}^n$ closer to each other.

Moreover, we performed a sensitivity analysis for the hyper-parameter σ to check that the results are robust to σ . The last two rows of Table 3 show the parameter estimation results for Model 5 with $\sigma = 10^{-3}$ and 10^0 respectively (more values of $\sigma \in [10^{-8}, 10^0]$ are tested and the results are similar as well). We observe that the variation of parameters estimated by Model 5 with σ varying from 10^{-6} to 10^0 does not exceed 1%. Therefore, the parameter estimation results are not sensitive to the choice of σ as long as σ is not too large.

Further model evaluation. The training and testing errors, $MAE_{(w)}^{[Tr]}$, $MAE_{(w)}^{[Te]}$, $MSE_{(w)}^{[Tr]}$, $MSE_{(w)}^{[Te]}$, as defined in Sect. B (in the Eq. (S5)) of Supplementary Information, are listed below in Table 4 with mean and standard deviation. We remind the readers that the superscripts $[Tr]$ and $[Te]$ represent the errors are computed on training and testing data respectively, and the subscript (w) denotes that the error is weighted average of the daily relative errors over time. It can be seen from Table 4 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 regularization leads to better prediction performance in average, which might not be obvious in this case due to the relatively large variance. The advantage of the proposed Model 5 is more evident for larger number of regions involved in the dynamic system, as shown in the next “Thirty provinces case”.

In addition, it can be seen from Table 4 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 values of λ_k 's are lower than the true values of λ_k 's for group 1, which leads to that the estimated newly confirmed cases are fewer than the true ones for provinces 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.

Furthermore, the last two rows of Table 4 report the training and testing errors for Model 5 with the same $\mu = 10^{2.7}$ while $\sigma = 10^{-3}$ and $\sigma = 10^0$ respectively. As a consequence of the robustness of the parameter estimation regarding σ , the errors of Model 5 are also robust to σ . The similar analysis is also performed for the other data sets and the similar results can be obtained which we do not report repetitively. Hereinafter, the results are presented with $\sigma = 10^{-6}$.

The plots of the mean of weighted and simply averaged testing errors $MAE_{(w)}^{[Te]}$, $MSE_{(w)}^{[Te]}$, $MAE_{(s)}^{[Te]}$ and $MSE_{(s)}^{[Te]}$ against varying μ are shown in Supplementary Figs. S2 and S3 respectively. Recall that the subscripts (w) and (s) denote the weighted and simple average respectively. Note that Model 5 with $\mu = 10^{2.7}$, at which the averaged validation errors over replicas are minimized, achieves the minimal values of testing errors $MAE_{(w)}^{[Te]}$ and $MSE_{(w)}^{[Te]}$ (also $MAE_{(s)}^{[Te]}$ and $MSE_{(s)}^{[Te]}$). This is because validation and testing errors have the same distribution in this case.

Thirty provinces case. In this simulation study, we present the results for the simulated data involving a larger number of regions.

Data description. We set the total number of regions $n = 30$, the total days considered $T = 30$, and set the threshold that separates the training and testing set as $T_{th} = 10$. The other prefixed parameters are listed as below:

- For $k \in \{1, 2, \dots, n\}$, $N_k(0) = 10^6$, $E_k(0) = 30$, $H_k(0) = 10$.
- For $l \in \{1, \dots, T\}$, $i, j \in \{1, 2, \dots, n\}$ and $i \neq j$, $(W_l)_{ij} = 5 \times 10^3$.

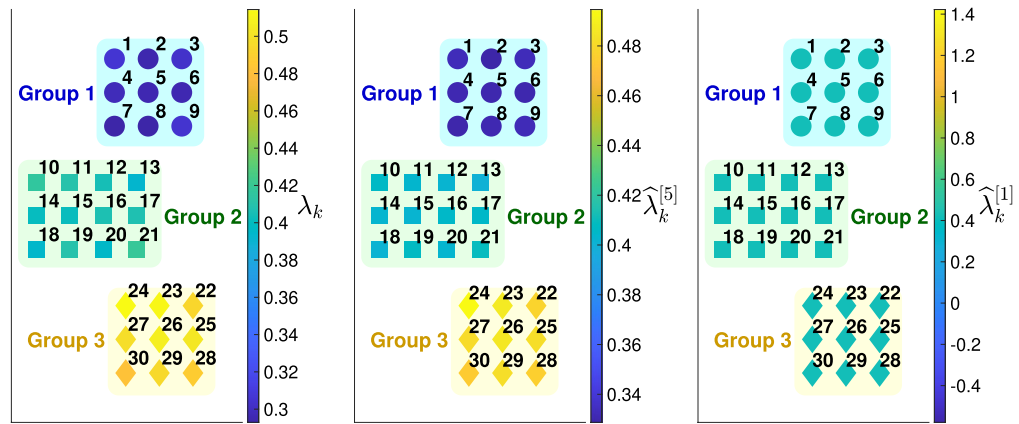


Figure 4. Maps of the thirty provinces divided into three groups, colored by transmission parameters or their estimates. In all the three panels, the circles on top of the panels denote the nine provinces in Group 1 (indexed by 1–9), the squares in the middle of the panels denote twelve provinces in Group 2 (indexed by 10–21), and the diamonds at the bottom of the panels denote nine provinces in Group 3 (indexed by 22–30). The provinces are colored by the ground truth $\{\lambda_k\}_{k=1}^{30}$ in the left panel, by the averaged estimates $\{\hat{\lambda}_k^{[5]}\}_{k=1}^{30}$ using Model 5 in the middle panel, and by the averaged estimates $\{\hat{\lambda}_k^{[1]}\}_{k=1}^{30}$ using Model 1 in the right panel. The superscripts $^{[5]}$ and $^{[1]}$ denote the models used to obtain the estimates of λ_k . In the experiments, the traveling volumes between provinces are taken to be constant.

- As in the four provinces case, we assign thirty provinces into different groups and reflect the similarity between provinces in the choice of λ_k 's, namely the values of λ_k 's being closer for provinces in the same groups. The three groups are denoted as D_1, D_2, D_3 below and the partition is denoted as P . Then, the proposed model (Model 5) can be applied as described in “[Estimation of model parameters](#)” with or without heterogeneity in the graph information (the ground truth partition P being fully known).

Figure 4 shows maps of the thirty provinces colored by transmission parameters or their estimates. It can be observed that λ_k 's of provinces in the same groups are closer. Furthermore, the estimates of λ_k 's using Model 5 with heterogeneity are close to the ground truth. In contrast, the estimates using Model 1 without heterogeneity are the same for all the provinces and deviate from the ground truth λ_k for most of the provinces. More results of trajectory prediction and parameter estimation can be found in “[Results of trajectory prediction](#)” and “[Results of parameter estimation](#)”.

Details of assignment of the thirty provinces, choice of transmission rates and their values are listed in Sect. C of Supplementary Information.

Models to compare. We compare the Models 1–5 listed in Table 2 as detailed in “[Models to compare](#)”.

Parameter inference of the five models. For Model 5, the estimation of Θ still follows the procedure in “[Estimation of model parameters](#)”. For the inference of $\bar{\Theta}$, when we assume that the ground truth group division P is known, the general formula (2.13) can be specified as follows:

$$\bar{\Theta}^* = \arg \min \left(-\log(p(y_{1:T_{th}}|\bar{\Theta})) + \frac{\mu}{2} \left(\sum_{m=1}^3 \sum_{i,j \in D_m} (\lambda_i - \lambda_j)^2 \right) + \beta \sum_{1 \leq m_1 < m_2 \leq 3} \sum_{i \in D_{m_1}, j \in D_{m_2}} (\lambda_i - \lambda_j)^2 \right) + \sigma \sum_i \lambda_i^2, \tag{3.2}$$

where $\beta = 0.1, \sigma = 10^{-6}$. The parameter inference for Models 1–4 is the same as described in “[Models to compare](#)”.

Mismatched partitions of regions. For Model 5, we also report results when there are mismatches in partition of the provinces, since in real-world applications the graph information may not be fully known. Specifically, since most of the results reported in the following sections are for Provinces 1 (in D_1), 10 (in D_2), and 22 (in D_3), we consider the other two partitions $P' = \{D'_1, D'_2, D'_3\}$ and $P'' = \{D''_1, D''_2, D''_3\}$, which are different from P in the assignments of Provinces 1, 10, and 22, and another three provinces respectively. Comparison between the results using the ground truth partition and the results of these two different kinds of mismatches may reflect the potentially different impact of these mismatches on the results. For Model 5 with partitions P' or P'' , $\bar{\Theta}$ is estimated by (3.2) with $P = \{D_1, D_2, D_3\}$ replaced by $P' = \{D'_1, D'_2, D'_3\}$ or $P'' = \{D''_1, D''_2, D''_3\}$.

The partition $P' = \{D'_1, D'_2, D'_3\}$ deviates from P in the assignment of Provinces 1 (in D_1), 10 (in D_2), and 22 (in D_3):

- Group 1 (D'_1): Provinces 22, 2–9.

True trajectory & Fitted trajectories

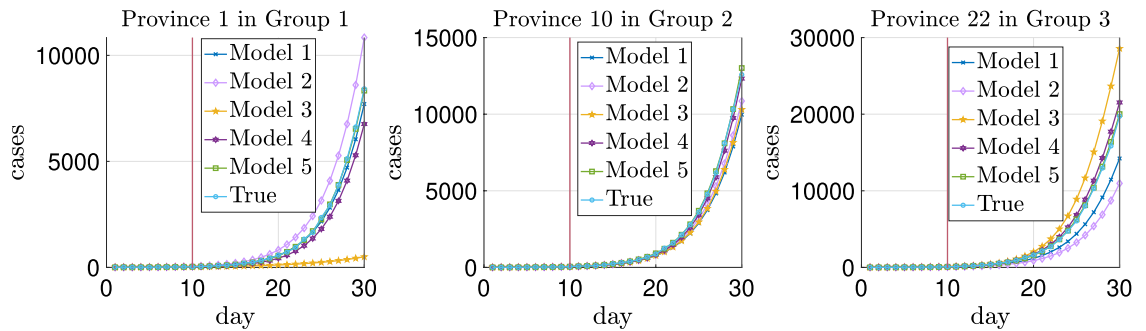


Figure 5. True and fitted trajectories for the simulated data in the three provinces chosen from the total thirty provinces. The vertical lines show the threshold of training-testing split. For Model 5, $\mu^* = 10^{1.8}$, at which the averaged validation errors over replicas are minimized (as shown in Fig. 7).

