



Research article

Spatial-temporal diffusion model of aggregated infectious diseases based on population life characteristics: a case study of COVID-19

Wen Cao^{1,*}, Siqi Zhao¹, Xiaochong Tong², Haoran Dai³, Jiang Sun⁴, Jiaqi Xu¹, Gongrun Qiu⁵, Jingwen Zhu¹ and Yuzhen Tian¹

¹ School of Geoscience and Technology, Zhengzhou University, Zhengzhou 450001, China

² School of Geospatial Information, University of Information Engineering, Zhengzhou 450001, China

³ Northern Information Control Research Institute Group Co. Ltd, Nanjing 211106, China

⁴ Beijing QMap Technology Co. Ltd, Beijing 100192, China

⁵ Beijing 61618 troops, Beijing 100088, China

* **Correspondence:** Email: zzdx_edifier@zzu.edu.cn.

Abstract: Outbreaks of infectious diseases pose significant threats to human life, and countries around the world need to implement more precise prevention and control measures to contain the spread of viruses. In this study, we propose a spatial-temporal diffusion model of infectious diseases under a discrete grid, based on the time series prediction of infectious diseases, to model the diffusion process of viruses in population. This model uses the estimated outbreak origin as the center of transmission, employing a tree-like structure of daily human travel to generalize the process of viral spread within the population. By incorporating diverse data, it simulates the congregation of people, thus quantifying the flow weights between grids for population movement. The model is validated with some Chinese cities with COVID-19 outbreaks, and the results show that the outbreak point estimation method could better estimate the virus transmission center of the epidemic. The estimated location of the outbreak point in Xi'an was only 0.965 km different from the actual one, and the results were more satisfactory. The spatiotemporal diffusion model for infectious diseases simulates daily newly infected areas, which effectively cover the actual patient infection zones on the same day. During the mid-stage of viral transmission, the coverage rate can increase to over 90%, compared to related research, this method has improved simulation accuracy by approximately 18%. This study can provide technical support for epidemic prevention and control, and assist decision-makers in developing more scientific and efficient epidemic prevention and control policies.

Keywords: infectious diseases; outbreak points; temporal evolution; diffusion model; epidemic prevention

1. Introduction

A novel coronavirus called SARS-CoV-2 (COVID-19) has had a severe global public health and economic impact since its outbreak in late 2019 [1]. The pandemic has spread rapidly across the globe in a short period of time, resulting in millions of infections. Governments have taken a series of emergency measures to deal with the outbreak [2], such as lockdowns, travel restrictions, quarantines, and social distancing. However, virus mutations, uneven vaccination rates, and varying implementation effects of public health policies have allowed the outbreak to persist.

To further improve public health policies and reduce the harm caused by the virus to humans, many scholars have conducted research on COVID-19 through mathematical modeling approaches. In the early stages of the COVID-19 outbreak, most studies utilized patient data from affected countries or regions to estimate epidemiological parameters of the virus, such as the basic reproduction number [3–5] and the incubation period [6,7]. Building upon this foundation, traditional and modified dynamic models were employed to predict the potential scale of the virus, further supporting epidemic prevention and control efforts [8–12]. However, these studies typically treated the epidemic areas as a whole, neglecting the impact of regional heterogeneity on viral transmission. In reality, the spread of infectious diseases is influenced by a complex interplay of external factors, including subjective factors (preventive measures and population mobility) and objective factors (geographical environment and population distribution). As a result, these studies could only depict the temporal sequence of changes in the infectious disease but could not provide better assistance for epidemic prevention and control. As early as 2005, Keeling and Eames [13] delved into the role of network models in describing the spread of infectious diseases, emphasizing the impact of individual contact patterns on viral transmission. Similarly, Christakis and Fowler [14] utilized social network models to anticipate and detect disease outbreaks. Additionally, some studies [15,16], considering both positive and negative factors within the network, have explored the influence of network models on the dynamics of epidemic propagation. Therefore, in subsequent studies on the transmission of COVID-19, some researchers [17–19] have introduced social network analysis models. These models aim to explore the impact of population movement on virus spread and to predict the course of the epidemic at the level of individual contact. Social network models simulate small urban clusters (such as schools, businesses, and areas surrounding city centers), but experiments typically involve random movement simulations based on hypothetical points within a spatial context. These simulations only incorporate the mobility characteristics of certain urban areas and population groups, lacking the integration of spatiotemporal big data of urban geography. Consequently, there is still a significant discrepancy between the simulated urban distribution and population movement patterns and the actual situation. As our understanding of the virus deepens, some scholars have conducted research on spatial evolution prediction [20], infection area identification [21], risk indicator assessment [22–24], spatial anomaly detection [25], and spatiotemporal visualization [26,27]. However, most of these studies are conducted at the provincial level or higher, lacking research on the transmission of infectious diseases within urban areas. In order to simulate the epidemic development within cities, some studies have utilized multi-agent models to construct urban population mobility by leveraging fine-grained data to

simulate the interaction process among the urban population [28]; these data include mobile phone location data [29,30] and public transportation card data [31]. However, due to the complexity of population mobility, the simulated population movement process exhibits significant errors compared to actual data. Moreover, most of these studies are based on simulations in developed regions. The applicability of these models to underdeveloped areas remains questionable, as obtaining fine-grained data may be hindered by outdated equipment. This issue inevitably reduces the universality of the models.

Therefore, in light of the aforementioned issues, to better assist governments and public health departments in formulating more efficient epidemic prevention and control strategies, this study proposes a spatiotemporal diffusion model for infectious diseases based on population living characteristics. The main contributions of this paper are as follows:

1) By conducting a fine-grained discrete grid division of the research area, we can achieve a more refined simulation of the virus propagation process within the city.

2) Utilizing patients' trajectory data, we have estimated the approximate location of virus outbreak points, addressing the issue of unclear outbreak locations due to population mobility. This provides a methodological foundation for other studies.

3) By employing multi-source data in place of population observation data, we have resolved the issue of data ambiguity in underdeveloped areas caused by inadequate facilities, thus enhancing the universality of the model.

