



Epidemic Modeling of the Spatiotemporal Spread of COVID-19 over an Intercity Population Mobility Network

Yuxi Liu¹(✉), Shaowen Qin¹, and Zhenhao Zhang²

¹ College of Science and Engineering, Flinders University,
Tonsley, SA 5042, Australia

{liu1356, shaowen.qin}@flinders.edu.au

² College of Life Sciences, Northwest A&F University,
Yangling 712100, Shaanxi, China
zhangzhenhow@nwfu.edu.cn

Abstract. Intercity traveling has been recognized as a leading cause for the continuation of the COVID-19 global pandemic. However, there lacks credible prediction of the spatiotemporal spread of COVID-19 with humans traveling between metropolitan areas. This study attempts to establish a novel framework to simulate human traveling and the spread of virus across an intercity population mobility network. A Markov process was introduced to capture the stochastic nature of travelers' migration. A backward derivation algorithm was adopted and the Nelder-Mead simplex optimization method applied to overcome the limitation of existing deterministic epidemic models, including the difficulties in estimating the initial susceptible population and the optimal hyper-parameters required for simulation. We conducted two case studies with data from 24 cities in China and Italy. Our framework yielded state-of-the-art accuracy while being modular and scalable, indicating the addition of population mobility and stochasticity significantly improves prediction performance compared to using epidemic data alone. Moreover, our results revealed that transmission patterns of COVID-19 differ significantly with different population mobility, offering valuable information to the understanding of the correlation between traveling activities and COVID-19 transmission.

Keywords: Epidemic modeling · COVID-19 · Spatiotemporal analysis · Population mobility

1 Introduction

It is a common belief that the spread of COVID-19 across the world is through humans traveling over an open and interwoven network of population flow. Recent studies confirmed that human mobility is a major factor driving the spatial spread of COVID-19 [1–3]. Similarly, [4] revealed that mobility patterns

strongly correlated with COVID-19 growth rates, based on data from the most affected counties in the U.S. However, it is challenging to predict the extent and speed of virus transmission with credible accuracy. Further and deeper investigation on the spatiotemporal spreading patterns of COVID-19 utilizing a population mobility network beyond the traditional Susceptible-Exposed-Infectious-Removed (SEIR) model [5] needs to be undertaken. Such an investigation is not only essential for understanding the mechanism of the spread in a population mobility network involving intercity traveling, but also beneficial for devising measures for spread prevention and control on a global basis.

The art, also the science, of mathematical modeling, is to construct the simplest model that captures the salient features of the system. Virus transmission is a dynamic process that involves many stochastic components. Consequently, it is necessary to seamlessly incorporate population mobility and stochastic processes into the classical SEIR model to predict the spatiotemporal effect of virus transmission. Additionally, COVID-19 shows some variation from the typical progression of infectious disease transmission, in the sense that it is also infectious during the incubation period [6], which needs to be considered through modification of the classical SEIR model.

To our best knowledge, there does not exist any applicable mainstream network topology structure to simulate COVID-19 spread associated with population mobility. Also, some inadequacies in the deterministic epidemic models remain unaddressed. First, previous network models with clustering are context-specific hence not scalable [7, 8]. Consequently, results generated from such models may not be generalizable [9]. Second, we propose an initial high-risk population to accurately represent the size of the initial susceptible population, which is pivotal for epidemic modeling [10]. Prior studies suggested using the city population base as the initial susceptible population [6, 11]. However, such methods overestimate the infection cases at the early stage of epidemic transmission, which varies based on the demographic background of different cities. Thus, the infection rate is at risk of being underestimated as the size of infected populations being the product of the initial susceptible population and the infection rate. Third, the epidemic models usually are not parameter-free. Besides the essential initial susceptible population, initial parameters like the infection rate and recovery rate also need to be provided [12]. However, due to the lack of a comprehensive review of historical incidence data, it is hard to derive specific parameters for different survey sites.

In this paper, we use fine-grained spatiotemporal population mobility data, in conjunction with epidemic data, to construct an urban network epidemic framework, which is closer to real-world scenarios. The use of the framework offers a scalable and credible solution compared with the traditional SEIR model and strong baseline models such as Metapopulation SIR model [13] and Susceptible-Undiagnosed-Infected-Removed (SUIR) model [14]. Additionally, some challenges of the deterministic epidemic model are better addressed. Extensive experiments were conducted to assess the performance of our approach, using a real-world COVID-19 epidemic dataset of 12 cities in Hubei Province, China, and 12 cities in Italy.