Absolute errors of predicted trajectories for all models

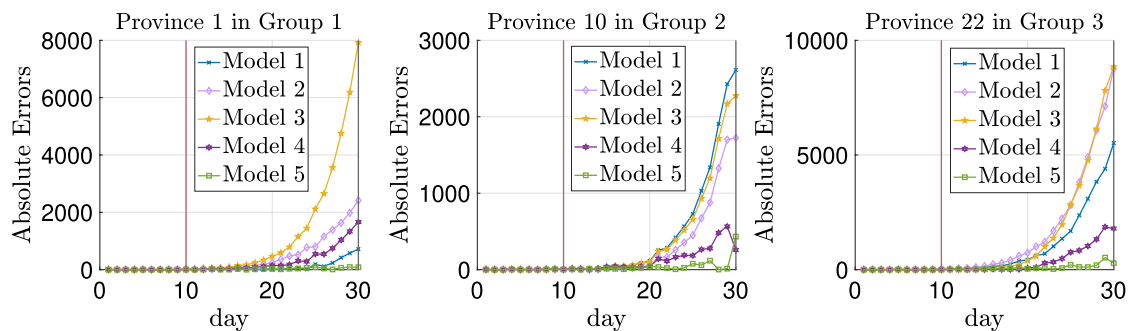


Figure 6. Absolute errors of the fitted trajectories for simulated data in the three provinces chosen from the total thirty ones. The vertical lines show the threshold of training-testing split.

- Group 2 (D'_2): Provinces **1**, 11–21.
- Group 3 (D'_3): Provinces **10**, 23–30.

The partition $P'' = \{D''_1, D''_2, D''_3\}$ deviates from P in the assignment of Provinces **2** (in D_1), **11** (in D_2), and **23** (in D_3):

- Group 1 (D''_1): Provinces **1**, **23**, 3–9.
- Group 2 (D''_2): Provinces **10**, **2**, 12–21.
- Group 3 (D''_3): Provinces **22**, **11**, 24–30.

For mismatched partitions P' and P'' , the provinces assigned to the wrong groups are marked in bold.

Results of trajectory prediction. Note that for Model 5 (with the ground truth partition P), we choose $\mu^* = 10^{1.8}$, at which the averaged validation errors over replicas are minimized, as marked with orange pentagrams in Fig. 7 and Supplementary Fig. S4. The following results for Model 5 are all obtained with $\mu^* = 10^{1.8}$.

We choose three provinces from the total thirty provinces (one province from each group), and plot the prediction of the trajectories and also the absolute errors from one specific replica as in Figs. 5 and 6. It can be observed from Fig. 6 that Model 5 (the green lines) achieve the most accurate prediction results due to the graph Laplacian regularization. Model 4 (the dark purple lines) also have good generalization performance, thanks to the heterogeneity and transportation involved in this model.

We note that Model 3 also behaves much worse than the best models, especially in Provinces **1** (in D_1) and **22** (in D_3). In fact, the estimates for the λ_k 's in these three provinces using Model 3 are 0.2908, 0.4198, and 0.4639, respectively, which are not much different from the ground truth values. Nevertheless, Model 3 does not involve transportation between provinces. This leads to that the confirmed cases of Province **22** in D_3 (which are much more than the cases of provinces in the other two groups) are not output to other provinces; hence the predicted newly confirmed cases are much more than the truth. A similar situation happens in the case of Province **1** (in

Model	M	H	GL prior	λ_1	λ_{10}	λ_{22}
1	×	×	×	0.4223 ± 0.0067		
2	✓	×	×	0.4227 ± 0.0067		
3	×	✓	×	0.3743 ± 0.0349	0.4101 ± 0.0284	0.4537 ± 0.0324
4	✓	✓	×	0.3051 ± 0.0383	0.4093 ± 0.0350	0.4902 ± 0.0326
(P)	✓	✓	✓	0.3332 ± 0.0172	0.4071 ± 0.0175	0.4784 ± 0.0209
5(P')	✓	✓	✓	0.3669 ± 0.0148	0.4335 ± 0.0180	0.4375 ± 0.0218
(P'')	✓	✓	✓	0.3430 ± 0.0174	0.4051 ± 0.0174	0.4761 ± 0.0210

Table 5. Estimated λ_i with standard deviation using Models 1–5 for simulated data with thirty provinces. The ground truth is $\lambda_1 = 0.307$ (in D_1), $\lambda_{10} = 0.416$ (in D_2), $\lambda_{22} = 0.493$ (in D_3) as listed in Sect. C of Supplementary Information. Remarks for Columns 1–4 and estimates of λ_k 's are the same as those in Table 3. In addition, for inference of parameters using Model 5, $\mu^* = 10^{1.8}$, at which the averaged validation errors over 100 replicas (computed with partition P) are minimized as shown in Fig. 7 and Supplementary Fig. S4. The partitions of Model 5 are P , P' , and P'' respectively. P is the ground truth underlying graph structure, and P' and P'' are mismatched partitions introduced in “Models to compare”.

Model	M	H	GL prior	$\text{MAE}_{(w)}^{[Tr]}$	$\text{MAE}_{(w)}^{[Te]}$	$\text{MSE}_{(w)}^{[Tr]}$	$\text{MSE}_{(w)}^{[Te]}$
1	×	×	×	0.1739 ± 0.0092	0.1704 ± 0.0543	0.2445 ± 0.0135	0.1744 ± 0.0534
2	✓	×	×	0.2427 ± 0.0136	0.3387 ± 0.0573	0.3128 ± 0.0171	0.3420 ± 0.0566
3	×	✓	×	0.1541 ± 0.0082	0.4920 ± 0.0594	0.2212 ± 0.0137	0.4973 ± 0.0599
4	✓	✓	×	0.1549 ± 0.0083	0.1960 ± 0.0647	0.2200 ± 0.0139	0.1994 ± 0.0649
(P)	✓	✓	✓	0.1648 ± 0.0083	0.1495 ± 0.0579	0.2312 ± 0.0135	0.1531 ± 0.0570
5(P')	✓	✓	✓	0.1702 ± 0.0084	0.1612 ± 0.0522	0.2364 ± 0.0135	0.1650 ± 0.0514
(P'')	✓	✓	✓	0.1707 ± 0.0085	0.1648 ± 0.0539	0.2372 ± 0.0134	0.1685 ± 0.0532

Table 6. Training and testing errors of Models 1–5 with standard deviation for simulated data with thirty provinces. The formulas of errors are detailed in Sect. B (Eq. (S5)) of Supplementary Information. Remarks for Columns 1–4 and presented results of errors are the same as those in Table 4. In addition, the remarks for the choice of $\mu^* = 10^{1.8}$ in Model 5 and the partitions P , P' , and P'' are the same as those in Table 5.

D_1). Namely, Model 3 does not take the imported cases from the provinces in the other two groups into account, causing the predicted newly confirmed cases to be much less than the truth.

Results of parameter estimation. The mean and standard deviation of λ_k 's in Provinces 1 (in D_1), 10 (in D_2), and 22 (in D_3) are reported in Table 5. We can still see that Models 3–5 estimate the transmission parameters more accurately than Models 1 and 2 especially for Provinces 1 and 22. In addition, the estimates by Model 5 of λ_k are closer than those by Model 4, due to the existence of the regularization term.

The last three rows show the parameter estimates using Model 5 with three partitions P , P' , and P'' respectively. Recall that P is the ground truth underlying graph structure, and P' and P'' are mismatched partitions introduced in “Models to compare”. It can be seen that compared with the estimated parameters using Model 5 with P' that differs from P in the grouping of Provinces 1, 10, and 22, the estimates using Model 5 with P'' that differs from P in the grouping of Provinces 2, 11, and 23 are closer to the estimates using Model 5 with the ground truth partition P . Hence, the results imply that for the parameter inference of simulated data generated from a certain partition, the incorrect division causes more discrepancy in the mismatched regions when comparing the estimates using the ground truth partition.

Further model evaluation. The mean and standard deviation of the weighted training and testing errors (formulas can be found in Sect. B of Supplementary Information, specifically in Eq. (S5)) are presented in Table 6. We see that Model 5 with $\mu^* = 10^{1.8}$ achieves the minimum testing errors among all the models, and hence have the best generalization performance.

The last three rows show the comparison of errors for Model 5 with three partitions P , P' , and P'' respectively. We may see that the testing errors of Model 5 with the incorrect partitions are basically the same and larger than those with the ground truth partitions. Nevertheless, these errors are still smaller than the other baseline models. Therefore, the incorrect division might worsen the generalization performance of Model 5; however, when the division does not deviate from the ground truth much, it still behaves better than Models 1–4 with uniform prior.