The remainder of this paper is organized as follows. In Section 2, we introduce the data sources used in the study and the proposed methodology. In Section 3, we provide validation and analysis of the experimental results. Section 4 discusses the virus outbreak points and the spatial diffusion of the virus. Finally, in Section 5, we summarize the contents of this research.

2. Data and methods

2.1. Data collection

1) Social environment data

The social environment data primarily consist of population distribution, man-made surface, nighttime lighting, transportation station, and road network traffic data. Notably, man-made surface data can be derived from land use data. These datasets, to a certain extent, represent the spatial aggregation characteristics of human mobility and daily life. The sources of the relevant data are presented in Table 1.

Table 1. The information of collected social environment data.

Variables	Year	Precision	Data sources
Population distribution data	2020	100 m	https://www.worldpop.org/
Land use data	2018	30 m	http://data.ess.tsinghua.edu.cn
Night lighting data	2018	130 m	http://59.175.109.173:8888/
Traffic station data	2021	\	https://download.geofabrik.de
Road network traffic data	2021	\	https://download.geofabrik.de

2) City point-of-interest(POI) data

City POI data is a concentrated representation of local culture, population density, and lifestyle, directly reflecting the spatial distribution of urban infrastructure and the primary areas where people gather for activities. The data is sourced from Gaode Map and was updated in November 2021. It has been re-categorized into seven groups based on their relevance to people's daily lives, including Business Offices, Shopping & Dining, Medical & Education, Financial Services, Residential Communities, Leisure & Entertainment, and Life Services.

3) Epidemic data

This study gathered epidemic data from Wuhan, Xi'an, Shanghai, and Hong Kong, which were officially released by each respective city. The specific data sources and time ranges are displayed in Table 2.

Table 2. Time range and sources of epidemic data.

City	Date	Data sources
Wuhan	2019/12/8/~2020/4/15	http://wjw.wuhan.gov.cn/ztzl_28/fk/tzgg/
Xi'an	2021/12/9/~2022/1/20	http://xawjw.xa.gov.cn/ztzl/fyfk/gzdt/
Shanghai	2022/3/1~2022/5/11	https://wsjkw.sh.gov.cn/yqtb/
Hong Kong	2022/3/1~2022/3/21	https://www.chp.gov.hk/sc/

4) Spatial distribution data of patients

The patient spatial distribution data consists of the spatial flow trajectory data for the day the patient was detected and the preceding days. Different infection statuses across cities result in varying pressure for epidemic prevention, which indirectly leads to differences in the level of detail in the published patient spatial distribution data. Due to the initial outbreak of the epidemic in Wuhan, early-stage patient spatial distribution data was incomplete. Consequently, this study collected images of selected community outbreak announcements published by relevant public accounts related to life in Wuhan, and compiled patient data for Wuhan based on these images. The patient spatial distribution data for the remaining cities was officially published, with the time range consistent with the epidemic data. The data sources are presented in Table 3.

Table 3. Patient spatial distribution data sources.

City	Data sources
Wuhan	/
Xi'an	https://weibo.com/huashangbao
Shanghai	https://wsjkw.sh.gov.cn/yqtb/
Hong Kong	https://data.gov.hk/sc-data/dataset/

2.2. Methods

2.2.1. Infectious disease dynamics model

To represent the progression of infectious diseases in time series, this study employed the SEAIR model to predict infection cases within cities. The model was proposed by Okuonghae and

Oname [32] based on the classical compartmental model, which assumes a completely random distribution of the population within the study area and does not consider population in-migration and out-migration. The total population (N) in the infected area is categorized into susceptible population (S), exposed population (E), asymptomatic infected population (A), symptomatic infected population (I), isolated population (ID), and recovered population (R), as demonstrated in Eqs (1)–(7). Therefore, at time t , $N(t) = S(t) + E(t) + A(t) + I(t) + ID(t) + R(t)$. The flowchart of the SEAIR infectious disease model is illustrated in Figure 1, and the model parameters are referenced from relevant literature [32] as shown in Table 4.

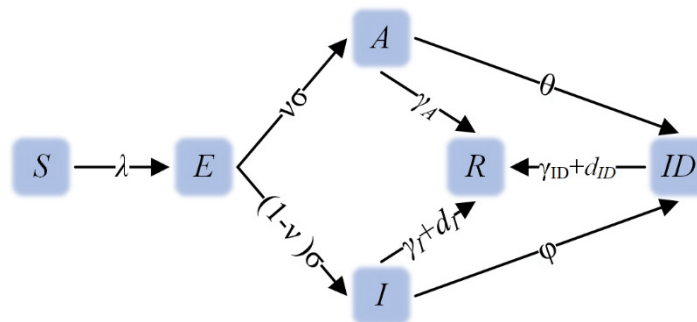


Figure 1. The flow chart of the SEAIR infectious disease model.

$$\lambda = \frac{\beta(\alpha A + I)}{(N - ID)} \quad (1)$$

$$\frac{dS}{dt} = -\lambda S \quad (2)$$

$$\frac{dE}{dt} = \lambda S - \sigma E \quad (3)$$

$$\frac{dA}{dt} = \nu \sigma E - (\theta + \gamma_A) A \quad (4)$$

$$\frac{dI}{dt} = (1 - \nu) \sigma E - (\varphi + \gamma_I + d_I) I \quad (5)$$

$$\frac{d(ID)}{dt} = \theta A + \varphi I - (\gamma_{ID} + d_{ID})(ID) \quad (6)$$

$$\frac{dR}{dt} = \gamma_{ID}(ID) + \gamma_A A + \gamma_I I \quad (7)$$

where β is the effective transmission rate, α is the modification parameter for the reduction in the transmission rate of asymptomatic infected individuals relative to symptomatic infected individuals, σ is the probability of conversion of latent state to infected state, ν is the proportion of asymptomatic infected individuals, γ is the recovery rate of the infected population, d is the lethality of disease, and θ and φ are the detection rates of asymptomatic and symptomatic infected individuals, respectively.

Table 4. Values of parameters in the model.

