



Modelling the epidemic dynamics of COVID-19 with consideration of human mobility

Bowen Du¹ · Zirong Zhao¹ · Jiejie Zhao¹ · Le Yu¹ · Leilei Sun¹  · Weifeng Lv¹

Received: 31 January 2021 / Accepted: 3 June 2021 / Published online: 24 June 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

Abstract

So far COVID-19 has resulted in mass deaths and huge economic losses across the world. Various measures such as quarantine and social distancing have been taken to prevent the spread of this disease. These prevention measures have changed the transmission dynamics of COVID-19 and introduced new challenges for epidemic modelling and prediction. In this paper, we study a novel disease spreading model with two important aspects. First, the proposed model takes the quarantine effect of confirmed cases on transmission dynamics into account, which can better resemble the real-world scenario. Second, our model incorporates two types of human mobility, where the intra-region human mobility is related to the internal transmission speed of the disease in the focal area and the inter-region human mobility reflects the scale of external infectious sources to a focal area. With the proposed model, we use the human mobility data from 24 cities in China and 8 states in the USA to analyse the disease spreading patterns. The results show that our model could well fit/predict the reported cases in both countries. The predictions and findings shed light on how to effectively control COVID-19 by managing human mobility behaviours.

Keywords COVID-19 · Epidemic dynamics · Human mobility · Social distancing

1 Introduction

COVID-19 has spread throughout the world since it was first reported in Wuhan, China [6]. By the end of August 2020, COVID-19 had become a pandemic causing more than 20 million infections and leading to massive economic loss [18]. To decelerate the spreading speed of COVID-19, interventions such as the declaration of national emergency and control of human movement have been adopted in many countries [4]. However, it is not immediately clear to what extent the interventions could help to prevent the spread of the disease [13,17]. In this paper, we investigate the effects of quarantine efforts and human mobility behaviours in modelling the infection scale and managing the epidemic dynamics of the disease.

An earlier yet qualitative study was conducted to utilize the spatial human transmission to predict the subsequent spread of disease [21]. Since the outbreak of COVID-19, great atten-

tion has been paid to the epidemic modelling and prediction for prevention. [8] analysed the temporal dynamics of the COVID-19 outbreak in China, Italy and France. [22] applied SIR model for epidemic simulation and estimated parameters by minimizing the negative log-likelihood function. [2] combined machine learning with the traditional SEIR model to predict the development of epidemic. Due to the large population movement during the outbreak, the modelling of COVID-19 has become complicated. Some scholars tried to consider the impact of human mobility on the epidemic [5,12,15,19,23,27]. Specifically, [27] utilized the exported infected cases between cities and the human mobility data to forecast the extent of the domestic and global risks of epidemic. [15] and [23] both investigated the effect of human mobility on the COVID-19 and analysed how the travel restriction measures mitigate the spread of COVID-19. [19] utilized telecommunication data in Switzerland to describe the human mobility patterns and construct regression models to predict the growth of COVID-19 cases. [12] proved that compared to pure synthetic model, simulated models based on pedestrian simulators make more contributions to the solution of COVID-19. However, the above research only considered inter-region mobility patterns, which can-

✉ Leilei Sun
leileisun@buaa.edu.cn

¹ State Key Laboratory of Software Development Environment, Beihang University, Haidian District, Beijing 100191, China

not fully capture the effects of human mobility on disease transmission. [5] focused on a variety of intervention strategies adopted since the outbreak of epidemic, which affected human activity to a certain extent and in turn had an effect on the spread of epidemic. Moreover, some studies considered the presence of asymptomatic, or infectious but not detected in COVID-19 [1,4,9,20]. However, in these studies, conventional epidemic prediction models (e.g. SIR [25] and SEIR [3]) were directly applied to estimate the spread of COVID-19. They did not take quarantine into account or simply treated the confirmed cases as the infected cases (i.e. the I in SIR or SEIR models).

To cope with the above challenges, this paper provides a novel method to model the epidemic dynamics of COVID-19 by taking quarantine and human mobility into account. Different from the previous models, we divide the population into three states: susceptible, infected and confirmed, where “confirmed” is a newly added state that coincides with the quarantine fact because once a person is diagnosed with COVID-19, the patient will be quarantined. To thoroughly model the impacts of human mobility on the spread of COVID-19, we consider inter-region and intra-region human mobility, which correspond to imported infectious cases and internal transmission speed of a focal area, respectively.

In summary, we highlight our contributions as follows:

- A comprehensive empirical study is conducted on the intra-region and inter-region mobility patterns during the epidemic. Actually, most of the prevention measures could be reflected by the changes of human mobility, and the transmission of COVID-19 largely depends on human mobility.
- Taking into account characteristics of the epidemic isolation, a “confirmed” status is firstly introduced into epidemic dynamic models, which can better resemble the real-world scenario.
- Two types of human mobility are incorporated with the model to study the epidemic dynamics of COVID-19. Particularly, the intra-region human mobility, which indirectly affects the epidemic spreading speed under a series of preventive measures, is important for the COVID-19 modelling, but was ignored in most of research.
- The proposed model is evaluated on the real-world data of China and the USA, and experimental results demonstrate the effectiveness and robustness of model.

The rest of the paper is organized as follows: an empirical study to reflect the relationship between the epidemic and human mobility patterns under a series of preventive measures is introduced in Sect. 2; the methodology used to model the epidemic dynamics of COVID-19 by taking quarantine and human mobility into account is introduced in Sect. 3; comprehensive experiments are conducted in Sects. 4 and 5

to verify the effectiveness of our proposed model; Sect. 6 gives the conclusion of our study.

2 Empirical study

In real-world scenario, the spread of an epidemic largely depends on human mobility, where the immigrants from epidemic area could bring in infectious sources and the social contact in a focal area is closely related to the spread speed [7,10,11]. The COVID-19 broke out during the 2020’s Chinese Spring Festival. Unlike the SARS outbreak in 2003, the scale of human mobility in China today is about six times that of 17 years ago [26]. Governments around the world have issued a series of preventive measures to alleviate the spread of COVID-19 [28]. Most of the measures target at cutting the spreading route by reducing human mobility. Here, we take China as an example to study the impacts of two types of human mobility on the spread of COVID-19.

The human mobility and epidemic data were collected from Baidu Qianxi¹ and Chinese Center for Disease Control and Prevention² (China CDC) respectively. The preventive measures were crawled from the State Council of the People’s Republic of China³ (SCPRC). Till to February 28, 2020, there are 24 cities totally whose confirmed cases were more than 200.

In this section, to observe human mobility changes, we select eight representative cities from them, including three cities with the most confirmed cases in Hubei province (Wuhan, Xiaogan and Huanggang), a city with the most confirmed cases outside Hubei province (Wenzhou), and four cities with the most frequent population migrations (Beijing, Shanghai, Guangzhou and Shenzhen).

2.1 Human mobility analysis during epidemic

The evolution of intra-city and inter-city human mobilities in the 8 selected cities during epidemic is shown in Fig. 1. From Fig. 1a, it can be observed that the scales of intra-city human mobility began to decrease in most of the cities in China. The reason is that people began to leave big cities to their hometowns in the last week before Chinese Spring Festival (from January 18 to 24). However, the variation of human mobility in Xiaogan and Huanggang is a bit distinct from others with a short-term increment. As is reported in China News, two-thirds of people who left from Wuhan before the closure on January 23 flowed to other cities in Hubei province, including Xiaogan and Huanggang, which promoted the intra-city human mobility. Moreover, as for Xiaogan and Huanggang,

¹ <http://qianxi.baidu.com/>.

² <http://www.chinacdc.cn/>.

³ <http://www.gov.cn/zhengce/>.

most of immigrants were from Hubei province, see Fig. 1b, which may be the reason for the epidemic breakout in them in the following several weeks.

From January 23 to February 9, local governments all over the country posed a series of measures to restrict human movements during the Spring Festival holiday [14], such as encouraging home rest, reducing mass gathering activities, cancelling or postponing large-scale public activities, closing schools and shutting down shuttle buses. The Chinese government also extended the Spring Festival of Hubei province to March 10 and the holiday of other provinces to February 9. As a result, both the intra-city and inter-city human mobility were extremely minimized.