2 The Proposed Approach

We propose an urban network epidemic framework (M-Urb-SEIR), a novel approach that incorporates population mobility and stochasticity for accurate COVID-19 confirmed case forecasting.

2.1 SEIR Model (Single-Network)

We adopt the SEIR model [5] to describe the dynamic process of epidemic propagation. Criteria for dividing the subjects are as follows: (P) represents the total population, Susceptible (S) is for the susceptible individuals, Exposed (E) for the exposed individuals, previously susceptible who have been exposed to the virus but may not be infected; Infected (I) for the infective individuals capable of transmitting the disease, this includes a non-symptomatic infectious period, and Recovered (R) for recovered individuals who were previously infected but have become immune. If the immune period is limited, R can be converted into S again. The relation between all variables is shown below:

$$\frac{dS}{dt} = -\alpha I(t)S(t)/N, \quad (1)$$

$$\frac{dE}{dt} = \alpha I(t)S(t)/N - \mu E(t), \quad (2)$$

$$\frac{dI}{dt} = \mu E(t) - \gamma I(t), \quad (3)$$

$$\frac{dR}{dt} = \gamma I(t), \quad (4)$$

$$S(t) + E(t) + I(t) + R(t) = N, \quad (5)$$

where α represents the rate of conversion for the exposed become infected; μ represents the rate of transformation for the incubation period to a patient; and γ represents the probability of recovery.

2.2 M-Urb-SEIR (Urban Network Epidemic Framework)

The traditional SEIR model assumes a single infection network among individuals and only model epidemic propagation in a single dimension. We extend the traditional SEIR model to the scenario of urban networks. We assume that there are N cities in a city network of interests. Eligibility criteria required individuals to be divided into four states: S, E, I, and R. To assess the city n , the city population base was used P_n , and the number of S, E, I, and R in the city at time t is $S_n(t)$, $E_n(t)$, $I_n(t)$, and $R_n(t)$. This study assumes a mobility intensity (W_{nm}) between city n and m , representing the average number of visitors from the city n to m . Recent evidence suggests that cases with the latent period are infectious [6]. We therefore set out to assess the effect of the infection rate of the

infected individual and the effect of infection rate of the latent individual (infectious and lag onset). α_1 represents the infection rate of the infected individual; α_2 represents the infection rate of the latent individual; μ represents the rate of transformation of the incubation period to patients; γ represents recovery rate of patients.

The transmission of and recovery from infection are intrinsically stochastic processes, and the deterministic epidemic model does not account for fluctuations [15]. To tackle this issue, we assume the process is Markovian on the relevant time scales, the dynamic variations between states of the four populations at t are summarized as follows:

$$\begin{aligned}
\frac{dS_n(t)}{dt} = & -\alpha_1 S_n(t) \sum_{m=1}^N \left(\frac{W_{mn}}{P_m} + \frac{W_{nm}}{P_n} \right) I_m(t) / P_n \\
& -\alpha_2 S_n(t) \sum_{m=1}^N \left(\frac{W_{mn}}{P_m} + \frac{W_{nm}}{P_n} \right) E_m(t) / P_n \\
& + \sqrt{\alpha_1 S_n(t) I_n(t) / P_n} \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \\
& + \sqrt{\alpha_2 S_n(t) E_n(t) / P_n} \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \\
& + \sqrt{\alpha_1 S_n(t) \sum_{m \neq n}^N \left(\frac{W_{nm}}{P_n} \right) I_m(t) / P_n} \\
& \quad \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \\
& + \left(\sqrt{\alpha_2 S_n(t) \sum_{m \neq n}^N \left(\frac{W_{nm}}{P_n} \right) E_m(t) / P_n} \right. \\
& \quad \left. \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \right) \\
& + \sqrt{\sum_{m \neq n}^N \alpha_1 S_n(t) \left(\frac{W_{nm}}{P_n} \right) I_m(t) / P_n} \\
& \quad \cdot P_t(S_n \xrightarrow{\alpha_1} E_n).
\end{aligned} \tag{6}$$