Moreover, as shown in Table 6, both Model 2, which allows transportation but not heterogeneity, and Model 3, that introduces heterogeneity but no transportation, behave worse than the baseline Model 1. Model 2 having more significant testing errors is due to the same reason as the case of four provinces, as has been analyzed in

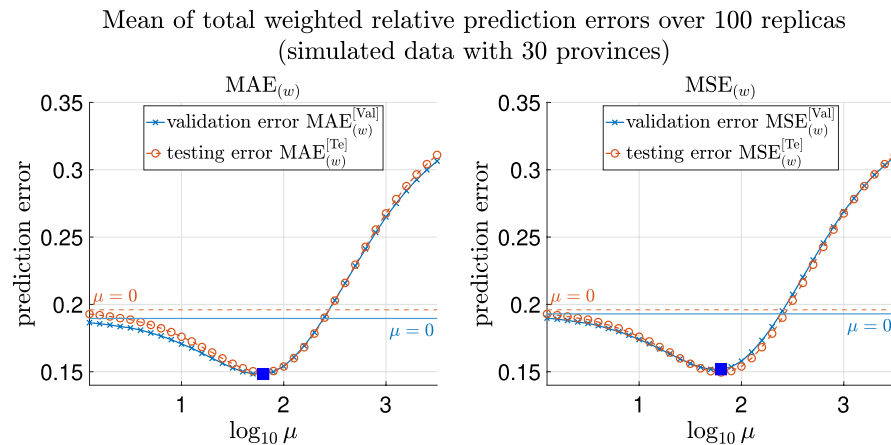


Figure 7. Testing and validation errors on simulated data with thirty provinces. Testing and validation errors on simulated data with four provinces. Left: The weighted prediction errors on validation set ($MAE_{(w)}^{[Val]}$) and testing set ($MAE_{(w)}^{[Te]}$) respectively, plotted vs. the values of μ . The errors are averaged over 100 replicas of experiment. Right: Same plot of MSE error. In each plot, the blue and red horizontal lines show the values of the averaged errors when $\mu = 0$. Note that both the MAE and MSE validation errors are minimized at $\mu^* = 10^{1.8}$, which is marked by blue squares in both plots. The construction of training/validation/testing data is detailed in Sect. A.1.1 of Supplementary Information, and the formulas of computing the errors can be found in Sect. B of Supplementary Information.

“Further model evaluation”. The reason for Model 3 not being able to predict well is explained in detail in “Results of trajectory prediction”.

However, we also note that Model 4, which involves both transportation and heterogeneity of parameters, has greater testing errors than Model 1 on average. This phenomenon may be partly explained by the fact that, on the one hand, there may be more infected cases imported from regions with higher λ_k 's to those with lower transmission rates. On the other hand, although the transmission parameters are the same and all-around 0.42 in Model 1, E_0 and H_0 are still allowed to be spatially heterogeneous, which makes the trajectories not that far away from the ground truth. Meanwhile, though the λ_k 's are allowed to be spatially heterogeneous in Model 4, it often underestimates λ_k 's for provinces in D_1 and overestimates λ_k 's for provinces in D_3 , making the trajectories deviate from the ground truth.

Figure 7 and Supplementary Fig. S4 plot the change of weighted and simply averaged testing errors $MAE_{(w)}^{[Te]}$, $MSE_{(w)}^{[Te]}$, $MAE_{(s)}^{[Te]}$ and $MSE_{(s)}^{[Te]}$ regarding μ (recall that the subscripts (w) and (s) denote weighted and simple average respectively). We can still observe that Model 5 with the minimizer of validation errors also achieves the minimal testing errors, since the validation and testing errors have the same distribution for the simulated data.

Experimental results on COVID-19 data

In this section, we apply our model to two real-world data sets, the COVID-19 data in China (from January to early February 2020) and Europe (from May to August 2020), in “Results for COVID-19 data in China” and “Results for COVID-19 data in Europe” respectively. We remark that the regions in “Methods” refer to provinces or municipalities for the data in China and refer to countries for the data in Europe. Before reporting results, we first introduce the data sources in “Data sources”. Then, we generalize our inference method to the case that the transmission rates have temporal heterogeneity in “Model extension by allowing time-varying parameters”, which might happen in real-world data due to travel restrictions or other reasons. For the two real-world data sets, we compare the proposed model with other baselines in the aspects of trajectory prediction and quantitative evaluation. We also show the estimation of transmission parameters with posterior distributions using our proposed model. Details can be found in “Results for COVID-19 data in China” and “Results for COVID-19 data in Europe”.

For the real-world data in Europe, since there is not consensus on proper divisions, we presented the results from two different partitions, both of which are based on geographical locations of the countries. The results imply that in comparison to the simulation data where the data are generated in the artificially divided regions, the generalization performances for the real-world data of both the partitions are similar and both much better than the baseline models.

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

Real-world data in China. The publicly available pandemic data in China include the number of confirmed cases and removed cases (consisting of recovered cases and fatalities) in China's major provinces and municipalities from January 21st to March 28th, 2020. These publicly available data are downloaded from the websites of the

National Health Commission of the People's Republic of China¹⁶ and Chinese Center for Disease Control and Prevention¹⁷. The corresponding total population in each province or municipality is from Ref.⁵¹.

We also utilize two transportation data sets in China extracted from Baidu Qianxi⁵², lasting from January 10th to February 10th in 2020: (1) the migration indexes which reflects the number of people moving out from the provinces, and (2) the migration percentage which reflects the percentage of the population moving to a destination province from the origin province. The traveling volume is estimated by combining these two data sets.

Real-world data in Europe. The publicly available pandemic data in Europe include the number of confirmed and removed cases (consisting of recovered cases and fatalities) in the following 11 countries from May 1st to August 31st in 2020: Denmark, Finland, Norway, Austria, Germany, Switzerland, Italy, Spain, Belgium, France, and Ireland. The data are downloaded from Refs.^{13,53}, where the data are collected from European Center for Disease Prevention and Control¹⁸. The data of total population in each country is obtained from Ref.⁵⁴. During the study, we are not able to obtain suitable transportation data between these European countries needed by our model, and thus we simplify the proposed model by incorporating only spatial and temporal heterogeneity but no transportation in this experiment, see details in “Results for COVID-19 data in Europe”.

We remark that the data in China are from the initial outbreak of the epidemic. On the contrary, the data in Europe are from when the epidemic has been ongoing for about six months. The choice of the aforementioned time period of interest is to illustrate that the proposed model could predict both the case when the epidemic breaks out from some seed region and the case when the epidemic has already established a sizable local spread within different regions.

Model extension by allowing time-varying parameters. To capture the trend that the transmission parameters might change in real-world data due to travel restrictions or other factors, we first extend the model and parameter estimation procedure described in “Methods” so that $\{\lambda_k\}$ are allowed to be time-varying.

Specifically, suppose that λ_k is a piece-wise constant function of t , and $\lambda_k = \lambda_k^{(1)}$ for days before some threshold T_k , and $\lambda_k = \lambda_k^{(2)}$ afterwards.

We denote $\lambda^{(1)} = (\lambda_1^{(1)}, \dots, \lambda_n^{(1)})^T$ and $\lambda^{(2)} = (\lambda_1^{(2)}, \dots, \lambda_n^{(2)})^T$ as the vectors of transmission rates in the two periods respectively. The stochastic dynamic model introduced in “Model description” remains unchanged except that the transmission rates will vary as time increases.

As for the parameter estimation, the first step of obtaining $\gamma_k^* = \arg \max_{\gamma_k} \prod_{i=1}^T \text{Pois}((\Delta R_a)_k(i) | ((C_a)_k(i-1) - (R_a)_k(i-1))\gamma_k)$ also stays the same as described in “Estimation of model parameters”.

Then, when estimating $\bar{\Theta} = \{E_k(0), H_k(0), \lambda_k^{(1)}, \lambda_k^{(2)}\}_{k=1}^n$ and the marginal posterior distributions of $\bar{\Theta}$, we modify the prior distribution $p(\bar{\Theta})$ as below,

$$P(\bar{\Theta}) = \frac{1}{C_{\sigma,A}} \exp(-\lambda^{(1)T} (D - A)\lambda^{(1)} - \sigma \|\lambda^{(1)}\|_2^2 - \lambda^{(2)T} (D - A)\lambda^{(2)} - \sigma \|\lambda^{(2)}\|_2^2), \quad (4.1)$$

where σ , D and A have the same meaning as in (2.10), and $C_{\sigma,A} = \int_{\mathbb{R}^{2n}} \exp(-\lambda^{(1)T} (D - A)\lambda^{(1)} - \sigma \|\lambda^{(1)}\|_2^2 - \lambda^{(2)T} (D - A)\lambda^{(2)} - \sigma \|\lambda^{(2)}\|_2^2) d\lambda^{(1)} d\lambda^{(2)}$ is still the normalizing constant. Consequently,

$$\begin{aligned} V(\bar{\Theta}) &= -\log \left(P \left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta} \right) \right) - \log(P(\bar{\Theta})) \\ &= -\log \left(P \left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta} \right) \right) + \lambda^{(1)T} (D - A)\lambda^{(1)} + \sigma \|\lambda^{(1)}\|_2^2 + \lambda^{(2)T} (D - A)\lambda^{(2)} \\ &\quad + \sigma \|\lambda^{(2)}\|_2^2 + \log C_{\sigma,A} \\ &= -\log \left(P \left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta} \right) \right) + \frac{1}{2} \sum_{ij} a_{ij} (\lambda_i^{(1)} - \lambda_j^{(1)})^2 + \sigma \|\lambda^{(1)}\|_2^2 + \frac{1}{2} \sum_{ij} a_{ij} (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \\ &\quad + \sigma \|\lambda^{(2)}\|_2^2 + \log C_{\sigma,A}. \end{aligned} \quad (4.2)$$

Therefore, by the definition of a_{ij} in (2.12), the estimation of $\bar{\Theta}$ by minimizing $V(\bar{\Theta})$ can be equivalently written as

$$\begin{aligned}
 \bar{\Theta}^* &= \arg \min V(\bar{\Theta}) \\
 &= \arg \min \left(-\log \left(P \left(\{(\Delta C_a)_k(i)\}_{k,i} \middle| \bar{\Theta} \right) \right) + \frac{1}{2} \sum_{ij} a_{ij} (\lambda_i^{(1)} - \lambda_j^{(1)})^2 + \sigma \|\lambda^{(1)}\|_2^2 \right. \\
 &\quad \left. + \frac{1}{2} \sum_{ij} a_{ij} (\lambda_i^{(2)} - \lambda_j^{(2)})^2 + \sigma \|\lambda^{(2)}\|_2^2 \right) \\
 &= \arg \min \left(-\log \left(P \left(\{(\Delta C_a)_k(i)\}_{k,i} \middle| \bar{\Theta} \right) \right) + \frac{\mu}{2} \left(\sum_m \sum_{ij \in D_m} W_{ij} \left((\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \right) \right) \right. \\
 &\quad \left. + \beta \sum_{m_1 < m_2} \sum_{i \in D_{m_1}, j \in D_{m_2}} W_{ij} \left((\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \right) \right) + \sigma \sum_i ((\lambda_i^{(1)})^2 + (\lambda_i^{(2)})^2).
 \end{aligned}
 \tag{4.3}$$

After choosing W and σ , $\bar{\Theta}$ can be inferred through (4.3). Then, we could estimate the marginal posterior distributions of $\bar{\Theta}$ by MCMC sampling with the initial point $\bar{\Theta}^*$.