Parameter	Reference	Range
E	Fitted	Estimated
A	Fitted	Estimated
I	Fitted	Estimated
β	Fitted	Estimated
θ	Fitted/day	Estimated
φ	Fitted /day	Estimated
α	0.5/day	[0,1]/day
ν	0.5/day	[0,1]/day
σ	0.19231/day	[1/14,1/3]/day
γ_C	0.06667/day	[1/30,1/3]/day
$\gamma_A = \gamma_I$	0.13978/day	[1/30,1/3]/day
$d_c = d_I$	0.015/day	[0.001,0.1]/day

2.2.2. Estimation methodology for infectious disease outbreak points, accounting for virus latency and population movement

The identification of virus outbreak points is important for effective control of the source of infection, cutting off the transmission route, and the scientific formulation of prevention and control measures. First, the SEAIR model was used to predict the number of new infections per day, and the mean(μ) and standard deviation(σ) of the predicted data were calculated. Second, assuming that the number of new infections per day conforms to a Gaussian distribution, the epidemic phase is divided according to the 3sigma criterion, with the left breakpoint noted as $T_0(\mu-3\sigma)$, and it is argued that after T_0 days, the movement of population increases the randomness of virus transmission, which interferes with the precise location of the outbreak site. Third, using the viral incubation period ($t_0 = 14$, 14 days were taken as the incubation period of the COVID-19) as the time window, the spatial distribution data of patients with a time range of $[T_0-t_0, T_0]$ days were selected ($T_0-t_0 = 0$ if $T_0-t_0 < 0$) and subjected the data to an Albers equal area conic projection. The division schematic is shown in Figure 2. As the first law of geography posits, the closer one is to an epidemic point, the greater the probability of infection. Consequently, individuals predicted to be infected are likely to display a clustered distribution around these points. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is particularly suited for this task, given its ability to identify clusters of arbitrary shapes without needing prior knowledge of the number of existing clusters [33]. Therefore, we adopted the DBSCAN algorithm to cluster the spatial distribution data of the patients after projection. The clustering outcomes of this algorithm are determined by the values of two parameters: Epsilon, which denotes the search radius, and MinPts, which signifies the minimum number of neighboring points within this radius. For our analysis, we set the Epsilon value to correspond to the typical daily living radius of the population [34], and MinPts was set as 1. Finally, the divided maximum data set is processed to find the center of gravity, which is considered as the outbreak point of the current round.

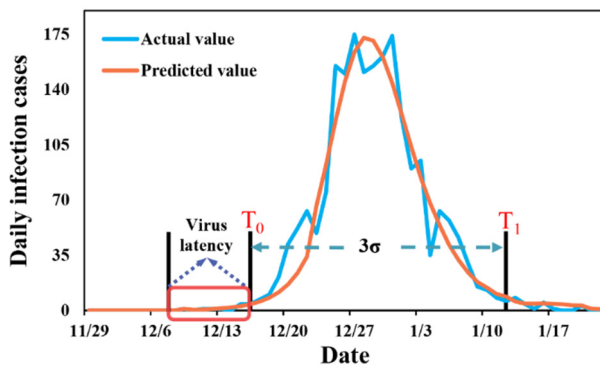


Figure 2. Schematic diagram of the epidemic curve division.

2.2.3. Quantitative analysis of the spatio-temporal aggregation characteristics of crowd activities

Spatial aggregation of crowds refers to a phenomenon where individuals actively or passively gather in specific areas due to their living needs. Quantitative analysis of spatial aggregation can reveal population interactions between different regions and provide guidance on the spatio-temporal trajectories of infectious disease spread and the extent of transmission. This study utilizes 11 types of data as quantitative indicators for analyzing spatial aggregation, which can be broadly classified into two main categories: social environment data and POI data.

First, an Albers equal area conic projection was performed on the 11 types of data using the WGS-84 coordinate system. Second, the resolution of the data was standardized by resampling or regional statistics. The distance of daily population activities is used as a parameter for kernel density analysis, which describes the intensity of daily life for the population. Third, the weights corresponding to each data type were calculated using the Analytic Hierarchy Process based on their relative importance, and the results are displayed in Table 5. Spatial overlay analysis of the data, based on the weights, was used to obtain the population life intensity weight matrix. The overall framework is shown in Figure 3. Finally, artificial surface data was employed to simulate the actual living areas of crowds, with the assumption that if the artificial surface area in the grid is small, the crowd activity within is considered low and the grid can be disregarded. To improve simulation accuracy, we set the threshold value for the minimum artificial surface area as 150 m × 150 m and removed grids with artificial surface areas smaller than the threshold value.

Table 5. Evaluation factor weighting value.

Factor	A	B	C	D	E	F	G	H	I	J	K	weight
A	1	1/9	1/7	1/3	1/5	1/9	1/5	1/9	1/8	1/10	1/7	3/250
B	9	1	3	7	5	13/10	5	2/3	3	1/2	3	157/1000
C	7	1/3	1	3	2	1/3	2	1/3	1/2	1/3	5/6	31/500
D	3	1/7	1/3	1	1/2	1/5	1/2	1/7	1/5	1/7	1/3	3/125
E	5	1/5	1/2	2	1	1/3	5/3	1/5	1/3	1/5	1/2	41/1000
F	9	10/13	3	5	3	1	3	2/3	2	2/3	3	127/1000
G	5	1/5	1/2	2	3/5	1/3	1	1/5	1/3	1/5	1/2	19/500

Continued on next page

Factor	A	B	C	D	E	F	G	H	I	J	K	weight
H	9	3/2	3	7	5	3/2	5	1	17/10	5/6	3	171/1000
I	8	1/3	2	5	3	1/2	3	10/17	1	2/5	1/3	93/1000
J	10	2	3	7	5	3/2	5	6/5	5/2	1	2	24/125
K	7	1/3	6/5	3	2	1/3	2	1/3	3	1/2	1	83/1000

Note: Factors A–J are financial services, catering and shopping, medical education, business offices, residential quarters, leisure and entertainment, life services, transportation stations, road network, population density and night lights. Consistency check: $CR = CI/RI = 0.041/1.520 = 0.027 < 0.1$.