2.2 Influence of immigrants from Hubei province

From Fig. 1b, it could be observed that the immigrants from Hubei province to the studied cities had significant peaks and then dropped sharply due to the preventive measures. To further investigate the correlations between the number of newly confirmed cases in each city and the corresponding immigrants from Hubei province, we compute the PCC (Pearson Correlation Coefficients) between these two factors, see Fig. 2. The horizontal axis represents the offset days between newly confirmed cases and immigrants from Hubei per day. For example, let us suppose that the period of newly confirmed cases ranges from February 1 to February 14, and 0 means the timestamps of newly confirmed cases and immigrants from Hubei per day are aligned (i.e. February 1 to February 14 vs. February 1 to February 14), -1 means the timestamps of immigrants are moved forward for 1 day (i.e. February 1 to February 14 vs. January 31 to February 13), and 1 means the timestamps of immigrants are delayed for 1 day (i.e. February 1 to February 14 vs. February 2 to February 15). Figure 2 illustrates that the largest PCC values of most cities (except for Wuhan) are distributed around offsets between -15 and -13 , which reflects the fact that the incubation period of COVID-19 is about 14 days. In particular, Huanggang, Xiaogan and Shenzhen, as the cities having more frequent interactions with Wuhan than other cities, have much higher PCC values than that of other cities. Moreover, for Wuhan, the correlation between immigrants from other cities in Hubei and the spread of the disease is not obvious since Wuhan is the birthplace of COVID-19 in China and the impact of immigrants to the disease is relatively limited.

2.3 Summary of observations

Based on the aforementioned analysis, the key observations could be summarized as follows.

Firstly, the preventive measures could effectively affect human mobility and thus help to slow down the spread of

epidemic, which could be known from the number of newly confirmed cases in China recently.

Secondly, with the implementation of preventive measures, both intra-city and inter-city human mobility have been dramatically changed, which makes the transmission dynamics of COVID-19 complex, which reveals that the conventional epidemic modelling methods without consideration of human mobility could not well fit/predict the real observed cases.

3 Methodology

In this section, we first briefly introduce two classical epidemic modelling methods (i.e. SIR and SEIR model) and explain why they are not suitable for directly modelling the COVID-19. Then we introduce the proposed Susceptible-Infectious-2-Confirmed (SI2C) model in detail.

3.1 Classical epidemic modelling method

SIR as a classic warehouse model was proposed by HH Weiss in 2013 [25], which is applicable to the types of epidemics that have no exposed period and that people will not be infected again after recovery. As Eq. (1) shows, in the SIR model, S represents susceptible status, I represents infectious status, R represents removed status (i.e. dead or recovered), both β and α represent the transition probability between three states, and p denotes the estimated number of people that an individual could contact physically.

Even though the SIR model is scientific in itself, it is unreasonable to directly regard the number of confirmed cases as I in the modelling of COVID-19. For this epidemic, patients will be quarantined as soon as they are confirmed, so the confirmed cases are not contagious and cannot play the role of I that can infect others. In fact, as for COVID-19, the role of confirmed cases is more similar to R . In order to avoid ambiguity, we refine I in the SIR as infected but not confirmed status and R as confirmed status C , thus a model named SIC obtained. It turns out that SIC model is more consistent with the actual situation.

$$\begin{cases} S_t &= S_{t-1} - \frac{p\beta\mathcal{I}_{t-1}S_{t-1}}{N} \\ \mathcal{I}_t &= \mathcal{I}_{t-1} + \frac{p\beta\mathcal{I}_{t-1}S_{t-1}}{N} - \alpha\mathcal{I}_{t-1} \\ \mathcal{R}_t &= \mathcal{R}_{t-1} + \alpha\mathcal{I}_{t-1} \\ N &= S_t + \mathcal{I}_t + \mathcal{R}_t \end{cases} \quad (1)$$

As shown in Eq. (2), SEIR model extends E on SIR model representing the exposed status, which is suitable for epidemics with exposed period (the exposed infections are not contagious). β , γ and α represent the transition probability between four states. As a derivative model of SIR, the SEIR

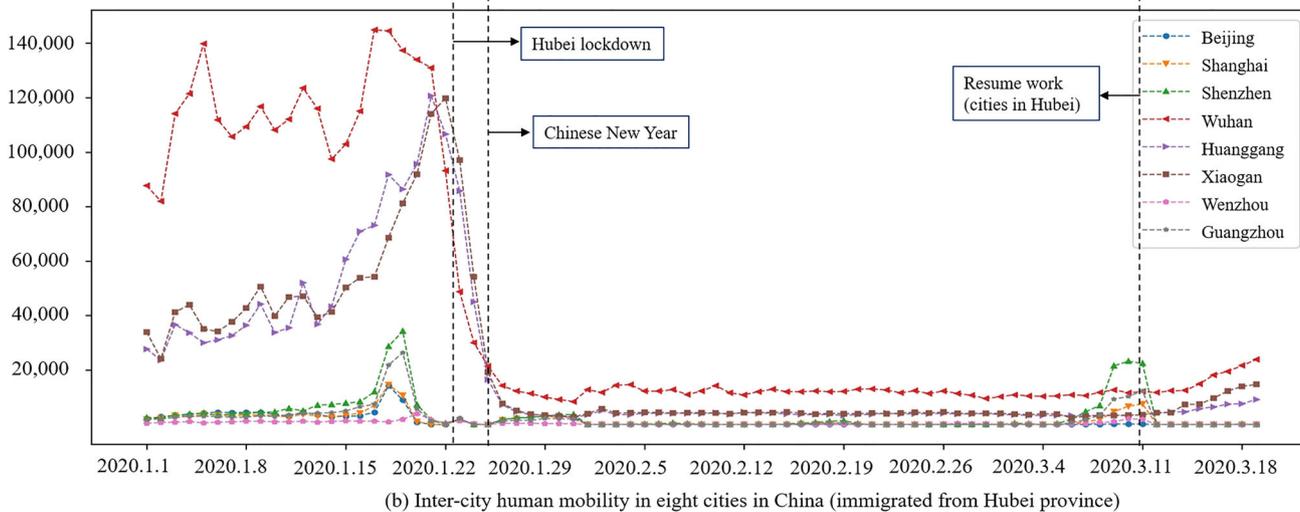
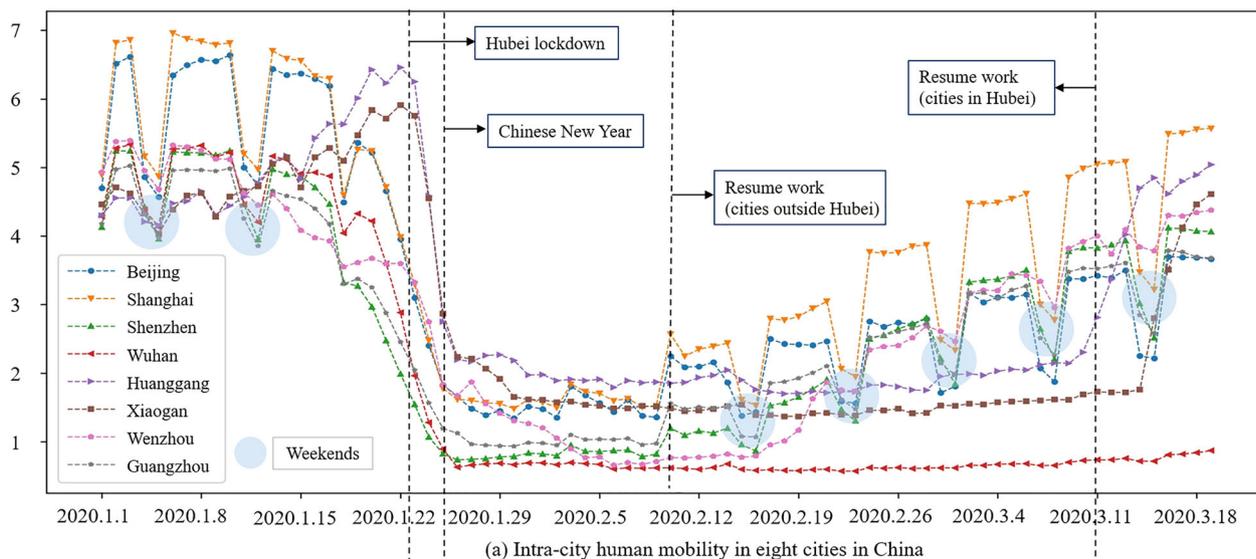


Fig. 1 The intra-city and inter-city human mobility during epidemic in 8 cities in China