We would like to remark that same as in the simulation study, since the proposed model depends on the penalty factor μ as shown in (4.3), μ is chosen so that the corresponding validation error achieves or is slightly higher than the minimum. Specifically, results from the multiple choices of μ are present, since for the real-world data, the validation data may not have the same distribution as the testing data (the choice of the training, validation, and testing sets is detailed in Sects. A, D, and E of Supplementary Information).

Results for COVID-19 data in China. *Data description.* After the selection process described in Sect. D of Supplementary Information, $n = 21$ provinces or municipalities are taken into consideration, which are Anhui, Beijing, Fujian, Gansu, Guangdong, Guangxi, Hebei, Henan, Hubei, Hunan, Jiangsu, Jiangxi, Liaoning, Ningxia, Shandong, Shan-Xi, Shanxi, Shanghai, Sichuan, Zhejiang, and Chongqing. Furthermore, since the epidemic data last from January 21st to March 28th, 2020, there are 20 days in total.

In addition, from observation of data from January 21st to February 10th, the transmission rates in the selected provinces or municipalities change after some specific time point. Thus, we allow $\{\lambda_k\}$ to be time-varying for this data set with appropriate changes to the model described in “Methods”, which are detailed in “Model extension by allowing time-varying parameters”. More details of the experimental settings can be found in Sect. D of Supplementary Information, including construction of training/validation/testing data and the choice of T_k , the day that λ_k changes.

Furthermore, we remark that 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 Baidu migration index (which reflects numbers of people departed from the stating point) and the percentage of population traveling from this origin to the destination. The corresponding scaling parameter α also needs to be inferred from the data for all the models.

Models to compare. As in “Models to compare”, 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.

For Model 5, the parameter inference adopts the method in “Estimation of model parameters”, while allowing $\{\lambda_k\}_{k=1}^n$ to be time-varying as described in “Model extension by allowing time-varying parameters”. Note that the medical resources were overwhelmed in Wuhan at the early stage of the pandemic⁵⁵. When applying the general formula (4.3), we divide the provinces into two groups, which are Hubei and other provinces except Hubei. Furthermore, the affinity matrix W is constructed as averaged traveling volumes, that is $W_{ij} = \bar{W}_{ij} / \max_{i,j} \bar{W}_{ij}$, where $\bar{W}_{ij} = \frac{1}{2} (\frac{1}{T} \sum_{l=1}^T (w_l)_{ij} + (\frac{1}{T} \sum_{l=1}^T (w_l)_{ji}))$. Then, for COVID-19 data in China, (4.3) can be specified as follows to estimate $\bar{\Theta} = \{E_k(0), H_k(0), \lambda_k^{(1)}, \lambda_k^{(2)}\}_{k=1}^n$ and scaling parameter α :

$$\begin{aligned}
 \bar{\Theta}^*, \alpha^* &= \arg \min \left(-\log P \left(\{(\Delta C_a)_k(i)\}_{k,i} \middle| \bar{\Theta}, \alpha \right) + \frac{\mu}{2} \left(\sum_{m=1}^2 W_{ij} \left((\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \right) \right) \right. \\
 &\quad \left. + \beta \sum_{i \in D_1, j \in D_2} W_{ij} \left((\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \right) \right) + \sigma \sum_i ((\lambda_i^{(1)})^2 + (\lambda_i^{(2)})^2),
 \end{aligned}
 \tag{4.4}$$

where we take $\beta = 0.1, \sigma = 10^{-6}$, and $\lambda_k^{(1)}$ and $\lambda_k^{(2)}$ are the transmission rates in the k -th province before and after the T_k -th day respectively.

For Models 1–4, if the model does not have heterogeneity of transmission parameters, then $\{\lambda_1^{(1)}, \dots, \lambda_n^{(1)}\}$ are forced to be the same and so do $\{\lambda_1^{(2)}, \dots, \lambda_n^{(2)}\}$, while the transmission rates in the two periods are allowed to be different; if the model does not allow transportation between provinces, then terms involving W_t disappear in (2.4) as described in “Models to compare”.

Results of trajectory prediction. We first remark that for Model 5, three values of μ are chosen, and the trajectory prediction results presented below are from these three choices to perform the sensitivity analysis of μ . The first

True and predicted newly confirmed cases in Hubei

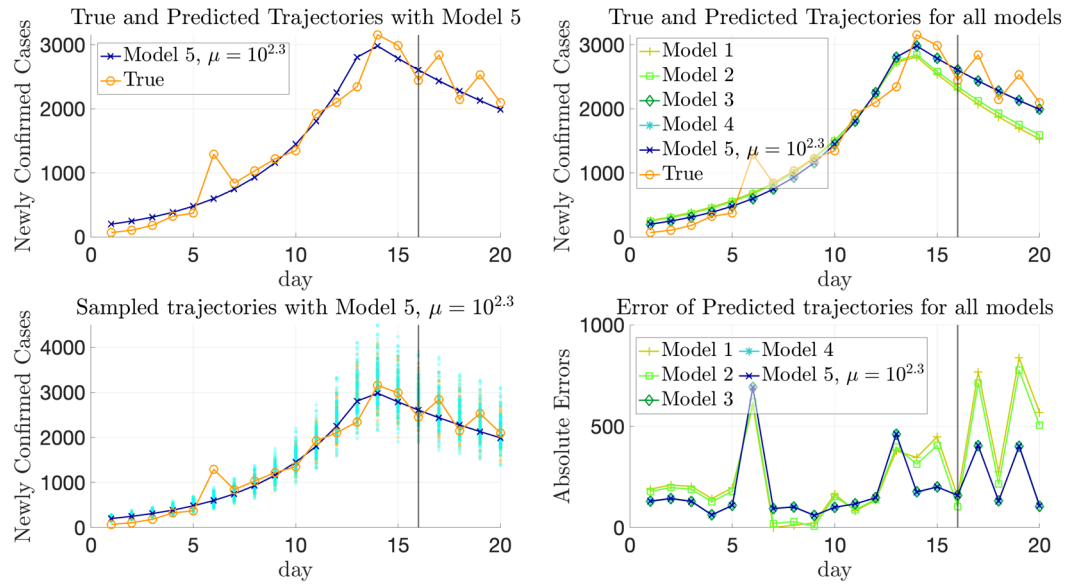


Figure 8. True and fitted trajectories in Hubei. The orange line with circles shows the true trajectory, the blue lines with crosses show the predicted deterministic trajectories using Model 5, blue and orange scatter plots show 100 stochastic trajectories with Θ^* inferred using Model 5 and sampled from the posterior distribution of Θ respectively. In each figure, the black vertical line shows the threshold of training-testing split. For Model 5, $\mu^* = 10^{2.3}$ is chosen, at which the validation errors achieves the minimum value, marked by blue squares in Fig. 12 and Supplementary Fig. S5.

True and predicted newly confirmed cases in Henan

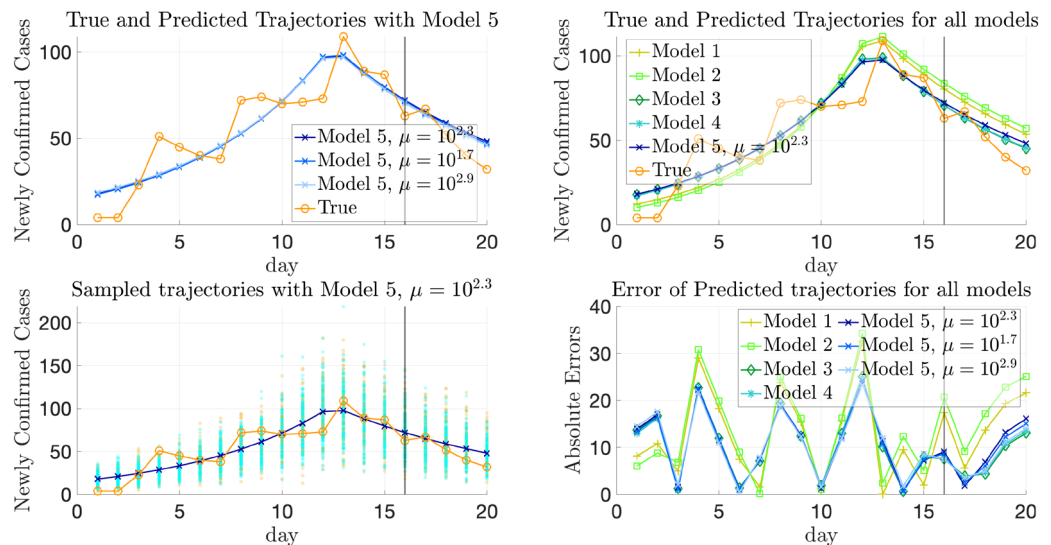


Figure 9. True and fitted trajectories in Henan. The remarks for the lines and scatter plots are the same as those in Figure 8. For Model 5, $\mu = 10^{1.7}$, $10^{2.3}$, and $10^{2.9}$ are chosen. As shown in Fig. 12 and Supplementary Fig. S5, the validation errors are minimized at $\mu^* = 10^{2.3}$. $\mu = 10^{1.7}$ and $10^{2.9}$ are obtained by perturbing the minimizer $\mu = 10^{2.3}$ without increasing validation errors much. $\mu = 10^{1.7}$, $10^{2.3}$, and $10^{2.9}$ are marked by yellow pentagrams, blue squares, and green diamonds respectively in Fig. 12 and Supplementary Fig. S5.