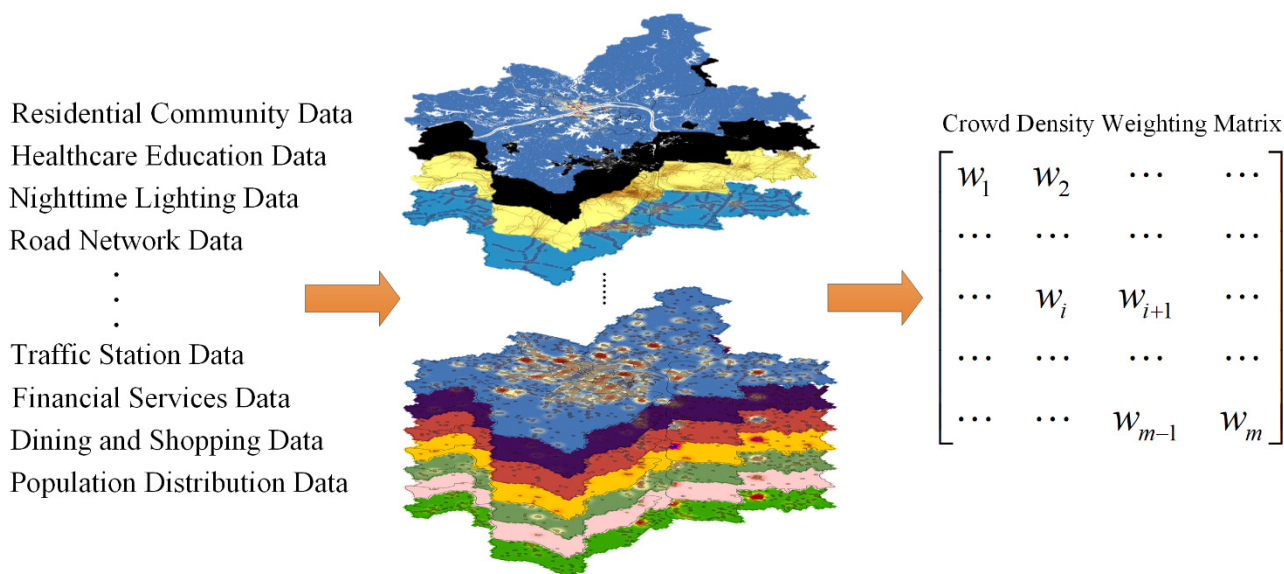


Figure 3. Framework diagram of the analysis of population life aggregation under the influence of multiple factors.

2.2.4. Spatial-temporal diffusion models of infectious diseases

Regional and virus species differences make the mode of transmission, transmission route and transmission capacity of infectious diseases vary somewhat, but they usually start from places close to the transmission route and gradually spread to surrounding areas (free state of transmission without human intervention) [35]. Figure 4 illustrates the spatial-temporal transmission pattern of infectious diseases under the life characteristics of the population. The figure uses a large scale dashed grid to generalize the area of the virus outbreak, with each grid considered as a whole, and the red grid indicates areas of infection due to the entry of an external infected population. The light blue grid indicates the area where the infection is caused by the input of infected patients in the red grid. This mode of transmission can be called ‘jump spread’, and usually this case requires the introduction of a large amount of patient spatial distribution data and refined population spatial migration data, which are difficult to obtain, so this model does not consider this case for the time being.

First, based on the mode of virus transmission, this study generalizes the spread of infectious diseases in a population using a tree structure of population lifestyles, where the spread radius can be expressed as the product of the population’s living distance and the time of virus outbreak. Second, the administrative area is gridded using the daily life range of the population as a distance parameter, and

each grid is considered as a whole, and the odds of infection within the grid are considered to be the same, and the interaction between the grids is quantified by the population density weight matrix. From the spatial transmission pattern of close contact with infectious diseases, it is clear that when the population movement in the infected area is not completely confined, the number of newly infected patients in a sub-region at $t + 1$ is influenced by the number of infected patients in that area and surrounding areas at t . Among them, patients in the infected area were most likely to be in contact with each other and infected with people in the surrounding 8 areas, but a variety of factors, such as crowd activity and distribution of facilities, led to different levels of infection in the areas of mutual contact. The specific representation is shown in the square blue area in Figure 4, where the depth of the color represents different levels of infection status. Third, in order to be able to more scientifically represent the daily infection area, assuming that the infection probability events conform to a Gaussian distribution, the actual daily infection area is filtered according to the 3sigma criterion and the weights corresponding to the daily suspected infection grid. Meanwhile, in order to be able to better quantify the probability of transmission of infectious diseases between regions, the weights of the actual infection grid are normalized, and the number of infections to be assigned in the infection grid is calculated based on the processed weights. Finally, the number of infections assigned to the grid is rounded down, where the rounded remainder is considered as an error and passed down to ensure that the number of infections assigned remains consistent with the predicted results.

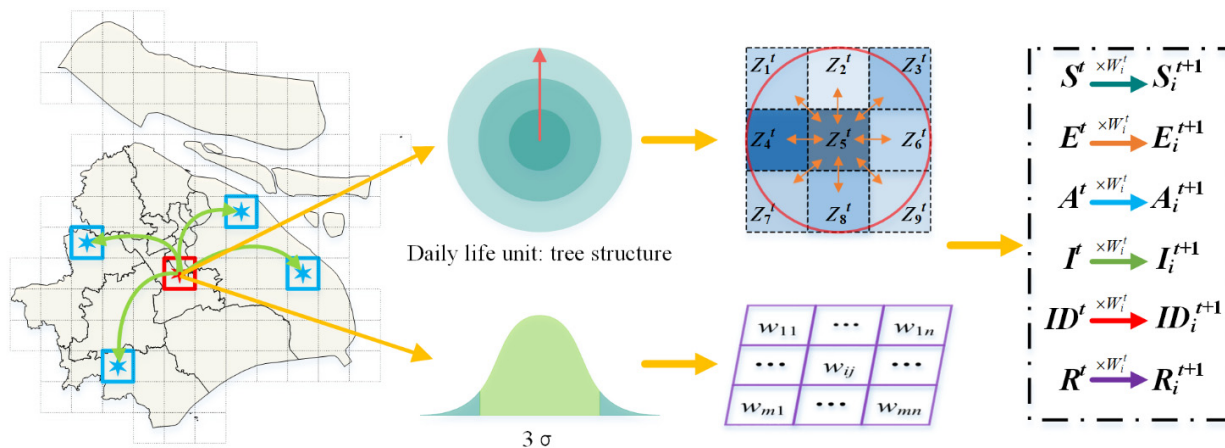


Figure 4. Spatial-temporal patterns of infectious disease transmission in the context of population lifestyle characteristics.