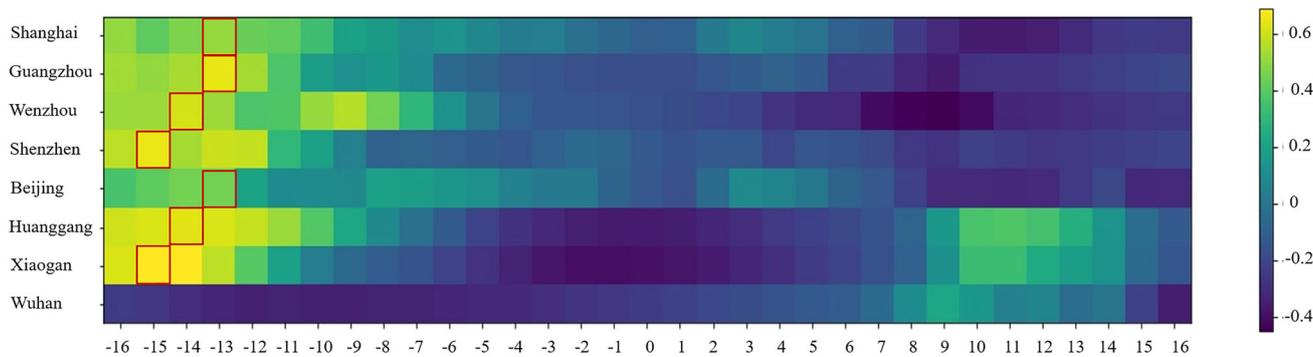


Fig. 2 Pearson correlation coefficients between the number of newly confirmed cases and the corresponding immigrants from Hubei province per day

Fig. 3 The architecture of the proposed SI2C model

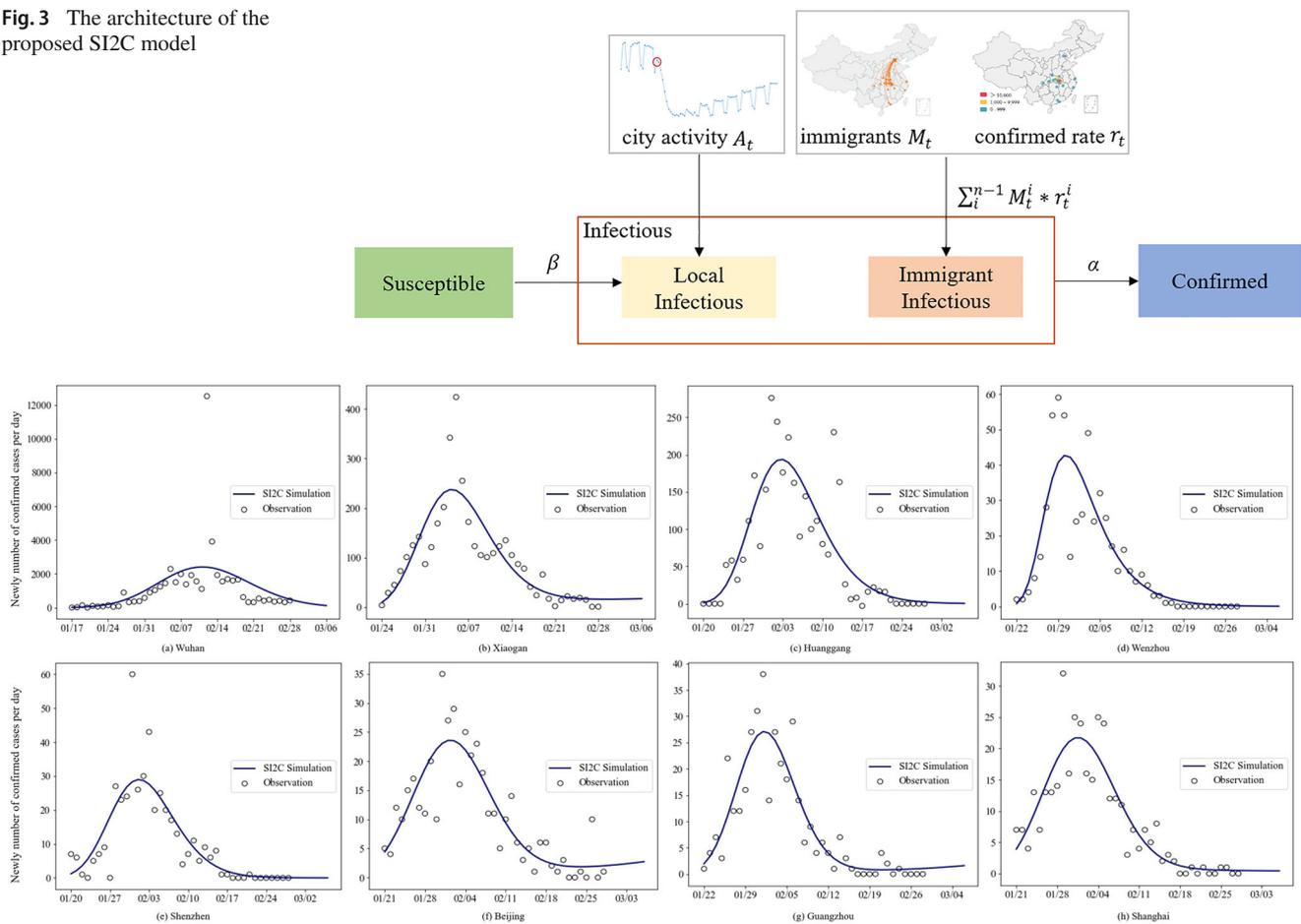


Fig. 4 Simulations of newly number of confirmed cases per day in cities of China on SI2C model

model has the same irrationality as SIR for COVID-19 (i.e. the number of confirmed cases cannot be taken as I). In addition, for this epidemic, there are a large number of studies prefer SEIR rather than SIR. They believe that COVID-19 has an incubation period and the patients in incubation period is infectious, so they reformed E on the SEIR model. Actually, those methods misinterpret the meaning of E in the original SEIR model whose E can only represent exposed status without infectious, because once E is considered to be infectious, it is equivalent to I .

$$\begin{cases}
 S_t &= S_{t-1} - \frac{p\beta I_{t-1} S_{t-1}}{N} \\
 E_t &= E_{t-1} + \frac{p\beta I_{t-1} S_{t-1}}{N} - \gamma E_{t-1} \\
 I_t &= I_{t-1} + \gamma E_{t-1} - \alpha I_{t-1} \\
 R_t &= R_{t-1} + \alpha I_{t-1} \\
 N &= S_t + E_t + I_t + R_t
 \end{cases} \quad (2)$$

3.2 The overview of SI2C model

Both intra-region and inter-region human mobilities have been dramatically changed with the intervention measures, which makes the transmission dynamics of COVID-19 complex. The traditional methods without consideration of human mobility could not well perform for epidemic modelling. Therefore, a novel epidemic spreading model SI2C is proposed. SI2C model is based on the dynamic partial differential equations of epidemic in mathematics. Taking the patient quarantine into account, we distinguish the infectious abilities of the infected and the confirmed individuals by introducing a “confirmed” status into the epidemic spreading model to better predict the spread of epidemic. Moreover, we leverage the intra-region and inter-region human mobility patterns into the proposed model to further improve the prediction performance. The architecture of SI2C model is shown in Fig. 3, and the corresponding dynamical equations

Table 1 The performance of 4 SICs models on China dataset

Chinese cities	MAPE				PCC			
	S12C	SIC@inter	SIC@intra	SIC	S12C	SIC@inter	SIC@intra	SIC
Wuhan	0.200	2.088	0.200	2.088	0.9914	0.9707	0.9914	0.9707
Xiaogan	0.054	0.244	0.051	0.695	0.9988	0.9911	0.9987	0.9753
Huanggang	0.119	1.131	0.108	2.637	0.9991	0.9850	0.9991	0.9674
Jingzhou	0.098	0.380	0.105	1.079	0.9973	0.9886	0.9973	0.9719
Ezhou	0.346	0.328	0.346	1.729	0.9966	0.9869	0.9966	0.9747
Suizhou	0.055	0.141	0.055	0.371	0.9993	0.9934	0.9993	0.9770
Xiangyang	0.023	0.063	0.022	0.241	0.9997	0.9969	0.9997	0.9829
Huangshi	0.019	0.119	0.021	0.311	0.9997	0.9929	0.9996	0.9758
Yichang	0.143	0.462	0.155	1.918	0.9988	0.9937	0.9988	0.9771
Jingmen	0.331	0.345	0.352	1.586	0.9961	0.9892	0.9961	0.9750
Xianning	0.085	0.104	0.085	0.148	0.9856	0.9816	0.9856	0.9727
Shiyan	0.037	0.060	0.055	0.403	0.9998	0.9980	0.9998	0.9872
Chongqing	0.050	0.052	0.049	0.241	0.9997	0.9972	0.9997	0.9884
Xiantao	0.051	0.243	0.050	0.619	0.9997	0.9914	0.9996	0.9755
Wenzhou	0.060	0.418	0.060	1.05	0.9989	0.9911	0.9989	0.9708
Tianmen	0.133	0.123	0.133	0.403	0.9914	0.9789	0.9914	0.9655
Shenzhen	0.115	0.355	0.115	0.697	0.9982	0.9814	0.9982	0.9591
Beijing	0.027	0.095	0.032	0.299	0.9996	0.9946	0.9994	0.9811
Guangzhou	0.035	0.235	0.039	0.666	0.9996	0.9899	0.9995	0.9702
Shanghai	0.033	0.118	0.036	0.304	0.9996	0.9926	0.9996	0.9775
Xinyang	0.053	0.114	0.050	0.286	0.9991	0.9749	0.9991	0.9780
Enshi	0.036	0.048	0.036	0.116	0.9969	0.9945	0.9969	0.9866
Changsha	0.051	0.199	0.065	0.519	0.9995	0.9699	0.9995	0.9699
Nanchang	0.081	0.198	0.102	0.656	0.9992	0.9957	0.9990	0.9801

are shown in Eq. (3). Each equation in Eq. (3) is explained in detail in the next two subsections.