True and predicted newly confirmed cases in Anhui

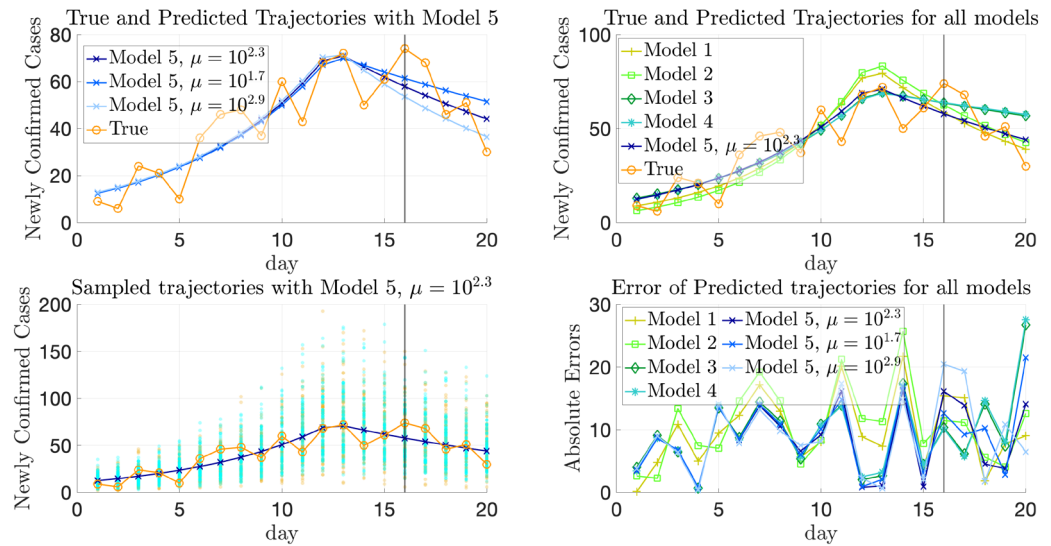


Figure 10. True and fitted trajectories in Anhui. The remarks for the lines and scatter plots are the same as those in Fig. 8. The choice of $\mu = 10^{1.7}$, $10^{2.3}$, and $10^{2.9}$ has been explained in Fig. 9.

	θ^*	Mean of estimated posterior	Standard deviation of estimated posterior
$\hat{\lambda}_{\text{Hubei}}^{(1)}$	0.3566	0.3566	3.9623×10^{-4}
$\hat{\lambda}_{\text{Hubei}}^{(2)}$	0.0723	0.0716	0.0076

Table 7. Estimated transmission rates in Hubei. θ^* is inferred using (4.4) with $\mu = 10^{2.3}$, at which the validation errors are minimized as shown in Fig. 12 and Supplementary Fig. S5. The posterior distribution is estimated by 5×10^5 MCMC iterations.

value is $\mu^* = 10^{2.3}$, at which the relative validation errors ($\text{MAE}_{(w)}^{[Val]}$, $\text{MSE}_{(w)}^{[Val]}$, $\text{MAE}_{(s)}^{[Val]}$, and $\text{MSE}_{(s)}^{[Val]}$) are minimized, marked with blue squares in Fig. 12 and Supplementary Fig. S5. The other two values are $\mu = 10^{1.7}$ and $\mu = 10^{2.9}$ obtained by perturbing the minimizer $\mu^* = 10^{2.3}$, whose validation errors are slightly larger, marked with orange pentagrams and green diamonds in Fig. 12 and Supplementary Fig. S5 respectively. Since for the real-world data in China, the validation error does not necessarily have the same distribution as the testing error, the results from various choices of μ are presented for better comparison.

Figure 8 shows the true and predicted trajectories for newly confirmed cases in Hubei. In Fig. 8,

- The orange line with circles shows the true trajectory.
- The blue lines with crosses show the predicted deterministic trajectories obtained by running (2.4) with Θ^* inferred using (4.4) with $\mu^* = 10^{2.3}$. The predicted trajectories with $\mu = 10^{1.7}$ and $10^{2.9}$ are close to the ones with $\mu^* = 10^{2.3}$, and are therefore not included in Fig. 8.
- The blue scatter plots show 100 stochastic trajectories with Θ^* inferred using Model 5 with $\mu^* = 10^{2.3}$.
- The orange scatter plots show 100 stochastic trajectories sampled from the posterior distribution of Θ , which is estimated by 5×10^5 MCMC iterations starting from Θ^* inferred using Model 5 with $\mu^* = 10^{2.3}$.

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 all the predicted deterministic trajectories generally capture the trend of the true trajectory. Besides, it can be seen from Fig. 8 that the performances of all the models that include heterogeneity of parameters 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. Moreover, sampling trajectories from the posterior distribution of Θ generates more randomness than sampling trajectories with Θ^* .

Similarly, Fig. 9 shows the predicted trajectories in Henan. When the transmission parameters 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. Thus, heterogeneity helps improve the performance of fitting and predicting. Additionally, although the validation errors achieves the minimum at $\mu^* = 10^{2.3}$, it can be seen

Estimated posterior distribution for λ in Hubei

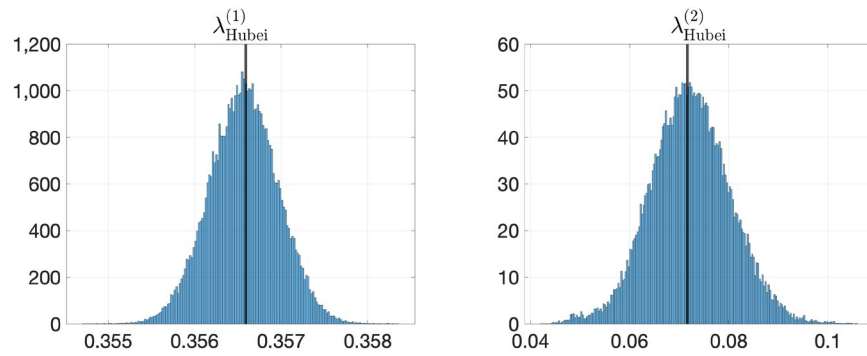


Figure 11. Estimated posterior distribution of λ in Hubei. The vertical black lines represent the values of corresponding θ^* using Model 5.

Model	M	H	GL prior	MAE _(w) ^[Tr]	MAE _(w) ^[Te]	MSE _(w) ^[Tr]	MSE _(w) ^[Te]
1	×	×	×	0.399	0.497	0.650	0.626
2	✓	×	×	0.388	0.459	0.645	0.590
3	×	✓	×	0.296	0.398	0.565	0.511
4	✓	✓	×	0.296	0.400	0.562	0.514
5	✓	✓	✓ ($\mu = 10^{1.7}$)	0.296	0.373	0.565	0.477
	✓	✓	✓ ($\mu^* = 10^{2.3}$)	0.298	0.343	0.572	0.438
	✓	✓	✓ ($\mu = 10^{2.9}$)	0.303	0.328	0.582	0.416

Table 8. Training and testing errors of Models 1–5 for real-world data in China. The formulas of errors are detailed in Sect. B (Eq. (S5)) of Supplementary Information. The remarks for Columns 1–4 are the same as those in Table 4. In addition, the choice of $\mu = 10^{1.7}$, $10^{2.3}$, and $10^{2.9}$ in Model 5 has been explained in Fig. 9. Note that the validation errors are minimized at $\mu^* = 10^{2.3}$ as shown in Fig. 12.

Total weighted relative prediction errors (real-world data in China)

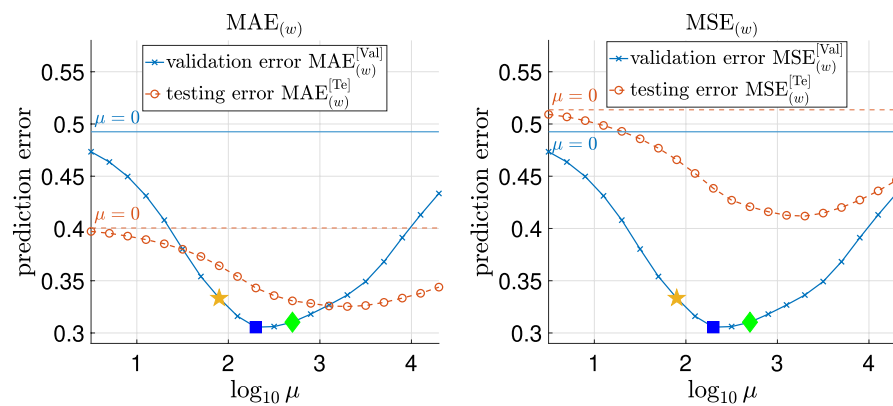


Figure 12. Testing and validation errors on real-world COVID-19 data in China. The two plots are similar to those in Figure 7, with errors computed on the real-world data in China instead of simulated data. Both the MAE and MSE validation errors are minimized at $\mu = 10^{2.3}$, which is marked by blue squares in both plots. $\mu = 10^{1.9}$ and $10^{2.7}$ are chosen by perturbing the minimizer $\mu = 10^{2.3}$, have slightly larger validation errors, and are marked with yellow pentagams and green diamonds in both plots. The construction of training/validation/testing data is detailed in Sects. A.1.2 and D of Supplementary Information.

from Fig. 9 that the deterministic trajectory obtained by Model 5 with $\mu = 10^{1.7}$ and $10^{2.9}$ achieves slightly better performance in prediction than the one obtained by Model 5 with $\mu^* = 10^{2.3}$. Therefore, the result suggests that the choice of μ is not necessarily limited to the minimizer of the validation error. Instead, we may also compare trajectories with μ that have slightly larger validation error for possibly better generalization performance.

Moreover, from Fig. 10 which shows the fitted trajectories in Anhui, it can be seen that heterogeneity helps the models fit the training trajectory better. Furthermore, adding regularization helps Model 5 capture the trend of the trajectory better than Model 4.

Results of parameter estimation. To illustrate the estimate of the posterior of parameters, Fig. 11 shows the estimated posterior distributions of $\lambda_k^{(1)}$ and $\lambda_k^{(2)}$ in Hubei using Model 5, in which the vertical black lines show the inferred $\lambda_k^{(1)}$ and $\lambda_k^{(2)}$ with $\mu = 10^{2.3}$, and Table 7 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.