3. Results

3.1. Estimation of early outbreak points of infectious diseases

3.1.1. Impact of epidemic data prediction accuracy on outbreak site location

In this subsection, using Shanghai as an example and maintaining consistent prevention and control measures, the daily number of new infections in Shanghai was predicted using 10 days (2022/3/1 to 2022/3/10), 15 days (2022/3/1 to 2022/3/15), 20 days (2022/3/1 to 2022/3/20), and 40 days (2022/3/1 to 2022/4/9) of epidemic data, respectively. A visualization comparing the various predicted data with

the actual data is presented in Figure 5. The optimal parameters of the SEAIR model for different prediction data are provided in Table 6.

Table 6. Optimal parameters of Shanghai SEAIR model with different time fitted data.

Days	E	A	I	β	θ	φ
10	3248.9970	66.7767	106.812	0.9033	0.0167	0.0307
15	2519.5694	31.2486	63.4667	0.9123	0.0637	1.79E-07
20	1902.192	125.7379	281.2181	0.9999	0.0667	0.0017
40	3445.1567	606.9305	954.0432	0.9984	0.0258	0.0081

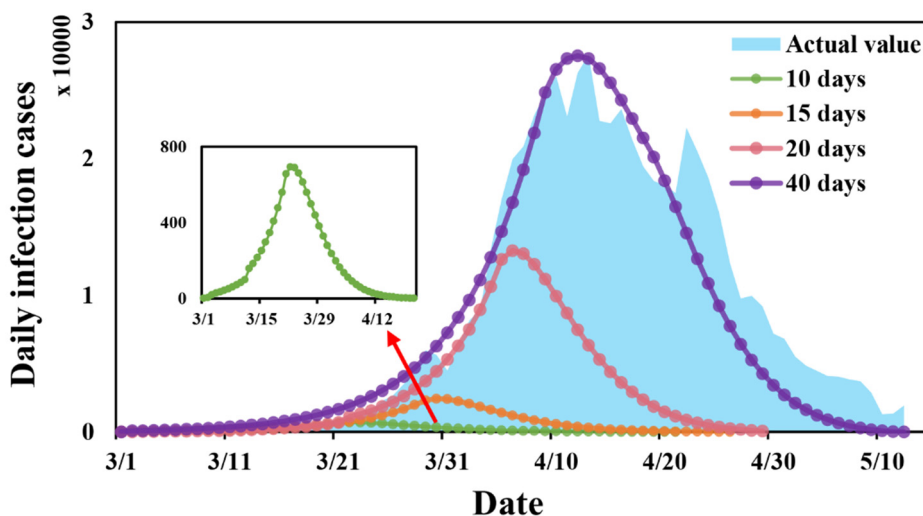


Figure 5. Forecast curves of the epidemic in Shanghai under different time fitting data.

As shown in Figure 5, the accuracy of the prediction results gradually increases with the expansion of the fitted data. This is because epidemic prediction inherently carries significant uncertainty, and any change in government decisions can cause fluctuations in the epidemic. However, as the fitted data increases, the more it reflects the current epidemic development trend, resulting in prediction results that better align with the actual situation.

Table 7 presents the calculated values of μ , σ , and T_0 derived from the projected data. However, since the spatial distribution data of patients in Shanghai is only available from March 6, 2022, it is not possible to estimate the outbreak point for the projected data with a T_0 value of less than 6.

Table 7. Parameters of Shanghai epidemic curve delineation at different times of fitted data.

Days	μ	σ	T_0
10	25.81	7.13	5
15	30.49	7.30	9
20	35.56	7.55	13
40	43.43	9.20	16

Figure 6 shows a comparison of the estimated outbreak points (subsequently referred to as

estimated outbreak points A, B, and C) using 15-, 20-, and 40-day forecast data with the officially announced outbreak point (Huating Hotel, Xuhui District, Shanghai). As seen in the figure, the inferred outbreak point A is southwest of the official outbreak point and is located at the junction of Xuhui District and Minhang District, with a distance of 5.922 km between them. The inferred outbreak points B and C are both situated within the Xuhui District of Shanghai, 3.508 and 2.767 km away from the official outbreak point, respectively.

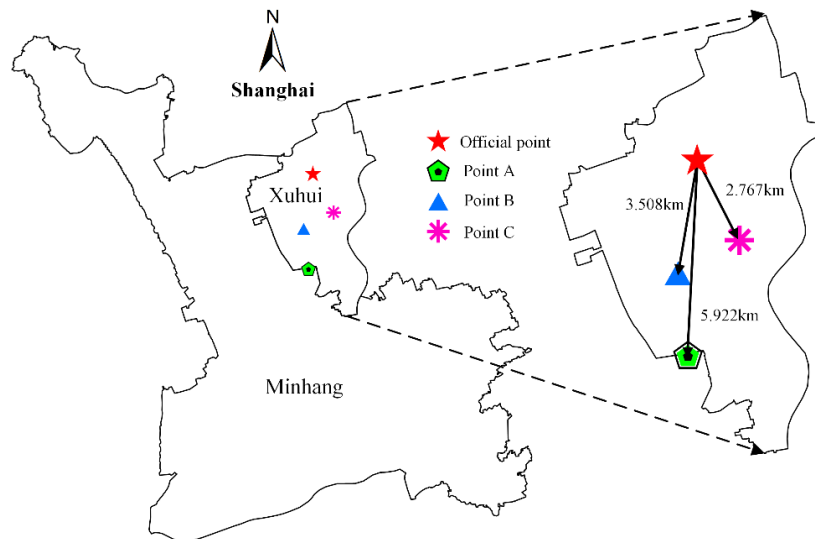


Figure 6. Estimated outbreak points in Shanghai under the fitted data of different times.