$$\begin{cases}
 p_t &= k_{\text{intra}} * A_t, \\
 S_t &= S_{t-1} - \frac{p_t \beta \mathcal{I}_{t-1} S_{t-1}}{N}, \\
 \mathcal{I}_t^{\text{loc}} &= \mathcal{I}_{t-1} + \frac{p_t \beta \mathcal{I}_{t-1} S_{t-1}}{N} - \alpha \mathcal{I}_{t-1}, \\
 \mathcal{I}_t^{\text{im}} &= k_{\text{inter}} * \sum_i^{n-1} M_t^i * r_t^i, \\
 \mathcal{I}_t &= \mathcal{I}_t^{\text{loc}} + \mathcal{I}_t^{\text{im}}, \\
 C_t &= C_{t-1} + \alpha \mathcal{I}_{t-1},
 \end{cases} \tag{3}$$

3.3 Local infectious dynamics

The spreading speed of an epidemic in a local region (e.g. a city) is strongly related with the active level of physical contacts in the community. Higher intra-region active level means a spreader has more chances to get in touch with susceptible people and then could infect more patients. Therefore, governments advocate the social distancing to reduce the frequency and extend of human movements. Accordingly, intra-region human mobility, as an indicator

of the active level of physical contacts, is useful to estimate how many people a spreader could contact physically within a local region. Specifically, we estimate the number of people that an individual could contact physically during time t as:

$$p_t = k_{\text{intra}} * A_t, \tag{4}$$

where A_t is the intra-region human mobility at time t and k_{intra} is a trainable parameter. Then we estimate the number of infected persons inside a local region during time t as:

$$S_t = S_{t-1} - \frac{p_t \beta \mathcal{I}_{t-1} S_{t-1}}{N}, \tag{5}$$

$$\mathcal{I}_t^{\text{loc}} = \mathcal{I}_{t-1} + \frac{p_t \beta \mathcal{I}_{t-1} S_{t-1}}{N} - \alpha \mathcal{I}_{t-1}, \tag{6}$$

where S_t and $\mathcal{I}_t^{\text{loc}}$ denote the number of susceptible individuals and infected persons, respectively. β represents the infection rate of an individual after contacting with an infected people, α is the confirmed rate of a disease, and N denotes the overall population of the focal region.

Table 2 Significant differences between the performances of 4 SICs models on China dataset

Metrics	S12C versus SIC@inter	S12C versus SIC@intra	SI2C versus SIC	SIC@intra versus SIC@inter	SIC@inter versus SIC	SIC@intra versus SIC
MAPE (<)	$p = 2.51e-03$	$p = 3.66e-02$	$p = 3.33e-05$	$p = 3.40e-03$	$p = 3.07e-05$	$p = 3.60e-05$
PCC (>)	$p = 3.53e-06$	$p = 1.72e-02$	$p = 2.49e-13$	$p = 3.84e-06$	$p = 2.71e-10$	$p = 2.58e-13$

3.4 Dispersal among different regions

Based on the observations from empirical study, we also investigate on the effect of the immigration among different regions on the propagation of epidemic. An assumption is that the actual influence of the population from a source region to a focal region is not only dependent on the amount but also related to the infected risk of the source region, which is calculated by,

$$\mathcal{I}_t^{im} = k_{inter} * \sum_i^{n-1} M_t^i * r_t^i, \tag{7}$$

where k_{inter} is a trainable parameter that describes the influence of inter-region immigration and n is the number of studied regions. \mathcal{I}_t^{im} is the estimated number of inter-region infected persons during time t , M_t^i represents the number of immigrants from source region i to the focal region during time t , and r_t^i denotes the confirmed rate of region i during time t .

Finally, we fuse the influences of intra-region and inter-region infectious sources together by,

$$\mathcal{I}_t = \mathcal{I}_t^{loc} + \mathcal{I}_t^{im}. \tag{8}$$

The number of confirmed cases is calculated by,

$$\mathcal{C}_t = \mathcal{C}_{t-1} + \alpha \mathcal{I}_{t-1}, \tag{9}$$

where \mathcal{C}_t denotes the number of confirmed individuals in the focal region during time t .

4 Experiments on China dataset

In this section, we show that the confirmed status and the two human mobility patterns in the proposed SI2C model contributes significantly to the performance of the model to estimate the spread of the disease. We conduct experiments on 24 cities in China to evaluate the performance of the proposed SI2C and related SICs models, including SIC@intra (solely considers the intra-city mobility), SIC@inter (solely considers the inter-city mobility) and SIC (without considering either intra-city or inter-city mobility). We have made the code publicly available on GitHub platform ⁴

4.1 Dataset description

The collected data include epidemic and human mobility data of 24 cities in China from January 16 to February

⁴ <https://github.com/waterybye/Modelling-the-Epidemic-Dynamics-of-COVID-19-With-Consideration-of-Human-Mobility>.