$$\begin{aligned}
\frac{dE_n(t)}{dt} = & \alpha_1 S_n(t) \sum_{m=1}^N \left(\frac{W_{mn}}{P_m} + \frac{W_{nm}}{P_n} \right) I_m(t) / P_n \\
& + \alpha_2 S_n(t) \sum_{m=1}^N \left(\frac{W_{mn}}{P_m} + \frac{W_{nm}}{P_n} \right) E_m(t) / P_n \\
& - \mu \cdot E_n(t) - \left(\sqrt{\alpha_1 S_n(t) I_n(t) / P_n} \right. \\
& \quad \left. \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \right) \\
& - \left(\sqrt{\alpha_2 S_n(t) E_n(t) / P_n} \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \right) \\
& - \left(\sqrt{\alpha_1 S_n(t) \sum_{m \neq n}^N \left(\frac{W_{nm}}{P_n} \right) I_m(t) / P_n} \right. \\
& \quad \left. \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \right) \\
& - \left(\sqrt{\alpha_2 S_n(t) \sum_{m \neq n}^N \left(\frac{W_{nm}}{P_n} \right) E_m(t) / P_n} \right. \\
& \quad \left. \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \right) \\
& - \left(\sqrt{\sum_{m \neq n}^N \alpha_1 S_n(t) \left(\frac{W_{nm}}{P_n} \right) I_m(t) / P_n} \right. \\
& \quad \left. \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \right) \\
& - \left(\sqrt{\sum_{m \neq n}^N \alpha_2 S_n(t) \left(\frac{W_{nm}}{P_n} \right) E_m(t) / P_n} \right. \\
& \quad \left. \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \right) \\
& + \left(\sqrt{\mu E_n(t)} \cdot P_t(E_n \xrightarrow{\mu} I_n) \right).
\end{aligned} \tag{7}$$

$$\begin{aligned}
\frac{dI_n(t)}{dt} = & \mu \cdot E_n(t) - \gamma \cdot I_n(t) \\
& \left(\sqrt{\mu E_n(t)} \cdot P_t(E_n \xrightarrow{\mu} I_n) \right) \\
& + \sqrt{\gamma I_n(t)} \cdot P_t(I_n \xrightarrow{\gamma} R_n).
\end{aligned} \tag{8}$$

$$\frac{dR_n(t)}{dt} = \gamma \cdot I_n(t) - \sqrt{\gamma I_n(t)} \cdot P_t(I_n \xrightarrow{\gamma} R_n). \quad (9)$$

$P_t(S_n \xrightarrow{\alpha_1} E_n)$, the probability that individuals in S state (city n) will be transformed into E state (city n) at t time, which is caused by the contact between S (city n) and I (city n). $P_t(S_n \xrightarrow{\alpha_2} E_n)$, the probability that individuals in S state (city n) will be transformed into E state (city n) at t time, which is caused by the contact between S (city n) and E (city n). $P_t(S_n \xrightarrow{\alpha_1} E_m)$, the probability that the individuals in S state (city n) will be transformed into E state (city m) at t time due to the contact between individuals in S (city n) and I (city m). $P_t(S_n \xrightarrow{\alpha_2} E_m)$, the probability that the individuals in S state (city n) will be transformed into E state (city m) at t time due to the contact between individuals in S (city n) and E (city m). $P_t(E_n \xrightarrow{\mu} I_n)$, the probability that individuals in E state (city n) transforms into I state (city n) at t time. $P_t(I_n \xrightarrow{\gamma} R_n)$, the probability of individuals in I state (city n) transforms into R state (city n) at time t.

In an urban network, there are three behaviors for susceptible populations in urban n to access the incubation period.

(1) Internal transmission of city n:

$$\begin{aligned} & \alpha_1 \cdot S_n(t) \cdot I_n(t)/P_n + \alpha_2 \cdot S_n(t) \cdot E_n(t)/P_n \\ & - \sqrt{\alpha_1 S_n(t) I_n(t)/P_n} \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \\ & - \sqrt{\alpha_2 S_n(t) E_n(t)/P_n} \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \end{aligned} \quad (10)$$

(2) Transmission caused by the flow of infected and exposed individuals from city m to n:

$$\begin{aligned} & \alpha_1 \cdot S_n(t) \cdot \sum_{m \neq n}^N \left(\frac{W_{mn}}{P_m}\right) \cdot I_m(t)/P_n \\ & + \alpha_2 \cdot S_n(t) \cdot \sum_{m \neq n}^N \left(\frac{W_{mn}}{P_m}\right) \cdot E_m(t)/P_n \\ & - \sqrt{\alpha_1 S_n(t) \sum_{m \neq n}^N \left(\frac{W_{nm}}{P_n}\right) I_m(t)/P_n} \cdot P_t(S_n \xrightarrow{\alpha_1} E_n) \\ & - \sqrt{\alpha_2 S_n(t) \sum_{m \neq n}^N \left(\frac{W_{nm}}{P_m}\right) E_m(t)/P_n} \cdot P_t(S_n \xrightarrow{\alpha_2} E_n) \end{aligned} \quad (11)$$