Points A and B are far from the actual outbreak site, primarily because the predicted number of infections for both differs significantly from the actual numbers, and the calculated T_0 is small. This, combined with the virus latency period, delays the time when infected patients are detected, which in turn affects the accuracy of the extrapolation of the site. Simultaneously, Shanghai officials did not publish data on the spatial distribution of patients before March 5, 2022, making the amount of data used in the extrapolation of points A and B limited and not a good representation of the spatial distribution of patients.

Compared to points A and B, outbreak point C is closer to the actual location, but there is still a gap of nearly 3 km. This is mainly because the spatial distribution data of patients published in Shanghai represents the patients' residence information. However, Shanghai, as an international city, has a massive flow of people and complex crowd activity trajectories. The location of residence is only one of many locations where patients move, which does not accurately reflect the spatial distribution trend of patients. Nevertheless, due to epidemic prevention pressure, it is still possible to extrapolate the outbreak point using the patients' residence information. As a result, outbreak point C can represent the location of the early outbreak point in Shanghai.

In summary, the higher the accuracy of the predicted data, the more consistent the delineated T_0 is with the actual situation, and the value of T_0 directly affects the range of patient spatial distribution data used. This, in turn, impacts the location of the inferred outbreak point. Although the longer the epidemic fit data, the better the accuracy of the inferred outbreak points, epidemic prevention and control work is urgent, and outbreak points A and B can also play an auxiliary role in decision-making for the early epidemic prevention and control efforts.

3.1.2. Estimation of outbreak points of the COVID-19 in different cities

This subsection uses Xi'an, Shanghai, and Hong Kong, where large-scale virus outbreaks have occurred, as examples to verify the applicability of the outbreak point projection method proposed in this paper. Outbreak point C from Subsection 3.1.1 is used as the extrapolated outbreak point for Shanghai. Since the epidemic in Xi'an was over at the time of the experiment, the SEAIR model was used to fit the daily number of new infections in Xi'an. Meanwhile, the daily number of new infections in Hong Kong was predicted, with the optimal parameters of the model shown in Table 8. The results of the comparison between the fitted or predicted number of new infections and the actual number of new infections are displayed in Figure 7.

Table 8. SEAIR model optimal parameters.

City	E	A	I	β	θ	φ
Xi'an	499.9991	77.2899	104.3007	1	0.0374	0.0262
Hong Kong	510.8527	43.8375	74.1575	1	0.0186	0.0469

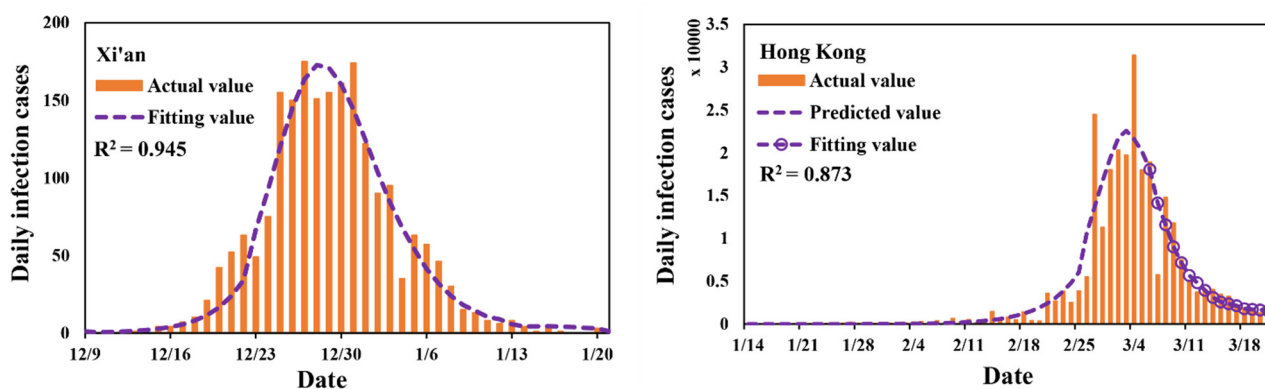


Figure 7. Comparison curve between predicted and actual number of new infections per day in Xi'an and Hong Kong.

Table 9 presents the calculated values of μ , σ , and T_0 derived from the predicted data for Xi'an and Hong Kong. The T_0 value of Xi'an is 6, which is much smaller than t_0 . This is large because Xi'an implemented physical prevention and control measures earlier and increased detection of infected individuals, which reduced the risk of people being infected from the transmission route and significantly shortened the duration of the epidemic, thus leading to a smaller T_0 value. The T_0 values in both Shanghai and Hong Kong were greater than t_0 , which was due to the late adoption of preventive and control measures in both cities and the low surveillance and detection of the infected population, resulting in a long incubation period of the virus in the early stage.

Table 9. Urban epidemic curve delineation parameters.

City	μ	σ	T_0
Xi'an	21.63	5.35	6
Hong Kong	50.04	6.68	30

Figure 8 shows a comparison of the projected outbreak sites and actual locations for Shanghai, Xi'an, and Hong Kong. The officially announced outbreak site in Xi'an is the family compound of Chang'an University in Yanta District, and the outbreak site in Hong Kong is Kwai Chung Village in Kwai Tsing District. The distance between the estimated outbreak site and the actual outbreak site in Xi'an was 0.965 km, which is very close to each other, and the result is satisfactory. The main reason for this is that the spatial distribution data published in Xi'an was more detailed, including the main activity trajectories of patients in recent days, which could better reflect the degree of aggregation of patients in space. The aggregation of early infected patients was strongly correlated with the outbreak site. At the same time, strict control measures shortened the time of virus transmission in the population and reduced the interference of outlier sites due to crowd movement in estimating the location of outbreak sites.