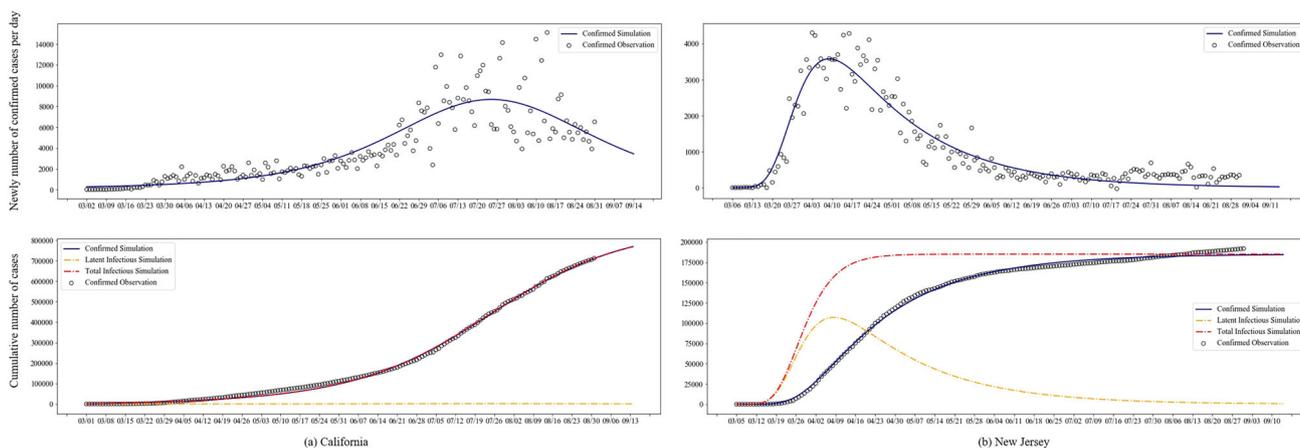


Fig. 5 Simulations of infections in California and New Jersey on SIC@intra model

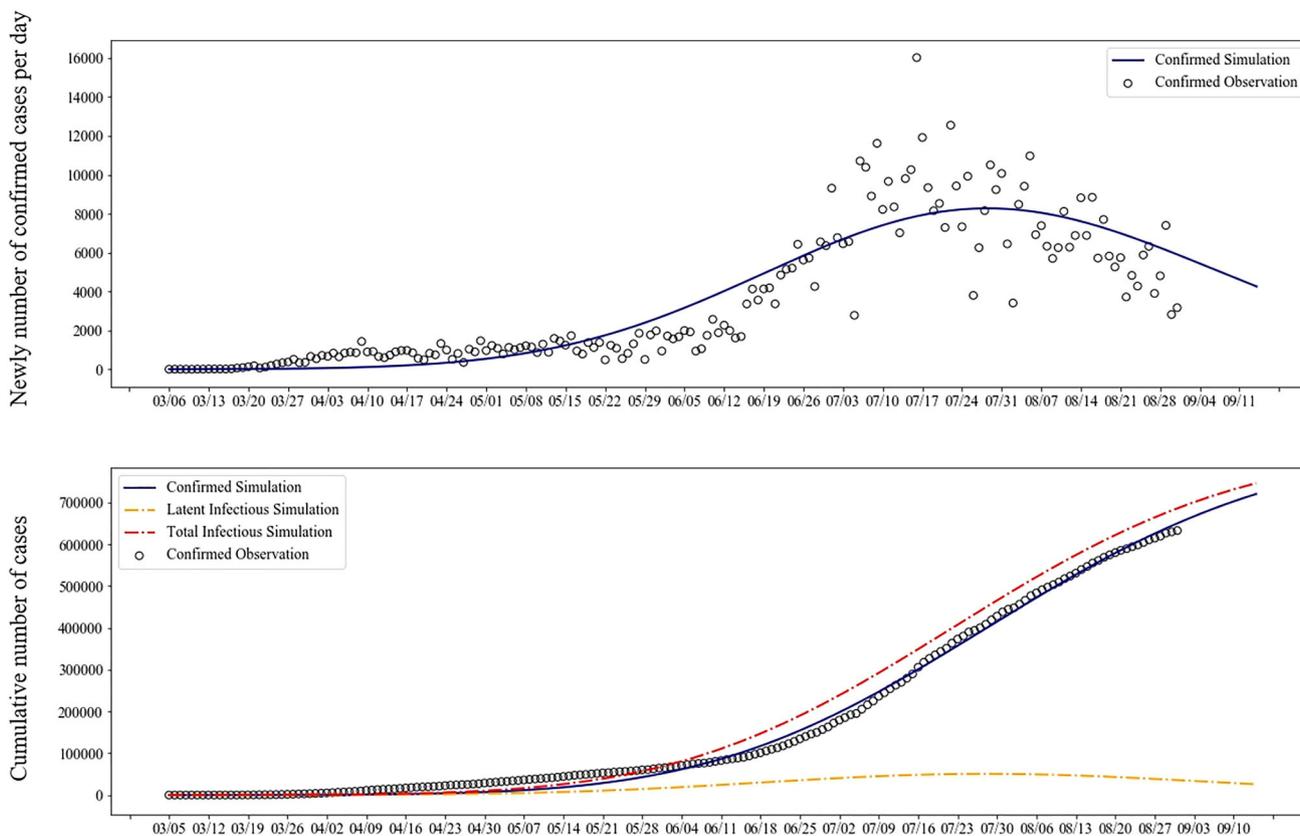


Fig. 6 Simulations of infections in Texas on SIC@intra model

28, 2020. The epidemic data record the cumulative confirmed cases, and the human mobility data provide daily travel intensity indices inside each city and daily migration scale indices between cities to reflect intra-city and inter-city human mobility respectively. More details of China dataset could be found in the beginning of Sect. 2. Since the epidemic period of China is relatively short and the epidemic development has stabilized, experimental results are directly

observed and compared on the fitting data.

4.2 Experimental setup

The subsection introduces the evaluation metrics and implementation details of experiments, which are the same on China dataset and the US dataset.

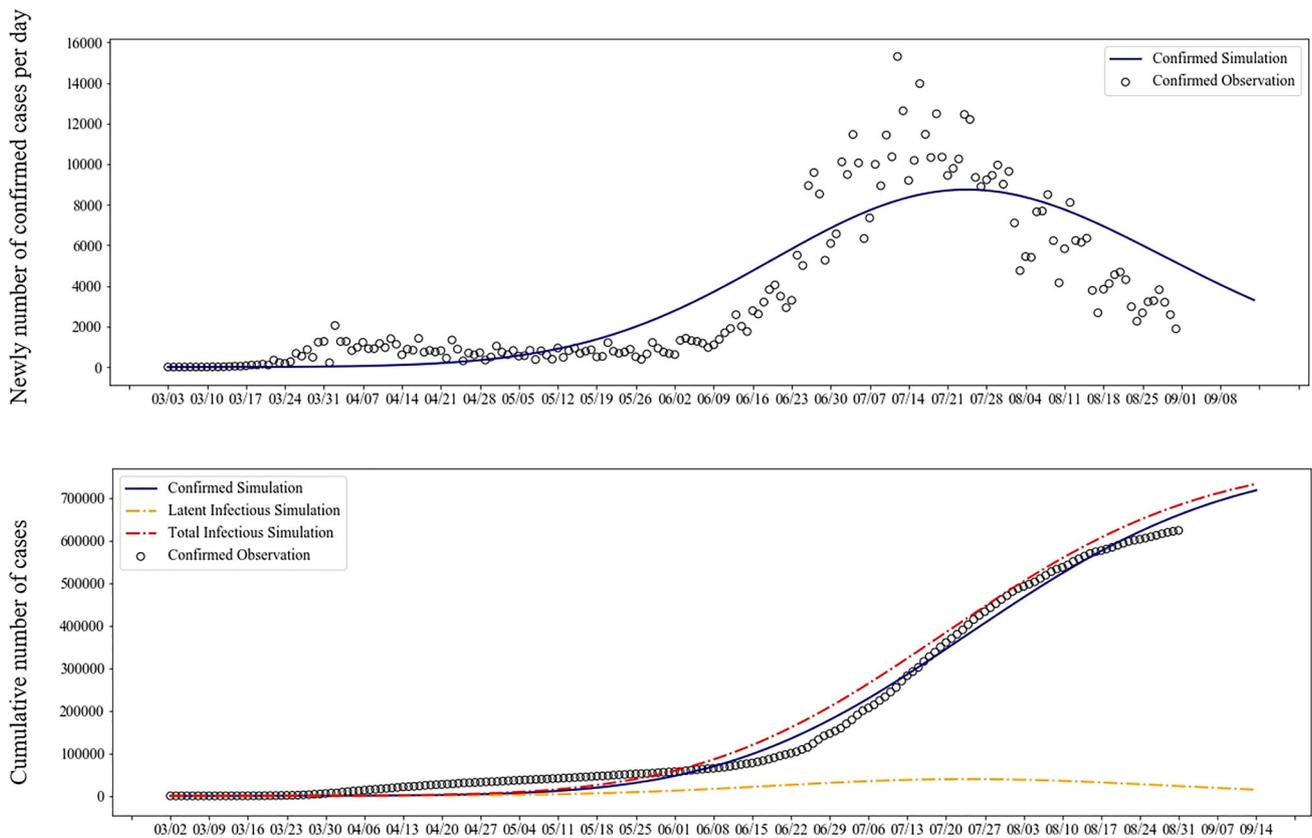


Fig. 7 Simulations of infections in Florida on SIC@intra model

4.2.1 Evaluation metrics

Three metrics, MAPE, PCC and p value, are used to evaluate the performance of the proposed methods, where MAPE and PCC measure the consistency of our results with real reported data, and p value assesses the significance of difference among different models.

- PCC: The Pearson correlation coefficient is used to measure the linear correlation between curve of reported infectious numbers with that generated by our method,

$$PCC = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (\hat{y}_t - \bar{\hat{y}})^2} \sqrt{\sum_{t=1}^T (y_t - \bar{y})^2}}, \quad (10)$$

where \hat{y}_t and y_t denote the predicted and actual confirmed cases at time slot t , respectively. $\bar{\hat{y}}$ and \bar{y} are the average values of all predicted and actual confirmed cases during the city time window T .

- MAPE: The mean absolute percentage error estimates the average of absolute percentage error,

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{\hat{y}_t - y_t}{y_t} \right|. \quad (11)$$

- p value: p value denotes the probability of observing results at least as extreme as the measured results of a statistical hypothesis test when the assumption of the null hypothesis is correct [16]. Here, we compute p value to show significant differences between the performances of proposed models.