(3) The susceptible population from city n flows into m and infected:

$$\begin{aligned} & \sum_{m \neq n}^N \alpha_1 \cdot S_n(t) \left(\frac{W_{nm}}{P_n}\right) \cdot I_m(t)/P_n \\ & + \sum_{m \neq n}^N \alpha_2 \cdot S_n(t) \left(\frac{W_{nm}}{P_n}\right) \cdot E_m(t)/P_n \\ & - \sqrt{\sum_{m \neq n}^N \alpha_1 S_n(t) \left(\frac{W_{nm}}{P_n}\right) I_m(t)/P_n} \cdot P_t^2(S_n \xrightarrow{\alpha_1} E_m) \\ & - \sqrt{\sum_{m \neq n}^N \alpha_2 S_n(t) \left(\frac{W_{nm}}{P_n}\right) E_m(t)/P_n} \cdot P_t^2(S_n \xrightarrow{\alpha_2} E_m) \end{aligned} \quad (12)$$

The proposed framework is implemented by an overall algorithm as follows:

Algorithm 1: M-Urb-SEIR

Require: \mathcal{R}_0 , (α', γ') , $(\alpha_1, \alpha_2, \mu, \gamma)$, W , P , cumulative confirmed $\hat{C}(t)$, T **Ensure:** $S(t)$, $E(t)$, $I(t)$, $R(t)$

- 1: **Initialize:** $\alpha', \gamma', \alpha_1, \alpha_2, \mu, \gamma$
 - 2: **Backward derivation \mathcal{E}_0 :**
 - 3: Apply $\mathcal{R}_0, \alpha', \gamma', P$ on Eq. (3.13), then use the Nelder-Mead simplex optimization method to obtain the optimal \mathcal{E}_0 from $I(T) + R(T) = \mathcal{E}_0$
 - 4: Apply $(\alpha_1, \alpha_2, \mu, \gamma)$, W , P on Markov stochastic process, and infer the $P_t (S_n \xrightarrow{\alpha_1} E_n)$, $P_t (S_n \xrightarrow{\alpha_2} E_n)$, $P_t (S_n \xrightarrow{\alpha_1} E_m)$, $P_t (S_n \xrightarrow{\alpha_2} E_m)$, $P_t (E_n \xrightarrow{\mu} I_n)$, $P_t (I_n \xrightarrow{\gamma} R_n)$ (see Table 3)
 - 5: **Estimation:**
 - 6: Apply $(\alpha_1, \alpha_2, \mu, \gamma)$, W , P , $\mathcal{E}_0(S(0) = \mathcal{E}_0)$ and $P_t(\cdot)$ on Eq. (3.6-3.9), obtain $I(t)$, $R(t)$, and $C(t) = I(t) + R(t)$
 - 7: Optimize $(\alpha_1, \alpha_2, \mu, \gamma)$ by using the Nelder-Mead simplex optimization method
 - 8: **Simulation:**
 - 9: Obtain the relative error from $C(t)$ and $\hat{C}(t)$
 - 10: **for** $t = 1$ to T **do**
 - 11: Apply $(\alpha_1, \alpha_2, \mu, \gamma)$ on Eq. (3.6-3.9)
 - 12: update $S(t)$, $E(t)$, $I(t)$, $R(t)$
 - 13: **end for**
-

2.3 Addressing the Challenges of a Deterministic Epidemic Model

(1) Our proposed framework is scalable. Once the original and target domain are located and marked, the actual cross-domain propagation of COVID-19 can be simulated. (2) We propose a backward derivation algorithm to derive the initial susceptible population \mathcal{E}_0 . Specifically, we first used the way of [16] to obtain \mathcal{R}_0 sequences from time-series data of confirmed cases. We then established a basic Susceptible-Infected-Recovered (SIR) model [11] as shown in Eq. (3.13).

$$\mathcal{R}_0 = \frac{\alpha \cdot P}{\gamma}, \quad (13)$$

where α represents the infection rate, P represents the total population in the area, and γ represents the recovery rate. Based on Eq. (3.13), we first initialized the infection rate (α) and recovery rate (γ). We then adopted the Nelder-Mead simplex optimization method to optimize α and γ to make the total number of infected individuals (I) and recovered individuals (R) as close as possible to the real number of confirmed cases. Last, the total number of infected individuals (I) and recovered individuals (R) were used as \mathcal{E}_0 of the urban network epidemic framework. Moreover, the optimal α and γ were used as the initial hyper-parameters of the urban network epidemic framework. Note that we used \mathcal{E}_0 instead of the city population base in the urban network epidemic framework, and the Markov process used the difference between the city population base and \mathcal{E}_0 to incorporate stochasticity. (3) We used the Nelder-Mead simplex optimization method again to optimize the α , γ , and μ (the rate of transformation of the incubation period to patients) of the urban network epidemic framework.