The extrapolated results for Shanghai and Hong Kong are poorer compared to Xi'an, with distance values of 2.767 and 6.151 km between them and the actual outbreak location, respectively. The reasons for this situation in the two cities can be summarized in two ways: 1) the pressure of epidemic prevention in the two cities and the coarse information on the trajectories of the patients counted, which obscured the connection between the spatial location of the patients and the outbreak sites; 2) both cities did not implement strict precautionary measures, which allowed a large amount of time for the virus to spread in the population. Moreover, as cosmopolitan cities with dense crowds and high pedestrian flow, the virus spreads relatively fast in space, increasing the possibility of bias in the extrapolation results.

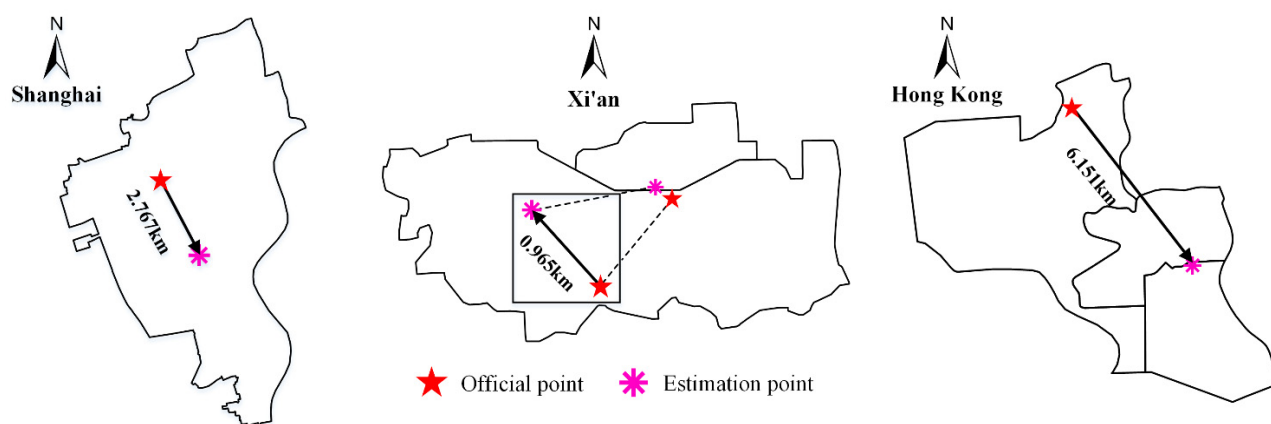


Figure 8. Projected outbreak point versus actual outbreak point for Shanghai, Xi'an and Hong Kong.

3.2. Simulation of spatial-temporal diffusion of infectious diseases

In the experiment, the Wuhan City Huainan Seafood Market was used as the initial outbreak point of the COVID-19 in Wuhan City, and the spatial-temporal diffusion process of the COVID-19 in Wuhan City from December 8, 2019 to April 5, 2020 was simulated, and the results are shown in Figure 10 and Table 10. To reflect the applicability of this diffusion model, this subsection also simulates the spatial-temporal diffusion process of COVID-19 in Shanghai and Hong Kong, and the results are shown in Figures 10 and 11, but mainly elaborates on the experimental results for Wuhan city.

Figure 9 illustrates the hotspot map of the spread of the COVID-19 in Wuhan at different time periods, and it can be seen that the COVID-19 has an obvious spatial aggregation. Table 10 gives the percentage of infected grids in Wuhan under different time periods, which correspond to each other in Figure 10. At the beginning of the outbreak, due to the lack of knowledge about the virus, a larger number of infected individuals were not counted, making the predicted number of infections less accurate, and the simulated infection grid was mainly concentrated near the outbreak site. With the spread of the virus in the population, scattered cases emerged in other areas of Wuhan, such as when the virus spread to day 20 (2019/12/27), infected areas appeared in and around the junction of Jiang'an, Jianghan, Qiaokou, and Wuchang districts. On day 40 (2020/1/16), infected patients also appeared in areas far from the outbreak site, such as Huangpi District, Xinzhou District and Hannan District, and the outbreak began to show a multi-point outbreak.

Due to the late intervention of human control measures, the virus spread among the population on a large scale and a “city closure” policy was adopted in Wuhan on day 47 (January 23, 2020). The strict “city closure” measures, while significantly reducing the movement of people and the spread of the virus, did not have an immediate and critical impact, as there were already large numbers of undetected infected people at the community level due to the prolonged spread of the virus. When the epidemic developed to day 50 (2020/01/26), the epidemic in Wuhan showed a comprehensive multi-point outbreak with increasingly obvious spatial spread effects, in which the number of infected grids surged to 1019, accounting for 16.02% of the total number of grids. By the 65th day of the outbreak development (2020/02/10), human prevention and control measures have come into play and the outbreak has reached an inflection point, with the number of new infections peaking daily and the spatial spread of the virus beginning to slow down, but the number of grids with serious infections is still increasing, and the mode of infection at this time is mainly community infection. On the 80th day of the epidemic development (2020/02/25), the number of grids with patients reached a maximum of 1936, accounting for 30.45% of the total number of grids. The epidemic was better controlled at this time, and the spatial aggregation effect of the epidemic became more obvious, showing a trend of gradual spread from urban areas to surrounding areas. Subsequently, under the effect of human prevention and control measures, the number of new infections per day in Wuhan continued to decrease, and the spread of the epidemic gradually leveled off, with the number of grids with patients not increasing and only the grids with more severe infections in some areas slowly increasing. Until the 120th day (2020/04/05), the epidemic was completely controlled in Wuhan, with 1936 total infected grids, accounting for 30.45% of the total grid data. Among them, 750 grids were at the level of primary infection (1~18), 661 grids were at the level of secondary infection (19~33), 379 grids were at the level of tertiary infection (34~50), and 146 grids were at the level of quaternary infection (51~76).