4.2.2 Implementation details

The models are implemented with Python. The initial experimental values are set as follows. S_0 is usually approximate to the total population of a region. I_0 is composed of I_0^{loc} and I_0^{im} . The product of C_0 and a coefficient k_c is taken as an approximation of I_0^{loc} , and k_c is obtained by grid search method. I_0^{im} can be calculated by Eq. (7). C_0 is the number of confirmed cases on the first day. Grid search is utilized to find the best initial settings of hyperparameters (i.e. k_{intra} , k_{inter} , β , α , k_c). The optimal solution of the dynamic partial differ-

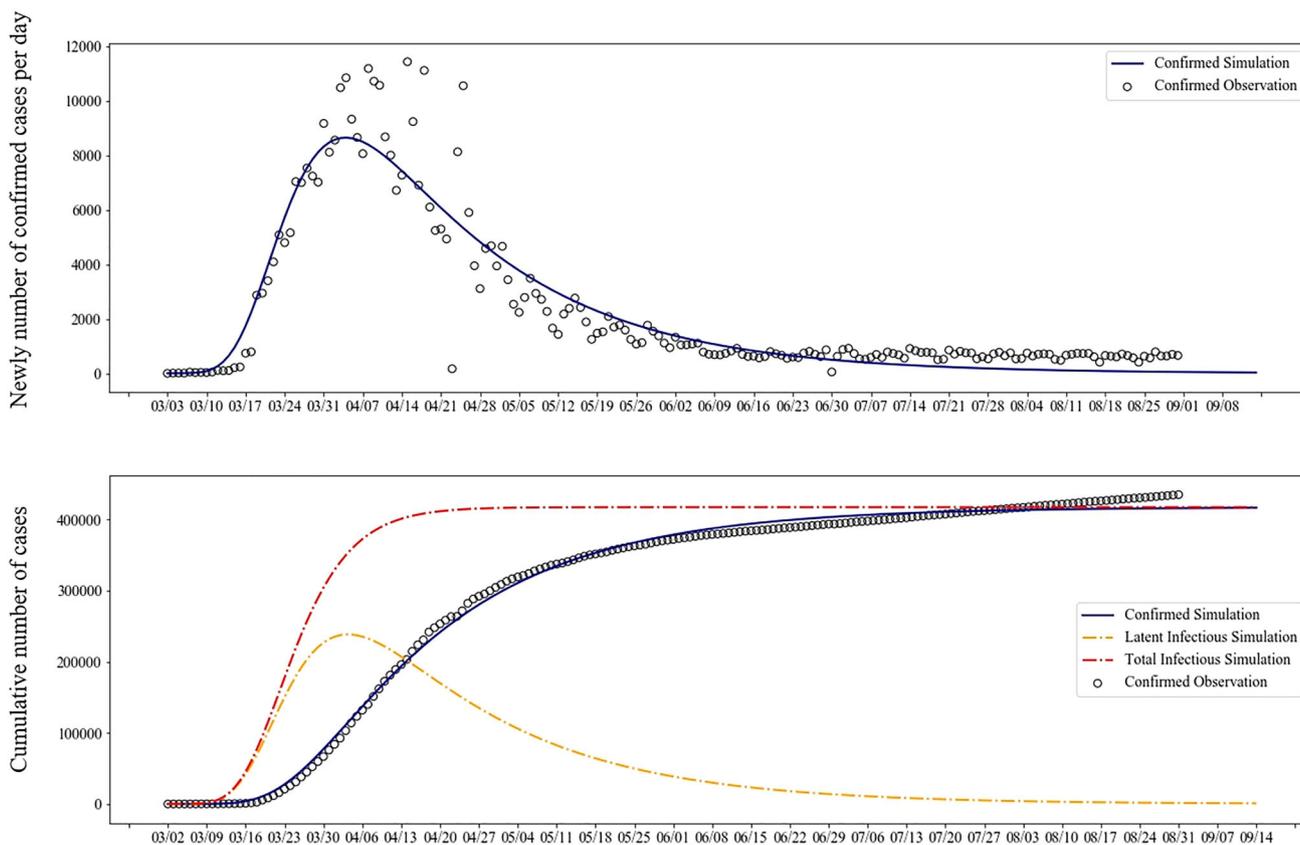


Fig. 8 Simulations of infections in New York on SIC@intra model

ential equations can be learned in the process of minimizing loss by the least squares method,

$$\mathcal{L} = \sum_{t=1}^T (\hat{C}_t - C_t)^2 + \sum_{t=1}^T (\hat{A}_t - A_t)^2 + \sum_{t=1}^T (\hat{\mathcal{I}}_t^{im} - \mathcal{I}_t^{im})^2, \quad (12)$$

where T represents the number of epidemic days in a region for training. The three penalty terms guide our model to approximate the real observed data as much as possible. Scipy optimization module is adopted to implement the least squares optimizer.

4.3 Experimental results

We first present a visual result to show that the proposed method could approximately fit the real reported data, see Fig. 4. Then, we compare four models to investigate the importance of introducing human mobility to the epidemic models as shown in Tables 1 and 2.

In Fig. 4, the blue curves are generated by the proposed SI2C model, the circles correspond to daily confirmed cases. The simulated curves of eight representative cities are shown in this figure; it can be known that, for whichever city, our model could well capture the epidemic dynamics of COVID-

19. To evaluate the effectiveness of the proposed methods, we compute the consistency of generated results with real-world reported data with regard to three metrics, see Table 1. It can be known that SIC@inter and SIC@intra achieve consistent lower MAPE and higher PCC than the SIC model, which suggests that taking either the inter-city or intra-city human mobility into consideration could improve the model performance. And SI2C model outperforms other models in most cases, which means that it is necessary to incorporate both intra-city and inter-city human mobility into the epidemic model.

Table 2 presents the significance tests of performance improvement by considering human mobility in epidemic models with regard to MAPE and PCC computed by p value, which is calculated by single sample T test. From Table 2, several key conclusions could be summarized.

Firstly, *both intra-city and inter-city human mobilities should be considered in the epidemic model of COVID-19.* As shown in the last two columns in Table 2, compared to SIC model, either the improvement of SIC@intra or that of SIC@inter is significant with regard to MAPE and PCC metrics.

Secondly, *the intra-city human mobility plays an even more important role than that of inter-city human mobility in*

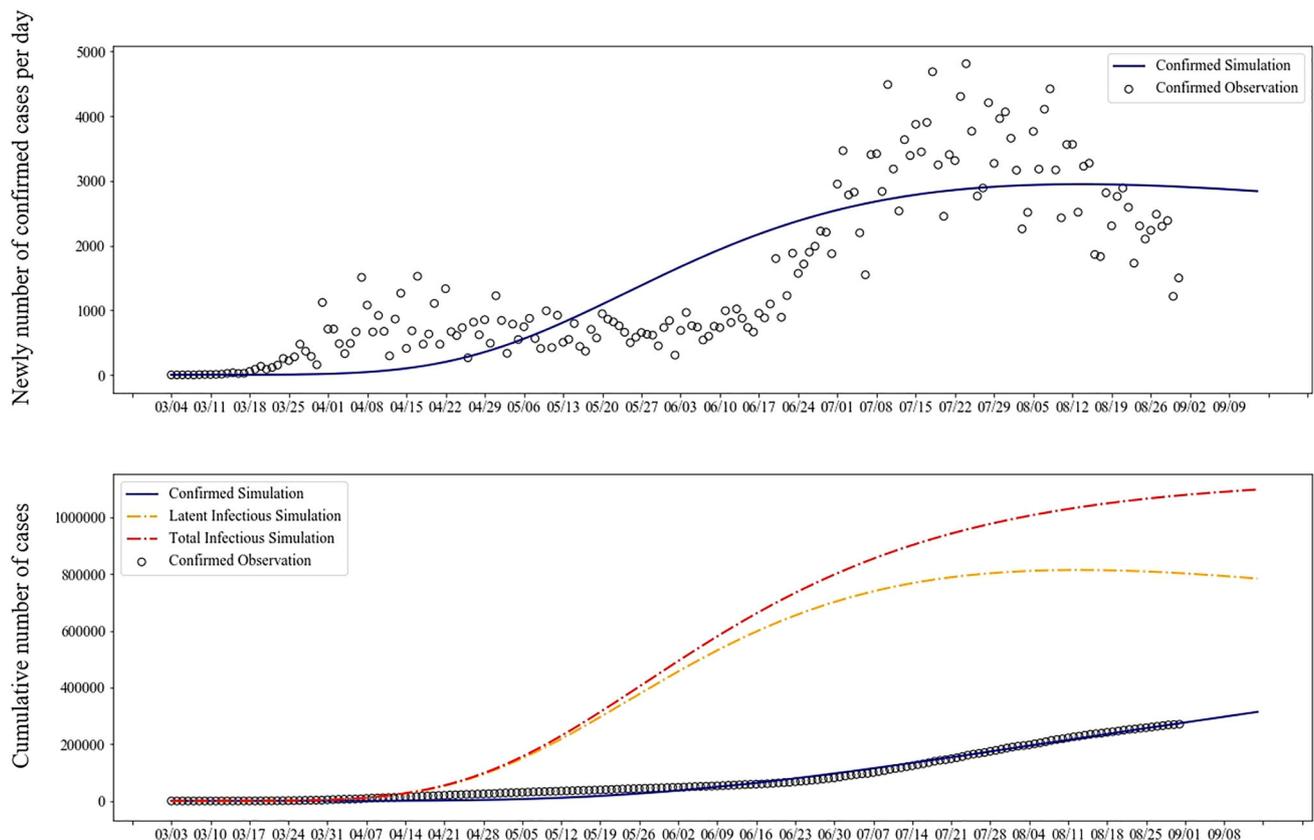


Fig. 9 Simulations of infections in Georgia on SIC@intra model

the study period. See the fourth column of Table 2; SIC@intra achieves significant better results than SIC@inter. Maybe the reason is that the lock-down measure of Hubei province stopped the exports of infectious sources to other cities. At the very initial period of epidemic, immigrant infectious sources took a large proportion of the disease spreader. After that, the local infectious sources become the main force of COVID-19 spreading. However, the intra-city human mobility has not been considered in the existing research [15,23,27].