3 Experimental Settings

We present datasets, competitors, and evaluation metrics for our experiments.

3.1 Datasets

We adopted the epidemic data from the National Health Commission of the People's Republic of China¹ (daily confirmed new cases for each city between January 23, 2020 and February 29, 2020) and Italian epidemic data² (daily confirmed new cases for each city between February 24, 2020 and April 15, 2020). The statistical data includes the cumulative number of infected, recovered, and death cases. Chinese migration data were obtained with consent from Baidu migration, and the most recent data are available on the website (<https://qianxi.baidu.com>). The dataset includes the immigration scale index, the emigration scale index, and intracity travel intensity. The migration scale index is converted according to the absolute value of the number of individuals who move in/out, reflecting the population scale of the cities. The intensity of intracity travel is the exponential result of the number of individuals who have traveled in the city and the city's resident population, reflecting the intracity mobility scale. Similarly, Italian migration data were obtained from [17].

3.2 Competitors

To fairly compare different approaches, we compare our approach with the following deterministic epidemic models.

1. SEIR is an epidemiological model used to simulate the spread of infectious disease.
2. Metapopulation SIR model [13] (SIGKDD 2018) extends the SIR model to a metapopulation SIR model that allows visitors transmission between any two sub-populations.
3. SUIR model [14] (SIGKDD 2020) incorporates a unique 'undiagnosed' state of the COVID-19 on the basis of the SIR model.
4. Urb-SEIR (without Markov process) is one variant of the proposed framework.

3.3 Evaluation Metrics

This study assumes that the prediction date between t and T , and the relative error is defined as:

$$e(t, T) = \frac{|C(t+T) - \hat{C}(t+T)|}{\hat{C}(t+T)}, \quad (14)$$

where $C(t)$ represents the cumulative confirmed cases obtained from the baselines and our proposed methods, and $\hat{C}(t)$ represents the cumulative truth cases based on the database.

¹ <https://github.com/CSSEGISandData/COVID-19>.

² <https://github.com/pcm-dpc/COVID-19>.

4 Experimental Results

Most research works use the number of city population base as the initial susceptible population, however, Table 1 shows the initial high-risk population base deduced by the backward derivation algorithm is much less than city population but more reasonable. Furthermore, the optimal hyper-parameter identified in these responses are summarized in Table 2. Table 3 shows the inference formulation by Markov stochastic process.

Figures 1, 2, 3 and 4 illustrate the prediction error bars of models in Chinese and Italian cities. The graphs show that our M-Urb-SEIR performs well in the 28 days forecast period of 12 cities in China compared with baseline models. Moreover, we find that the M-Urb-SEIR outperforms the Urb-SEIR. This benefits from incorporating random effects into epidemic models, which is critical for improving prediction accuracy. Besides, we had three critical findings from the experimental results in Italian cities. First, the Metapopulation SIR model prediction result performs the best on most days of the first week, and the sub-optimal model is the traditional SEIR model. Second, by contrast, Urb-SEIR and M-Urb-SEIR perform better with a longer prediction horizon. Usually, they will perform better than other approaches when the prediction horizon is longer than 1 or 2 weeks. Third, the performance of the Urb-SEIR is better than M-Urb-SEIR. Compared to China, where ‘Chunyun’ leads to dramatic population migration, Italy’s strength of population movement is much milder. Therefore, M-Urb-SEIR, which considers more about the stochastic effect, performs worse than Urb-SEIR. These figures suggest that the prediction of COVID-19 should be customized, and contextual information should be considered. In different application scenarios, the model should be able to be extended and modularized. Specifically, for the high randomness effect such as the ‘Chunyun’ event (the largest periodic human migration in China), models should take the Markov

Table 1. Number of the initial high-risk population obtained from backward derivation algorithm