Further model evaluation. Table 8 shows the training errors and testing errors of the five models computed as detailed in Sect. B of Supplementary Information. As can be seen from Table 8, 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 23rd⁵⁶, transportation no longer has much influence on the spread of the epidemic.

Figure 12 and Supplementary Fig. S5 plot weighted and simply averaged testing errors against varying μ respectively. We can see from Table 8 and Fig. 12 that though $\mu^* = 10^{2.3}$ achieves the minimal validation error, it has larger testing error than $\mu = 10^{2.9}$.

Results for COVID-19 data in Europe. *Data description.* The dataset contains numbers of daily COVID-19 cases in $n = 11$ countries in Europe, which are Denmark, Finland, Norway, Austria, Germany, Switzerland, Italy, Spain, Belgium, France, and Ireland. After preprocessing and removing the first and last three days of the original data, the dataset spans over a period of $T = 116$ days. More details of experimental settings can be found in Sect. E of Supplementary Information, including the selection of the countries, the preprocessing of the data, the construction of training/validation/testing data, and the choice of the day that λ_k changes in the selected countries.

Models to compare. **Three models for comparison.** For COVID-19 data in Europe, only three models detailed below are compared, which allow heterogeneity of parameters but do not include transportation. The last model is the proposed one in this paper, and the first two are baseline models. 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 restrictions⁵⁶.

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.

Two different partitions of the countries. For Model 3', 11 countries are partitioned into $d = 4$ groups according to geographical locations:

- D_1 (Northern Europe): Denmark, Finland, Norway;
- D_2 (Central Europe): Austria, Germany, Switzerland;
- D_3 (Southern Europe): Italy, Spain;
- D_4 (Western Europe): Belgium, France, Ireland.

We denote this partition as P . Since this might not be the unique appropriate partition for the $n = 11$ countries, we also present results from another partition denoted as P' in the following sections. P' also groups the countries that are geographically close together:

- D'_1 : Finland, Norway;
- D'_2 : Denmark, Austria, Germany;
- D'_3 : Spain, Belgium, France, Ireland;
- D'_4 : Switzerland, Italy.

The results for Model 3' in “Further model evaluation” are presented with both the partitions P and P' .

True and predicted newly confirmed cases in Austria

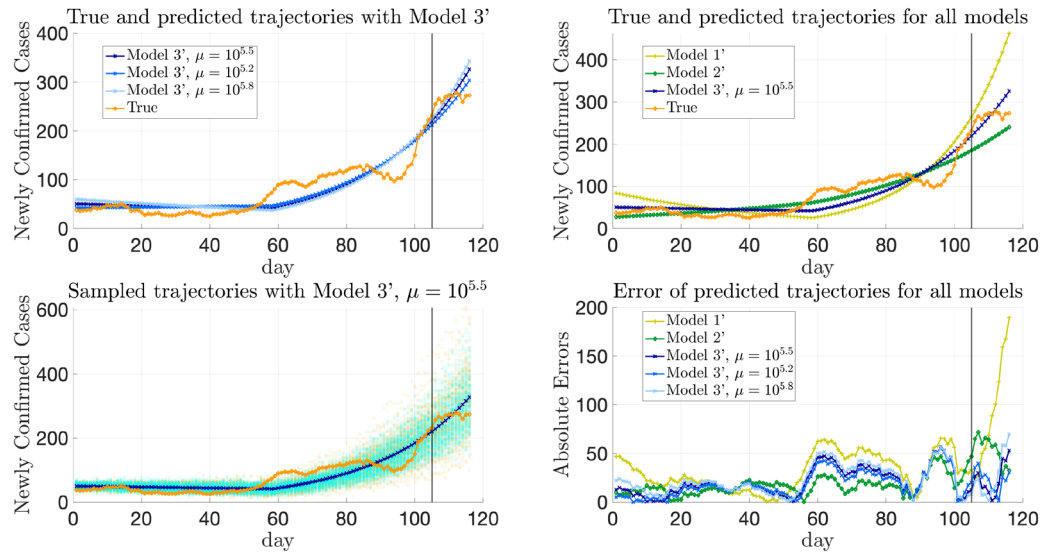


Figure 13. True and fitted trajectories in Austria. The remarks for the lines and scatter plots are the same as those in Fig. 8. The vertical lines show the threshold $T_{th} = 102$ of training-testing split of COVID-19 data in Europe. For Model 3, $\mu = 10^{5.2}, 10^{5.5},$ and $10^{5.8}$ are chosen. As shown in Fig. 17 and Supplementary Fig. S6, the validation errors are minimized at $\mu = 10^{5.5}$, and $\mu = 10^{5.2}$ and $10^{5.8}$ are obtained by perturbing the minimizer $\mu = 10^{5.5}$. $\mu = 10^{5.2}, 10^{5.5},$ and $10^{5.8}$ are marked by yellow pentagrams, blue squares, and green diamonds respectively in Fig. 17 and Supplementary Fig. S6.

True and predicted newly confirmed cases in Germany

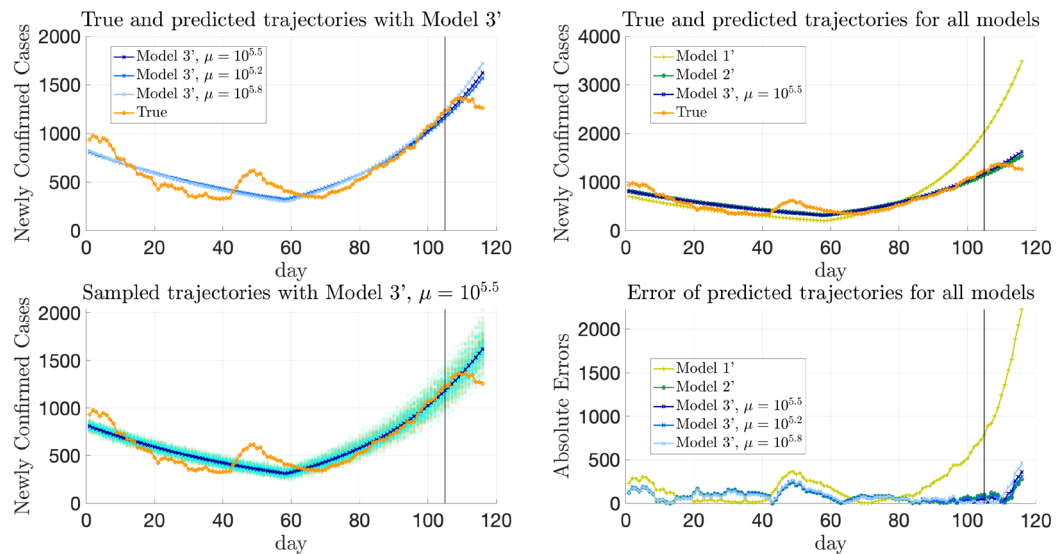


Figure 14. True and fitted trajectories in Germany. The remarks for the lines and scatter plots are the same as those in Fig. 8. The choice of $\mu = 10^{5.2}, 10^{5.5},$ and $10^{5.8}$ has been explained in Fig. 13.

Parameter inference of the three models. By the time-varying extension (4.3) detailed in “Model extension by allowing time-varying parameters”, for Model 3, $\bar{\Theta} = \{E_k(0), H_k(0), \lambda_k^{(1)}, \lambda_k^{(2)}\}_{k=1}^n$ are estimated by the optimization problem in (4.5):

True and predicted newly confirmed cases in Italy

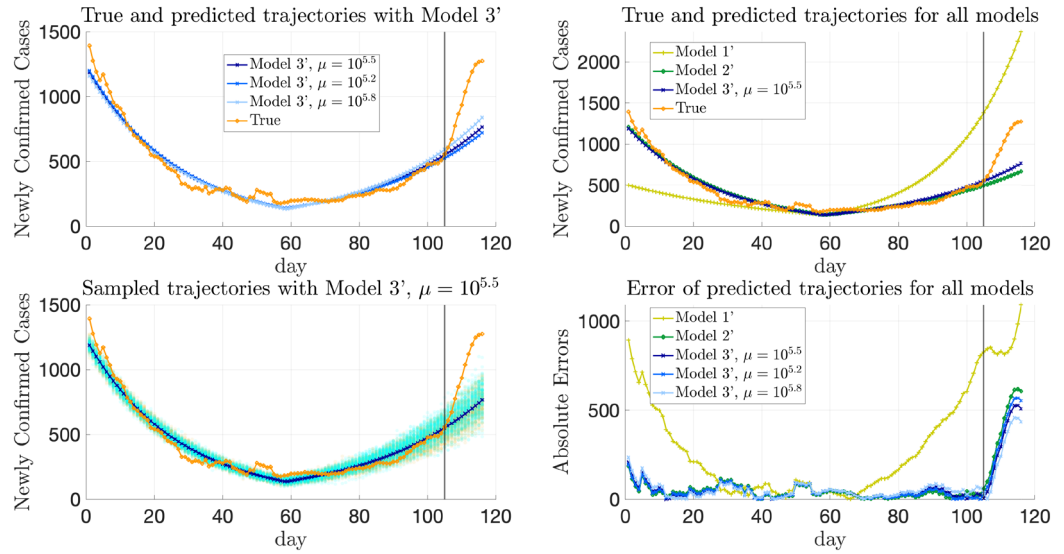


Figure 15. True and fitted trajectories in Italy. The remarks for the lines and scatter plots are the same as those in Fig. 8. The choice of $\mu = 10^{5.2}, 10^{5.5},$ and $10^{5.8}$ has been explained in Fig. 13.