Finally, the difference between SI2C and SIC is the most significant of all comparisons, which means that considering the impact of both population migration between regions and human activity within regions on the spread of epidemic is necessary. The remarkable superiority of SI2C also exactly proves our model design is optimal.

5 Experiments on the US dataset

We then conduct experiments on eight states in the USA to evaluate the performance of our models. As the inter-state human mobility data are not available, only SIC@intra and SIC models are implemented.

5.1 Dataset description

The collected data include epidemic and intra-state human mobility data of the USA from March 1 to August 31, 2020. The top eight states with most cumulative number of confirmed cases till to August 31, 2020 are studied. The epidemic data are obtained from CDC. The intra-state human mobility data are from Descartes Labs [24]. Up to now, since the spread of COVID-19 in the USA is not well controlled, the number of confirmed cases is still growing. Hence, in addition to fitting historical confirmed records, it is also meaningful to predict the spread of COVID-19 in the USA in the near future. Therefore, the whole US dataset is divided into training set and test set for comprehensive experiments. Specifically, the data of last 30 days are utilized as test set, and the remaining data are utilized as training set. The experimental results are observed and compared on both fitting data and prediction data.

5.2 Experimental results

Our model could not only fit the collected data by a learned curve, but also predict the future evolution of epidemic according to the curve. Additionally, our model could esti-

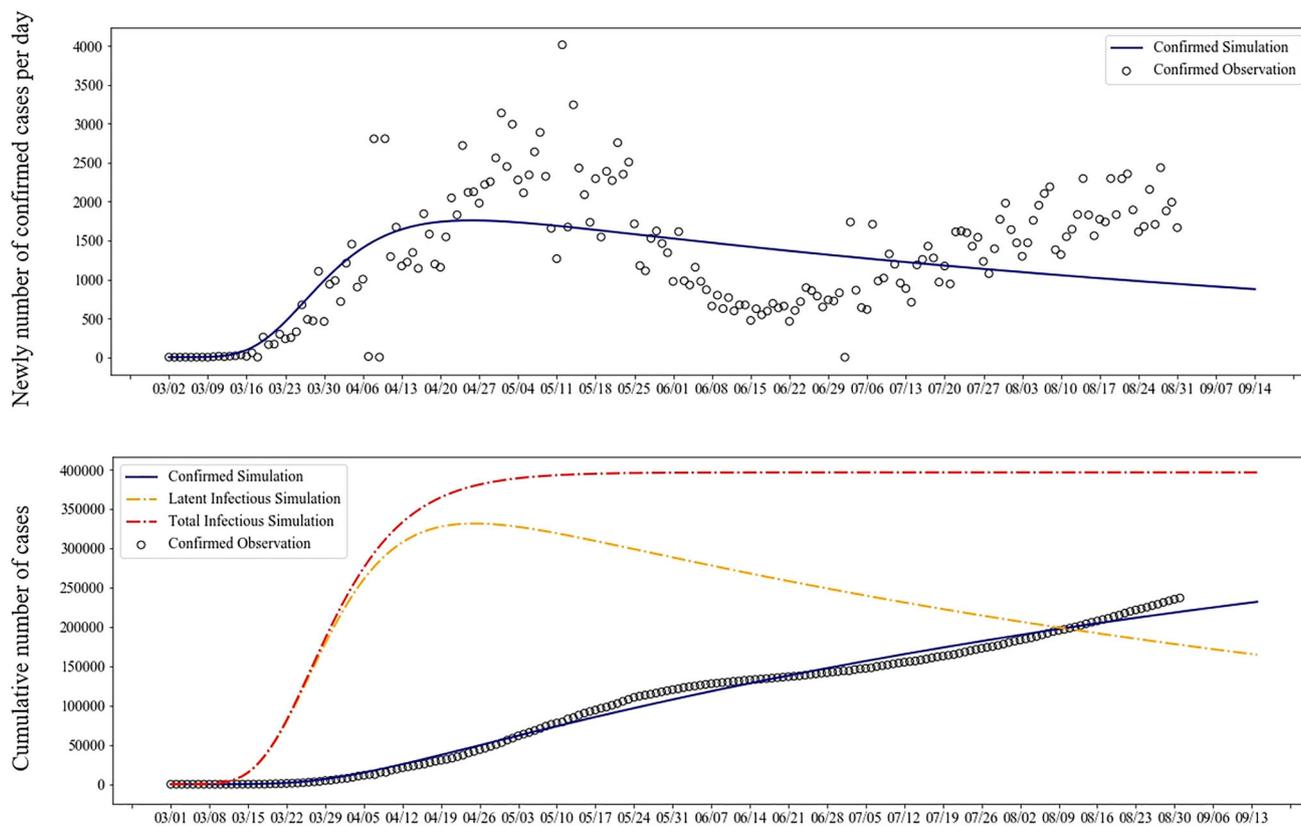


Fig. 10 Simulations of infections in Illinois on SIC@intra model

mate the latent infectious persons and the total number of infectious persons, which could help make decisions of preventive measures.

Figure 5 shows the simulations of SIC@intra model on California and New Jersey, which are states with the largest and least number of total confirmed cases separately among the selected eight states as of August 31. (The simulation results of other six states are shown in Appendix, Figs. 6, 7, 8, 9, 10, 11.) The upper sub-figures show results on the daily new confirmed curve, while the lower sub-figures show results on the cumulative confirmed curve. It can be seen that our model could approximately the real numbers of reported cases. The yellow line in lower sub-figures shows the evolution of numbers of latent infectious persons, which could illustrate in which period a state is undergoing, and the red line describes the evolution of total infectious people, which is achieved by adding the blue curve with the yellow one. For New Jersey, COVID-19 is gradually under control with the decreasing growth rates of confirmed cases and the numbers of confirmed cases and total infections tend to reach the peak. For California, the government is expected to strengthen epidemic prevention and take control measures.

Tables 3 and 4 show the quantitative evaluation of SIC@intra and SIC models on the fitting data and prediction data respectively. It could be observed that SIC@intra

outperforms SIC significantly, demonstrating the importance of incorporating human mobility into the epidemic models. And the experimental results in Table 4 fully prove the accuracy of our model predictions, which could provide scientific references for decision-makers to formulate epidemic management strategies.

6 Conclusion

It is well known that patient quarantine and social distancing could reduce the spreading of COVID-19, but how to analyse the epidemic dynamics of COVID-19 quantitatively with consideration of these factors is still a challenging problem. This paper proposes a novel infectious disease spreading model to cope with this problem by first introducing a confirmed status into model to coincide with the patient quarantine and then incorporating two types of human mobility with the above model such that intra-region human mobility describes the disease spreading speed due to the social distancing and inter-region human mobility reflects the immigrant infectious sources. Our model can be widely applied for the simulation and prediction of various infectious diseases which have no incubation period or the incubation period is also infectious. Note that our model further takes