Site	Ezhou	Enshi	Huanggang	Huangshi	Jingzhou	Shiyan
City Population	1059.7k	3390k	6333k	2471.7k	5570.1k	3398k
High-Risk Population	663	180	460250	682	7474	538
Site	Suizhou	Wuhan	Xiantao	Xiangyang	Xiaogan	Yichang
City Population	2221k	11212k	1140.1k	5680k	4921k	4137.9k
High-Risk Population	919	304800	687	582	4355	1770
Site	Roma	Milano	Brescia	Torino	Monza	Bologna
City Population	4342.2k	3293.7k	1266k	2259.7k	830.5k	1014.6k
High-Risk Population	17513	55678	129944	10214	62505	35390
Site	Firenze	Catania	Verona	Bergamo	Trieste	Napoli
City Population	1011.3k	1107.7k	926.5k	1114.6k	234.5k	3084.9k
High-Risk Population	4117	29981	19017	21485	1443	2865

Table 2. Optimal hyper-parameters of outbreak cities in China and Italy

Site	Ezhou	Enshi	Huanggang	Huangshi	Jingzhou	Shiyan
α_1	0.0971	0.0801	0.0965	0.0958	0.0967	0.0897
α_2	0.0764	0.0615	0.0751	0.0756	0.0637	0.0699
μ	0.0833	0.2882	0.1667	0.1053	0.4082	0.1942
γ	0.4261	0.3012	0.3661	0.4052	0.4753	0.3521
Site	Suizhou	Wuhan	Xiantao	Xiangyang	Xiaogan	Yichang
α_1	0.0828	0.1534	0.0459	0.0963	0.0732	0.0735
α_2	0.0512	0.0785	0.0403	0.0855	0.0614	0.0632
μ	0.1333	0.1429	0.2857	0.101	0.295	0.3846
γ	0.3253	0.3656	0.3081	0.4233	0.3031	0.4146
Site	Roma	Milano	Brescia	Torino	Monza	Bologna
α_1	0.1525	0.1585	0.1371	0.1734	0.1513	0.1539
α_2	0.1317	0.1457	0.1164	0.1611	0.1261	0.1372
μ	0.1316	0.1887	0.1811	0.0813	0.1942	0.1361
γ	0.4131	0.4373	0.4513	0.4461	0.4333	0.4717
Site	Firenze	Catania	Verona	Bergamo	Trieste	Napoli
α_1	0.1683	0.1746	0.1768	0.1476	0.1853	0.1736
α_2	0.1505	0.1681	0.1367	0.1346	0.1641	0.1623
μ	0.1357	0.3142	0.1392	0.2665	0.1047	0.0899
γ	0.4338	0.4162	0.4743	0.4526	0.4297	0.4529

Table 3. Inference formulation by Markov stochastic process

Site	S_n	$E_{n\alpha_1}$	$E_{n\alpha_2}$	I_n	R_n
S_m	$\frac{W_{mn}}{P_m}$	$\frac{W_{mn}}{P_m} \alpha_1$	$\frac{W_{mn}}{P_m} \alpha_2$	$\frac{W_{mn}}{P_m} (\alpha_1 + \alpha_2 \mu)$	$\frac{W_{mn}}{P_m} (\alpha_1 + \alpha_2) \mu \gamma$
$E_{m\alpha_1}$	0	$\frac{W_{mn}}{P_m}$	0	$\frac{W_{mn}}{P_m} \mu$	$\frac{W_{mn}}{P_m} \mu \gamma$
$E_{m\alpha_2}$	0	0	$\frac{W_{mn}}{P_m}$	$\frac{W_{mn}}{P_m} \mu$	$\frac{W_{mn}}{P_m} \mu \gamma$
I_m	0	0	0	$\frac{W_{mn}}{P_m}$	$\frac{W_{mn}}{P_m} \gamma$
R_m	0	0	0	0	$\frac{W_{mn}}{P_m}$

stochastic process into account; however, in the context of regular population movements (Italian cities), the results highlighted the need to use the Urb-SEIR as a predictive tool.

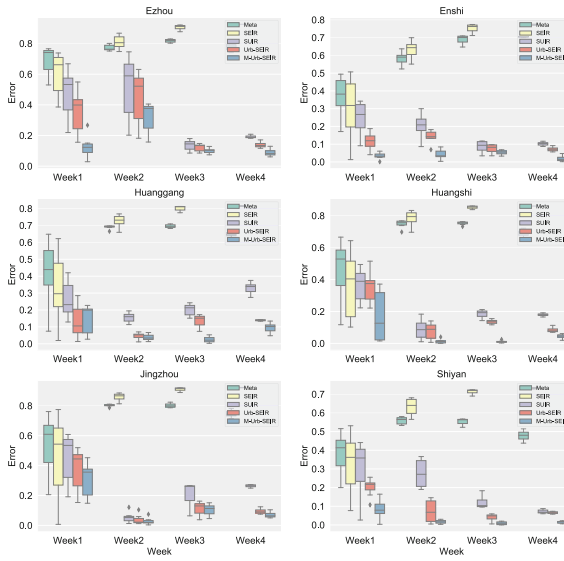


Fig. 1. Prediction error bars of models after T days.