Estimated posterior distribution for λ in Italy

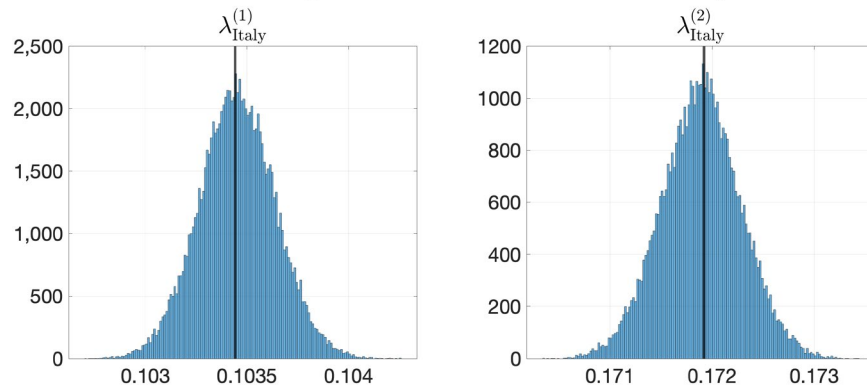


Figure 16. Estimated posterior distributions of λ in Italy. The vertical black lines represent the values of corresponding θ^* using Model 3'.

	θ^*	Mean of estimated posterior	Standard deviation of estimated posterior
$\lambda_{Italy}^{(1)}$	0.1034	0.1035	1.8768×10^{-4}
$\lambda_{Italy}^{(2)}$	0.1719	0.1719	3.8791×10^{-4}

Table 9. Estimated transmission rates in Italy. θ^* is inferred using (4.5) with $\mu = 10^{5.5}$, at which the validation errors are minimized as shown in Fig. 17 and Supplementary Fig. S6. The posterior distribution is estimated by 5×10^5 MCMC iterations.

$$\begin{aligned} \bar{\Theta}^* = \arg \min & \left(-\log P\left(\{(\Delta C_a)_k(i)\}_{k,i} \mid \bar{\Theta}\right) + \frac{\mu}{2} \left(\sum_m \sum_{i,j \in D_m} \left((\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \right) \right. \right. \\ & \left. \left. + \beta \sum_{m_1 < m_2} \sum_{i \in D_{m_1}, j \in D_{m_2}} \left((\lambda_i^{(1)} - \lambda_j^{(1)})^2 + (\lambda_i^{(2)} - \lambda_j^{(2)})^2 \right) \right) + \sigma \sum_i \left((\lambda_i^{(1)})^2 + (\lambda_i^{(2)})^2 \right) \right), \end{aligned} \tag{4.5}$$

Model	M	H	GL prior	MAE _(w) ^[Tr]	MAE _(w) ^[Te]	MSE _(w) ^[Tr]	MSE _(w) ^[Te]
1'	×	×	×	0.330	0.639	0.424	0.691
2'	×	✓	×	0.171	0.543	0.228	0.599
3' (P)	×	✓	✓ ($\mu = 10^{5.2}$)	0.203	0.450	0.268	0.489
	×	✓	✓ ($\mu^* = 10^{5.5}$)	0.214	0.447	0.285	0.488
	×	✓	✓ ($\mu = 10^{5.8}$)	0.225	0.441	0.302	0.486
3' (P')	×	✓	✓ ($\mu = 10^{5.2}$)	0.209	0.436	0.273	0.511
	×	✓	✓ ($\mu^* = 10^{5.5}$)	0.222	0.415	0.291	0.492
	×	✓	✓ ($\mu = 10^{5.8}$)	0.237	0.399	0.311	0.480

Table 10. Training and testing errors of Models 1'–3' for real-world data in Europe. The formulas of errors are detailed in Sect. B (Eq. (S5)) of Supplementary Information. The remarks for Columns 2–4 are the same as those in Table 4. In addition, the choice of $\mu = 10^{5.2}, 10^{5.5}$, and $10^{5.8}$ in Model 3' has been explained in Fig. 13. Note that the validation errors are minimized at $\mu^* = 10^{5.5}$ as shown in Fig. 17. The partitions in Model 3' are P and P' respectively. P and P' are two different partitions of countries in Europe introduced in “Models to compare”.

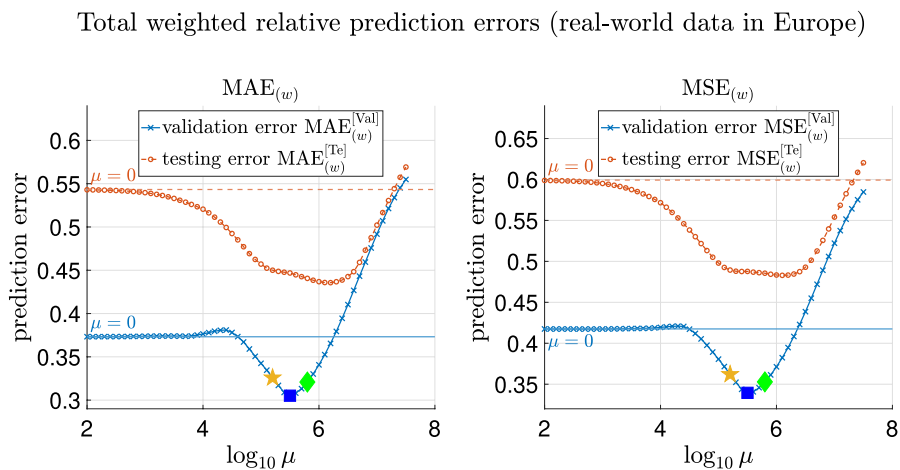


Figure 17. Testing and validation errors on real-world COVID-19 data in Europe. The two plots are similar to those in Fig. 7, with the errors computed on the real-world data in Europe instead of simulated data. Both the MAE and MSE validation errors are minimized at $\mu = 10^{5.5}$, which is marked by blue squares in both plots. $\mu = 10^{5.2}$ and $10^{5.8}$ are chosen by perturbing the minimizer $\mu = 10^{5.5}$, have slightly larger validation errors, and are marked with yellow pentagams and green diamonds in both plots. The construction of training/validation/testing data is detailed in Sects. A.1.2 and E of Supplementary Information.

or by (4.5) but with the partition $P = \{D_1, D_2, D_3, D_4\}$ replaced by the partition $P' = \{D'_1, D'_2, D'_3, D'_4\}$. Here, we still take $\sigma = 10^{-6}, \beta = 0.1$.

Results of trajectory prediction. We still remark that for Model 3' (with P), as in the case of real-world data in China, part of the results reported are from three choices of μ for careful consideration, since the validation error may not have the same distribution as the testing error. Note that Fig. 17 and Supplementary Fig. S6 plot the weighted and simply averaged relative validation errors, $MAE_{(w)}^{[Val]}, MSE_{(w)}^{[Val]}, MAE_{(s)}^{[Val]}$ and $MSE_{(s)}^{[Val]}$ respectively. One choice is $\mu^* = 10^{5.5}$, at which the validation errors (computed with P) achieve the minimum, marked with blue squares in Fig. 17 and Supplementary Fig. S6. The other two choices are $\mu = 10^{5.2}$ and $\mu = 10^{5.8}$, which have larger validation errors and are marked in orange pentagams and green diamonds respectively in Fig. 17 and Supplementary Fig. S6.

Figures 13, 14 and 15 present the true and predicted trajectories in Austria, Germany, and Italy, respectively. As can be seen from Figs. 13, 14 and 15, heterogeneity of transmission parameters in Model 2' (green lines) and Model 3' (blue lines) helps improve the performance of fitting and generalization. Furthermore, utilization of correlation between countries in Model 3' further improves the prediction of the trajectories as shown in Fig. 13. We also note that for Model 3', compared to the results for $\mu^* = 10^{5.5}$, at which the validation errors are minimized, $\mu = 10^{5.2}$ and $10^{5.8}$ achieve slightly more accurate prediction for Italy.

Results of parameter estimation. Same as before, Fig. 16 shows the estimated posterior distributions of $\lambda_{\text{Italy}}^{(1)}$ and $\lambda_{\text{Italy}}^{(2)}$ using Model 3', in which the vertical black lines show the inferred $\lambda_{\text{Italy}}^{(1)}$ and $\lambda_{\text{Italy}}^{(2)}$ with $\mu^* = 10^{5.5}$. Table 9 shows the inferred $\lambda_{\text{Italy}}^{(1)}$ and $\lambda_{\text{Italy}}^{(2)}$ and the mean and standard deviation of their estimated posterior distribution from MCMC. The estimated $\lambda_{\text{Italy}}^{(2)}$ is larger than the estimated $\lambda_{\text{Italy}}^{(1)}$, which is consistent with the trend that the newly confirmed cases in Italy first decrease and then increase.

Further model evaluation. Table 10 shows training and testing errors of the three models. It can be seen from Table 10 that Models 2' and 3' have smaller both training and testing errors than Model 1', due to the introduction of heterogeneity of transmission parameters. We can also see that compared with Model 2', Model 3' that adds Graph Laplacian prior has better generalization performance. Especially, Model 3' with $\mu = 10^{5.8}$ has smaller testing error than with $\mu = 10^{5.5}$, although $\mu = 10^{5.8}$ does not achieve the minimal validation errors.

Recall that P and P' are two different partitions of countries in Europe introduced in “Models to compare”, both based on geographical locations. By comparing the corresponding results in Table 10 for Model 3' with partitions P and P' respectively, it can be seen that the testing errors are close. This is different from the implications of the simulated data. A possible explanation for this might be that the regions in the real world have more complicated correlations than the simulated regions where clustering information is uniquely and artificially prefixed. Although reasonable groupings may not always be unique for real-world cases, the proposed model could still predict the trajectories more accurately than the baseline models.

Figure 17 below and Supplementary Fig. S6 plot the weighted and simply averaged testing errors (whose definitions are in Supplementary Information) against varying μ respectively, from which one may see that by imposing regularization properly through choosing a moderate μ helps improve the generalization performance. By comparing the blue solid lines and red dashed lines in Fig. 17 (and Supplementary Fig. S6), we see that the change in validation and testing errors concerning increasing μ are not the same. This further verifies the indications from the previous findings that it would be better to examine and compare the results from multiple choices of μ with relatively low validation errors for Model 3' to choose a better μ for the specific data sets.