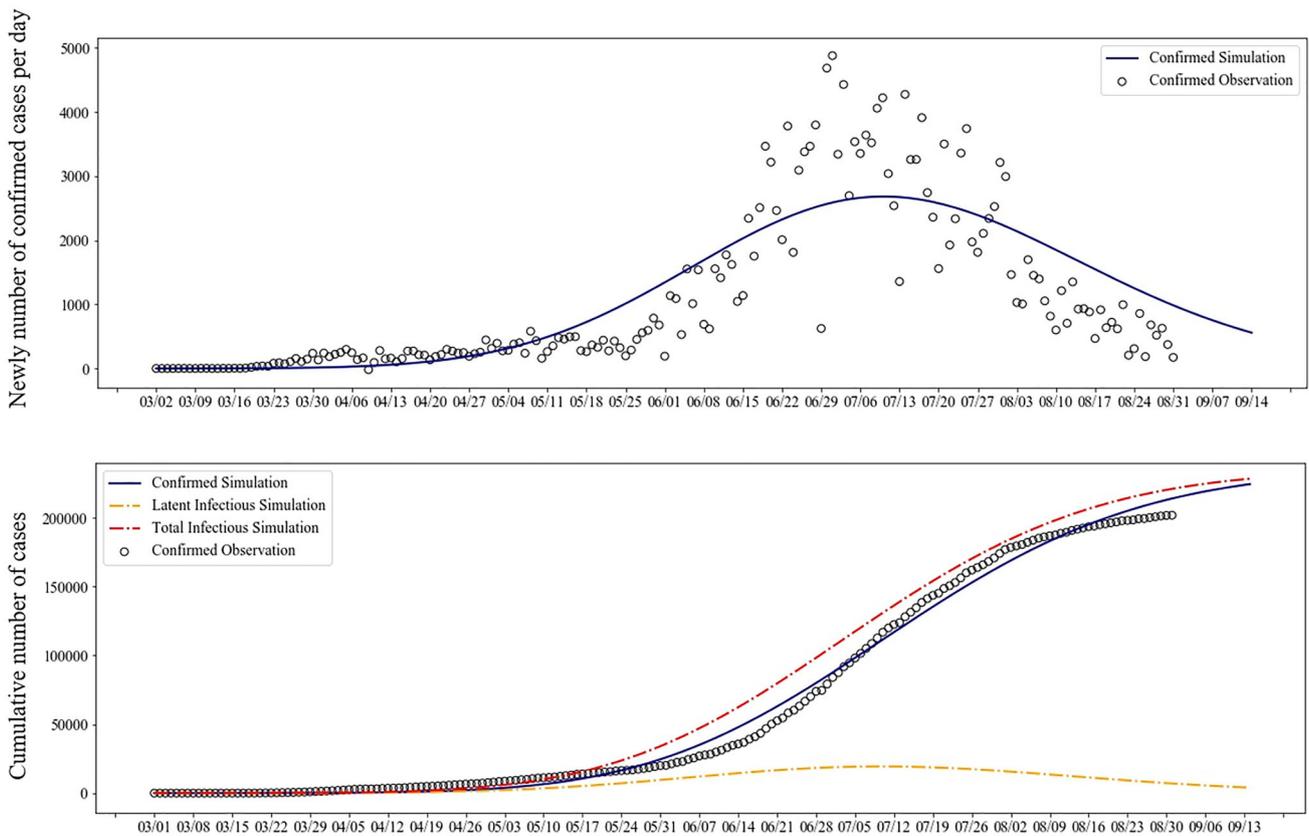


Fig. 11 Simulations of infections in Arizona on SIC@intra model

Table 3 The performance of SIC@intra and SIC models on the US fitting data

American states	MAPE		PCC	
	SIC@intra	SIC	SIC@intra	SIC
California	0.327	0.883	0.9981	0.9841
Texas	0.347	1.121	0.9980	0.9988
Florida	0.411	0.893	0.9957	0.9979
New York	0.153	0.319	0.9985	0.9766
Georgia	0.420	0.667	0.9922	0.9910
Illinois	0.184	0.264	0.9950	0.9927
Arizona	0.324	0.764	0.9968	0.7956
New Jersey	0.092	0.333	0.9989	0.9747

Table 4 The performance of SIC@intra and SIC models on the US prediction data

American states	MAPE		PCC	
	SIC@intra	SIC	SIC@intra	SIC
California	0.006	0.090	0.9982	0.9904
Texas	0.005	0.006	0.9993	0.9993
Florida	0.020	0.030	0.9980	0.9972
New York	0.021	0.090	0.9918	0.8447
Georgia	0.021	0.028	0.9954	0.9952
Illinois	0.032	0.037	0.9986	0.9836
Arizona	0.027	0.343	0.9985	0.9405
New Jersey	0.022	0.095	0.9963	0.8682

into account the impact of human mobility and prevention measures on the spread of epidemic. Hence, when the above information is available and affects the spread of epidemic, our approach can model the development of epidemic in real scenarios more accurately. We have conducted a comprehensive empirical study to reveal the characteristics of intra-city and inter-city mobility patterns under a series of preventive measures. To verify the robustness and general applicability of our model, experiments are conducted on COVID-19

data of 24 cities in China and 8 states in the USA, and the experimental results demonstrated its effectiveness and superiority. Additionally, our research also provides a new insight into study the epidemic dynamics of COVID-19 from the perspective of human mobility. Particularly, the intra-city human mobility plays an important role in epidemic modelling of COVID-19, but it was ignored in most of the existing research.

Acknowledgements We thank the reviewers for their constructive comments on this research work. This work was supported by Beijing Science and Technology Project (Z181100009018010) and the National Natural Science Foundation of China (U1811463, 71901011, 51822802, 51778033, 51991395).

A Simulations on six states of the USA

In the appendix, the simulations of infections in other six states (except California and New Jersey) of the USA on SIC@intra model are shown.

References

- Annas, S., Isbar Pratama, M., Rifandi, M., Sanusi, W., Side, S.: Stability analysis and numerical simulation of Seir model for pandemic COVID-19 spread in Indonesia. *Chaos Solitons Fractals* **139**, 110072 (2020)
- Arik, S., Li, C.L., Yoon, J., Sinha, R., Epshteyn, A., Le, L., Menon, V., Singh, S., Zhang, L., Nikoltchev, M., et al.: Interpretable sequence learning for COVID-19 forecasting. *Adv. Neural Inf. Process. Syst.* **33** 18807–18818 (2020)
- Biswas, M.H.A., Paiva, L.T., De Pinho, M.: A Seir model for control of infectious diseases with constraints. *Math. Biosci. Eng.* **11**(4), 761–784 (2014)
- Cadoni, M., Gaeta, G.: How long does a lockdown need to be? *arXiv preprint arXiv:2004.11633* (2020)
- Chang, S.L., Harding, N., Zachreson, C., Cliff, O.M., Prokopenko, M.: Modelling transmission and control of the COVID-19 pandemic in Australia. *Nat. Commun.* **11**(1), 1–13 (2020)
- Chen, S., Yang, J., Yang, W., Wang, C., Bärnighausen, T.: COVID-19 control in china during mass population movements at new year. *The Lancet* **395**(10226), 764–766 (2020)
- Cohen, J.E.: Infectious diseases of humans: dynamics and control. *JAMA* **268**(23), 3381 (1992)
- Fanelli, D., Piazza, F.: Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* **134**, 109761 (2020)
- Gaeta, G.: A simple sir model with a large set of asymptomatic infectives. *Mathematics in Engineering* **3**(2), 1–39 (2021)
- Giles, P.: The mathematical theory of infectious diseases and its applications. *Journal of the Operational Research Society* **28**(2), 479–480 (1977)
- Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C.L., Hui, D.S., et al.: Clinical characteristics of 2019 novel coronavirus infection in China. *MedRxiv* (2020)
- Hernández-Orallo, E., Armero-Martínez, A.: How human mobility models can help to deal with COVID-19. *Electronics* **10**(1) 33 (2021)
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al.: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**(10223), 497–506 (2020)
- Kissler, S.M., Tedijanto, C., Goldstein, E., Grad, Y.H., Lipsitch, M.: Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science* **368**(6493), 860–868 (2020)
- Kraemer, M.U., Yang, C.H., Gutierrez, B., Wu, C.H., Klein, B., Pigott, D.M., du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., et al.: The effect of human mobility and control measures on the COVID-19 epidemic in China *Science* **368**(6490), 493–497 (2020)
- Krzywinski, M., Altman, N.: Points of significance: importance of being uncertain *Nature Methods* **10**(9), 809–810 (2013)
- Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J.: The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine* **172**(9), 577–582 (2020)
- Ndairou, F., Area, I., Nieto, J.J., Torres, D.F.M.: Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solitons Fractals* **135**, 109846 (2020)
- Persson, J., Parie, J.F., Feuerriegel, S.: Monitoring the COVID-19 epidemic with nationwide telecommunication data. *arXiv preprint arXiv:2101.02521* (2021)
- Piovella, N.: Analytical solution of Seir model describing the free spread of the COVID-19 pandemic. *Chaos Solitons Fractals* **140**, 110243 (2020)
- Riley, S.: Large-scale spatial-transmission models of infectious disease. *Science* **316**(5829), 1298–1301 (2007)
- Thapa, P.: Predicating COVID-19 epidemic in Nepal using the sir model. *Department of Geomatics Engineering* (2020)
- Tian, H., Liu, Y., Li, Y., Wu, C.H., Chen, B., Kraemer, M.U., Li, B., Cai, J., Xu, B., Yang, Q., 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)
- Warren, M.S., Skillman, S.W.: Mobility changes in response to COVID-19. *arXiv preprint arXiv:2003.14228* (2020)
- Weiss, H.H.: The Sir model and the foundations of public health (2013). *Materials Mathematics*, 1–17
- Wilder-Smith, A., Chiew, C.J., Lee, V.J.: Can we contain the COVID-19 outbreak with the same measures as for sars? *The Lancet Infectious Diseases* **20**(5), 102–107 (2020)
- Wu, J.T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet* **395**(10225), 689–697 (2020)
- Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., He, J.: Modified Seir and Ai prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **12**(3) 165–174 (2020)

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