With regard to the research methods, some limitations need to be acknowledged. The principal limitation of this analysis was the variance to estimate R_0 [18]. Another major source of uncertainty is in the backward derivation algorithm used to calculate the initial susceptible population. The latent infectivity population and other external factors were not accounted for in the derivation process. Although there are limitations in the backward derivation, it contributed to the infinite approach to the real world’s transmission state. An additional uncontrolled factor is the effect of ‘Chunyun’ [19], which is hard to be measured in the prediction process. Furthermore, the summary of error results is subject to inevitable fluctuation. The fluctuation phenomenon is intrinsically one of the challenges of the deterministic epidemic model, which reflects the likelihood that the precision of a deterministic epidemic model will vary across the ‘life cycle’ of an epidemic outbreak when analyzed using a set of fixed parameters. This will clearly influence the results across a long forecast period; Therefore, we provided a 28 days prediction horizon, which is much longer than most prior studies. Future studies that adopt a ‘stage’ forecast in the life cycle of an epidemic are clearly indicated. Despite its limitations, this study indicates the effectiveness of incorporating population mobility and random effects into the epidemic simulation.

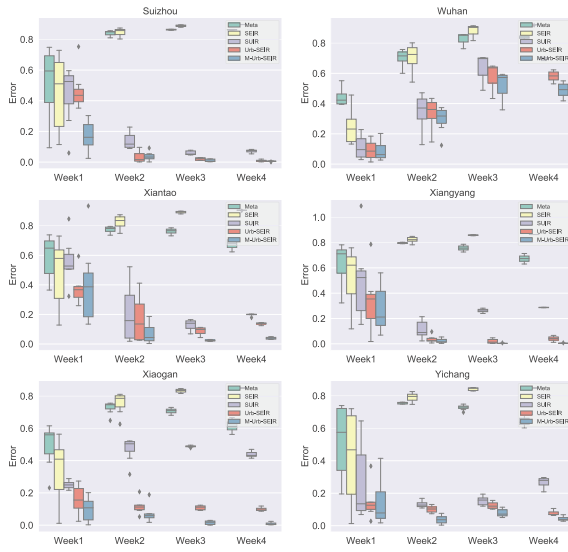


Fig. 2. Prediction error bars of models after T days.

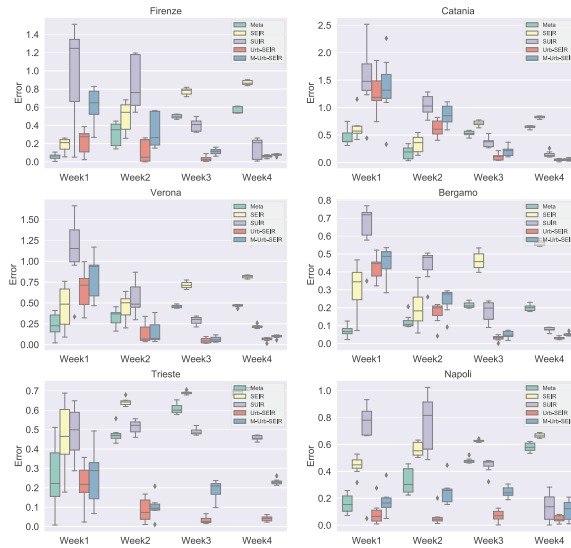


Fig. 3. Prediction error bars of models after T days.

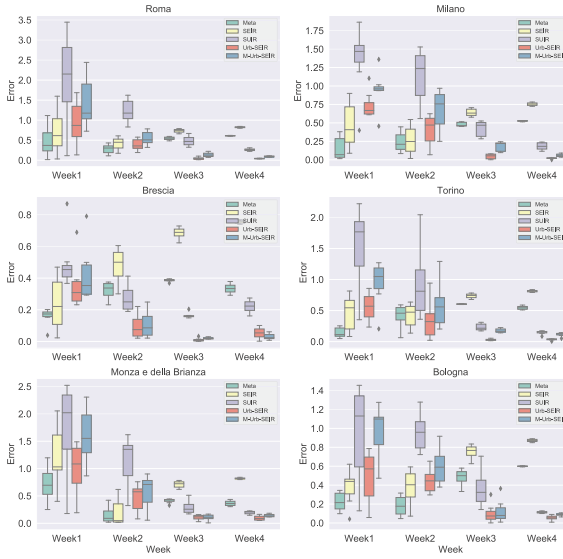


Fig. 4. Prediction error bars of models after T days.