Discussion

In this paper, we propose a stochastic dynamic model inspired by Ref.², with considerations on the inter-district transportation and as well as spatially and temporally heterogeneous transmission parameters, which can model the ongoing and lasting spread of the epidemic in multiple districts. Based on the proposed model, we also introduce a two-step procedure for estimating the parameters, which 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 proposed model improves the performance of fitting and generalization.

We acknowledge that there are limitations to 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 ascertainment rate of the positive cases, the asymptomatic virus carriers, and the infectious latent period. Second, the proposed model does not consider the heterogeneity with respect to transmission risk in different populations, for example, populations with different ages, jobs, or health conditions, and needs further modification to incorporate the large-scale COVID-19 vaccination as well as waning of immunity over time. Finally, the parameter inference depends on the correlation between regions determined by the clustering patterns among regions. However, such division is usually not fully known in the real-world cases, and the inference of graph structure is yet to be explored in this paper.

The current methods have possible extensions in the following several directions in our future study. First, the dynamic model could be refined to be closer to reality, for example, taking the asymptomatic virus carriers and contact tracking into account and modeling the change of the transmission parameters over time in a more sophisticated way. Second, the model can be modified to accommodate the heterogeneity of various populations, including different age groups and vaccination statuses. For example, the compartments are to be further divided into subgroups, and the parameters such as transmission rates and mortality rates are allowed to be different; and the dynamic model is also to be adjusted accordingly. Third, the method could further include detecting the underlying graph structure of the regions so that the construction of the graph Laplacian matrix could be more self-contained and systematic. At last, the method could be further used to assess the influence of containment measures taken by different countries, for example, by adjusting the traveling volume to make it different from the actual transportation data and then analyzing the corresponding influence on the size and trend of the pandemic.

Data availability

The data and code used in the simulation study (“Experimental results for simulated data”) and the application to COVID-19 data in Europe (“Results for COVID-19 data in Europe”) are publicly available in the repository https://github.com/Yixuan-Tan/Statistical_Inference_Using_GLEaM_with_Spatial_Heterogeneity_and_Correlation. The COVID-19 data in China (“Results for COVID-19 data in China”) are not publicly available due to that the official websites of health commissions of some provinces in China no longer maintain the COVID-19 data before March of 2020, but are available from the corresponding author on request.

Received: 21 December 2021; Accepted: 18 August 2022

Published online: 05 October 2022

References

- World Health Organization (WHO). <https://covid19.who.int>.
- Balcan, D. *et al.* Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *J. Comput. Sci.* **1**, 132–145 (2010).
- Keeling, M. J. & Rohani, P. *Modeling infectious diseases in humans and animals* (Princeton university press, 2011).
- Gomes, M. F. *et al.* Assessing the international spreading risk associated with the 2014 west african ebola outbreak. *PLoS Curr.* **6** (2014).
- 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 (Springer, 2016).
- Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395–400 (2020).
- 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 Curr.* **1**, RRN1129–RRN1129 (2009).
- Balcan, D. *et al.* Modeling the critical care demand and antibiotics resources needed during the fall 2009 wave of influenza a(h1n1) pandemic. *PLoS Curr.* **1**, RRN1133–RRN1133 (2009).
- Bajardi, P. *et al.* Modeling vaccination campaigns and the fall/winter 2009 activity of the new a(h1n1) influenza in the northern hemisphere. *Emerg. Health Threats J.* **2**, e11–e11 (2009;2008;2010;).
- Tizzoni, M. *et al.* Real-time numerical forecast of global epidemic spreading: Case study of 2009 a/h1n1pdm. *BMC Med.* **10**, 165 (2012).
- 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 (2014).
- 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 Med.* **7**, 1–12 (2009).
- Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track covid-19 in real time. *Lancet. Infect. Dis* **20**, 533–534 (2020).
- 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 (2017).
- Vespignani, A. Predicting the behavior of techno-social systems. *Science* **325**, 425–428 (2009).
- National Health Commission of the People's Republic of China. <http://en.nhc.gov.cn/antivirusfight.html>.
- Chinese Center for Disease Control and Prevention. <https://weekly.chinacdc.cn/news/TrackingtheEpidemic.htm>.
- European Centre for Disease Prevention and Control. <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.
- Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A* **115**, 700–721 (1927).
- Kendall, D. G. Deterministic and stochastic epidemics in closed populations. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Health*, 149–165 (University of California Press, 1956).
- Bailey, N. T. A simple stochastic epidemic. *Biometrika* **193**–202 (1950).
- Bartlett, M. Some evolutionary stochastic processes. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **11**, 211–229 (1949).
- Britton, T. *et al.* *Stochastic epidemic models with inference* (Springer, Berlin, 2019).
- Zhang, Y. *et al.* Prediction of the covid-19 outbreak in china based on a new stochastic dynamic model. *Sci. Rep.* **10**, 1–10 (2020).
- Donnat, C. & Holmes, S. Modeling the heterogeneity in covid-19's reproductive number and its impact on predictive scenarios. *J. Appl. Stat.* 1–29 (2021).
- Tkachenko, A. V. *et al.* Time-dependent heterogeneity leads to transient suppression of the covid-19 epidemic, not herd immunity. *Proc. Natl. Acad. Sci.* **118** (2021).
- Moreno, Y., Pastor-Satorras, R. & Vespignani, A. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B Condens. Matter Complex Syst.* **26**, 521–529 (2002).
- Boschi, T., Di Iorio, J., Testa, L., Cremona, M. A. & Chiaromonte, F. Functional data analysis characterizes the shapes of the first covid-19 epidemic wave in italy. *Sci. Rep.* **11**, 1–15 (2021).
- Carroll, C. *et al.* Time dynamics of covid-19. *Sci. Rep.* **10**, 1–14 (2020).
- Cremona, M. A. & Chiaromonte, F. Probabilistic *k*-mean with local alignment for clustering and motif discovery in functional data. *arXiv preprint arXiv:1808.04773* (2018).
- 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 Comput. Biol.* **16**, e1008031 (2020).
- Szapudi, I. Heterogeneity in sir epidemics modeling: superspreaders and herd immunity. *Appl. Netw. Sci.* **5**, 1–12 (2020).
- Volpert, V., Banerjee, M. & Sharma, S. Epidemic progression and vaccination in a heterogeneous population. application to the covid-19 epidemic. *Ecol. Complex.* 100940 (2021).
- Hou, X. *et al.* Intracounty modeling of covid-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proc. Natl. Acad. Sci.* **118** (2021).
- 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. *Sci. Rep.* **10**, 1–9 (2020).
- Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china. *Science* **368**, 638–642 (2020).
- Kraemer, M. U. *et al.* The effect of human mobility and control measures on the covid-19 epidemic in china. *Science* **368**, 493–497 (2020).
- World Health Organization (WHO). Enhancing response to omicron sars-cov-2 variant. [https://www.who.int/publications/m/item/enhancing-readiness-for-omicron-\(b.1.1.529\)-technical-brief-and-priority-actions-for-member-states](https://www.who.int/publications/m/item/enhancing-readiness-for-omicron-(b.1.1.529)-technical-brief-and-priority-actions-for-member-states).
- Kurtz, T. G. Solutions of ordinary differential equations as limits of pure jump markov processes. *J. Appl. Probab.* **7**, 49–58 (1970).
- Kurtz, T. G. Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *J. Appl. Probab.* **8**, 344–356 (1971).
- He, S., Peng, Y. & Sun, K. Seir modeling of the covid-19 and its dynamics. *Nonlinear Dyn.* **101**, 1667–1680 (2020).
- López, L. & Rodo, X. A modified seir model to predict the covid-19 outbreak in spain and italy: simulating control scenarios and multi-scale epidemics. *Res. Phys.* **21**, 103746 (2021).
- Yang, Z. *et al.* Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *J. Thorac. Dis.* **12**, 165 (2020).
- He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of covid-19. *Nat. Med.* **26**, 672–675 (2020).
- He, G. *et al.* When and how to adjust non-pharmacological interventions concurrent with booster vaccinations against Covid-19—Guangdong, China, 2022. *China CDC Week.* **4**, 199 (2022). <https://weekly.chinacdc.cn/article/id/397ce3f9-9388-46c1-862e-d6e8bee63a56>.
- Zhang, M. *et al.* Transmission dynamics of an outbreak of the covid-19 delta variant b.1.617.2 - guangdong province, china, may-june 2021. *China CDC Week.* **3**, 584 (2021). <https://weekly.chinacdc.cn/article/id/eb772589-1584-4ef9-beac-cac3ab2fbb12>.

47. Broyden, C. G. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA J. Appl. Math.* **6**, 76–90 (1970).
48. Fletcher, R. A new approach to variable metric algorithms. *Comput. J.* **13**, 317–322 (1970).
49. Goldfarb, D. A family of variable-metric methods derived by variational means. *Math. Comput.* **24**, 23–26 (1970).
50. Shanno, D. F. Conditioning of quasi-newton methods for function minimization. *Math. Comput.* **24**, 647–656 (1970).
51. National Bureau of Statistics of China. Annual data by province (2019). <http://www.stats.gov.cn/english/Statisticaldata/AnnuaIData/>.
52. Baidu Qianxi. <https://qianxi.baidu.com/#/2020chunyun>.
53. 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>.
54. Wikipedia. List of european countries by population. https://en.wikipedia.org/wiki/List_of_European_countries_by_population.
55. 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.
56. Blavatnik School of Government. Covid-19 government response tracker. <https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker>.

Acknowledgements

The authors thank Dr. Huaiyu Tian, Dr. Chong You, Omar Melikechi for helpful discussions, and Qiushi Lin and Feng Zhou for assistance with the collection of real-world data. Y. Z. and X.-H. Z. acknowledge support by National Natural Science Foundation of China grant 82041023 and The Bill & Melinda Gates Foundation (INV-016826). Y. T. and X. C. acknowledge support by NSF (DMS-2007040), and X.C. is also partially supported by the Alfred P. Sloan Foundation.

Author contributions

Y.T., Y.Z., and X.C. contributed to the development of the model and algorithm, as well as numerical experiments, data analysis and interpretation of results. All authors contributed to the study design and writing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-18775-8>.

Correspondence and requests for materials should be addressed to X.C. or X.-H.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022