5 Conclusion

In this paper, we propose a novel urban network epidemic framework to study the spread pattern of COVID-19 in different cities. We applied the framework to simulate and predict the COVID-19 confirmed cases in ‘epicenter’ Wuhan and other 11 cities in Hubei Province of China and 12 cities in Italy with severe epidemic situations, which outperforms other deterministic state-of-the-art models in the COVID-19 spreading prediction task. We also demonstrated that incorporating population mobility and random effect into epidemic models is necessary. Our findings provide new scientific evidence for further epidemic model design and offer a foundation for conducting other research studies, such as assessing the long-term social and economic effects of COVID-19.

References

1. Zhong, P., Guo, S., Chen, T.: Correlation between travellers departing from Wuhan before the spring festival and subsequent spread of Covid-19 to all provinces in China. *J. Travel Med.* **27**(3), taaa036 (2020)
2. Tian, H., et al.: An investigation of transmission control measures during the first 50 days of the Covid-19 epidemic in china. *Science* **368**(6491), 638–642 (2020)
3. Du, Z., et al.: Risk for transportation of coronavirus disease from Wuhan to other cities in China. *Emerg. Infect. Diseases.* **26**(5), 1049 (2020)
4. Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M., Gardner, L.M.: Association between mobility patterns and Covid-19 transmission in the USA: a mathematical modelling study. *Lancet Infect. Disea.* **20**(11), 1247–1254 (2020)

5. Cooke, K.L., Van Den Driessche, P.: Analysis of an SEIRS epidemic model with two delays. *J. Math. Biol.* **35**(2), 240–260 (1996)
6. Li, R., et al.: Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-COV-2). *Science* **368**(6490), 489–493 (2020)
7. Ball, F., Britton, T., Sirl, D.: A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon. *J. Math. Biol.* **66**(4), 979–1019 (2013)
8. Maki, Y., Hirose, H.: Infectious disease spread analysis using stochastic differential equations for sir model. In: 2013 4th International Conference on Intelligent Systems, Modelling and Simulation, pp. 152–156. IEEE (2013)
9. Pellis, L., et al.: Eight challenges for network epidemic models. *Epidemics* **10**, 58–62 (2015)
10. O’Dea, E.B., Pepin, K.M., Lopman, B.A., Wilke, C.O.: Fitting outbreak models to data from many small norovirus outbreaks. *Epidemics* **6**, 18–29 (2014)
11. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond. Ser. A (Containing Papers of a Mathematical and Physical Character)* **115**(772), 700–721 (1927)
12. Wang, N., Fu, Y., Zhang, H., Shi, H.: An evaluation of mathematical models for the outbreak of Covid-19. *Precis. Clin. Med.* **3**(2), 85–93 (2020)
13. Wang, J., Wang, X., Wu, J.: Inferring metapopulation propagation network for intra-city epidemic control and prevention. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 830–838 (2018)
14. Wang, J., Lin, X., Liu, Y., Feng, K., Lin, H., et al.: A knowledge transfer model for Covid-19 predicting and non-pharmaceutical intervention simulation. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2020)
15. Hufnagel, L., Brockmann, D., Geisel, T.: Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci.* **101**(42), 15124–15129 (2004)
16. Cintrón-Arias, A., Castillo-Chávez, C., Betencourt, L., Lloyd, A.L., Banks, H.T.: The estimation of the effective reproductive number from disease outbreak data. Technical report, North Carolina State University, Center for Research in Scientific Computation (2008)
17. Pepe, E., et al.: Covid-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Sci. Data* **7**(1), 1–7 (2020)
18. Adam, D.: A guide to R - the pandemic’s misunderstood metric. *Nature* **583**(7816), 346–349 (2020)
19. Jia, J.S., Lu, X., Yuan, Y., Xu, G., Jia, J., Christakis, N.A.: Population flow drives spatio-temporal distribution of Covid-19 in china. *Nature* **582**(7812), 389–394 (2020)