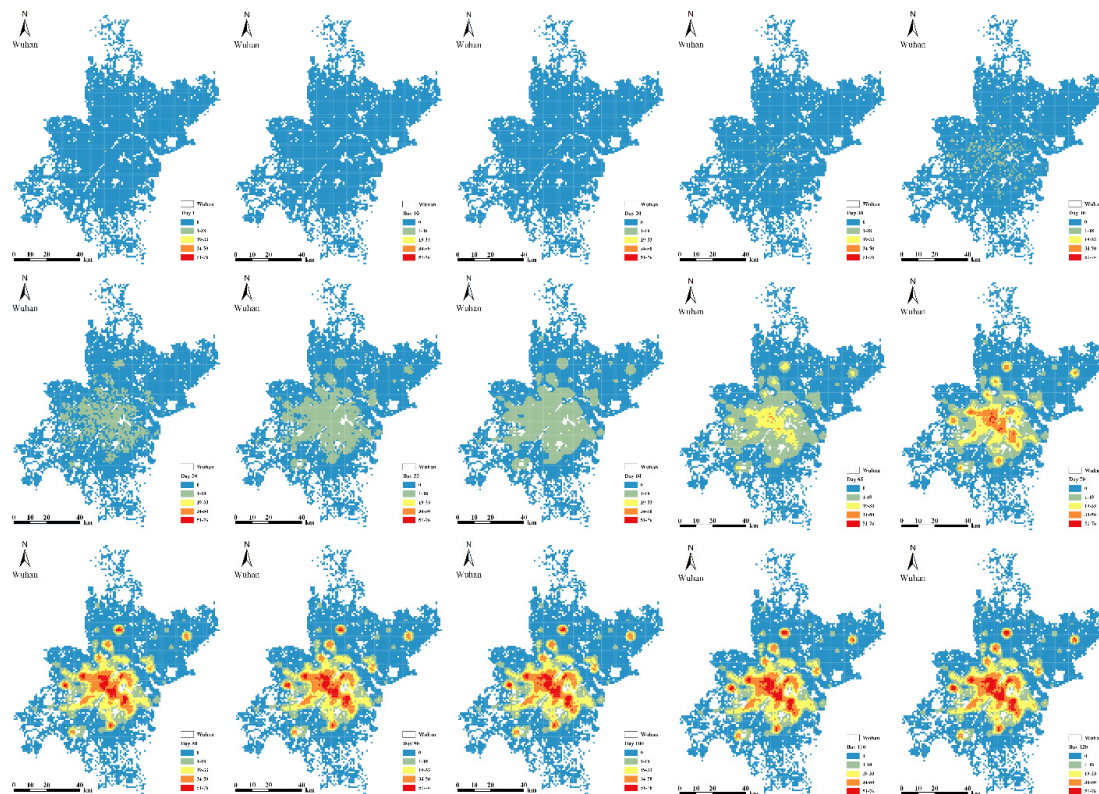


Figure 9. Spatial-temporal dispersal hotspot map of COVID-19 outbreak in Wuhan. COVID-19 in Wuhan was divided into three transmission stages (early, middle, and late) and five infection levels (0, 1~18, 19~33, 34~50, and 51~76) for presentation. Each row in the figure corresponds to a transmission stage, and considering the relatively small number of infections in the early stage of the epidemic and the strong randomness of the spread, and the relatively constant number of infected areas in the late stage, mainly the change of the number of infections in the infected areas, they are all displayed at an interval of 10 days; the number of infections in the middle stage is higher (73% of the total number of infections), and the change of the spread results is more obvious, so they are displayed at an interval of 5 days.

Table 10. Statistical table of grids with different levels of infection in Wuhan (total 6359).

Date	Infected Grid (%)	Grid of 1–18 patients (%)	Grid of 19–33 patients (%)	Grid of 34–50 patients (%)	Grid of 51–76 patients (%)
2019/12/08	1(0.02)	1(0.02)	0(0.00)	0(0.00)	0(0.00)
2019/12/17	1(0.02)	1(0.02)	0(0.00)	0(0.00)	0(0.00)
2019/12/27	9(0.14)	9(0.14)	0(0.00)	0(0.00)	0(0.00)
2020/01/06	59(0.93)	59(0.93)	0(0.00)	0(0.00)	0(0.00)
2020/01/16	279(4.39)	279(4.39)	0(0.00)	0(0.00)	0(0.00)
2020/01/26	1019(16.02)	1019(16.02)	0(0.00)	0(0.00)	0(0.00)
2020/01/31	1593(25.05)	1593(25.05)	0(0.00)	0(0.00)	0(0.00)
2020/02/05	1858(29.22)	1856(29.19)	2(0.03)	0(0.00)	0(0.00)

Continued on next page

Date	Infected Grid (%)	Grid of 1–18 patients (%)	Grid of 19–33 patients (%)	Grid of 34–50 patients (%)	Grid of 51–76 patients (%)
2020/02/15	1931(30.37)	1084(17.05)	589(9.26)	246(3.87)	12(0.19)
2020/02/25	1936(30.45)	803(12.63)	664(10.44)	372(5.85)	97(1.53)
2020/03/06	1936(30.45)	780(12.27)	651(10.24)	388(6.10)	117(1.84)
2020/03/16	1936(30.45)	768(12.08)	657(10.33)	385(6.05)	126(1.98)
2020/03/26	1936(30.45)	762(11.98)	656(10.32)	386(6.07)	132(2.08)
2020/04/05	1936(30.45)	750(11.79)	661(10.39)	379(5.96)	146(2.30)

Figure 10 shows the spatial and temporal spread of COVID-19 in Shanghai and Hong Kong. From the figure, it can be seen that the infected areas in both Shanghai and Hong Kong have obvious spatial aggregation, mainly concentrated in the central area of the city and gradually spreading to the surrounding areas. The more seriously infected areas in Shanghai are mainly in Jing'an, Huangpu, Changning and Xuhui districts, and are spreading to other urban areas. The more seriously infected areas in Hong Kong are mainly the Sham Shui Po district, Yau Tsim Mong district and Kowloon City district, but due to the special geographical conditions of Hong Kong, making its infected areas more scattered compared to Shanghai, the middle and late stages of the spread, the Tuen Mun district, Yuen Long district and North District also appeared infected areas one after another. As of the last day of the simulation, 27.86% of the total grids were infected in Shanghai and 34.89% of the total grids were infected in Hong Kong, which is close to the percentage of infected grids in Wuhan.

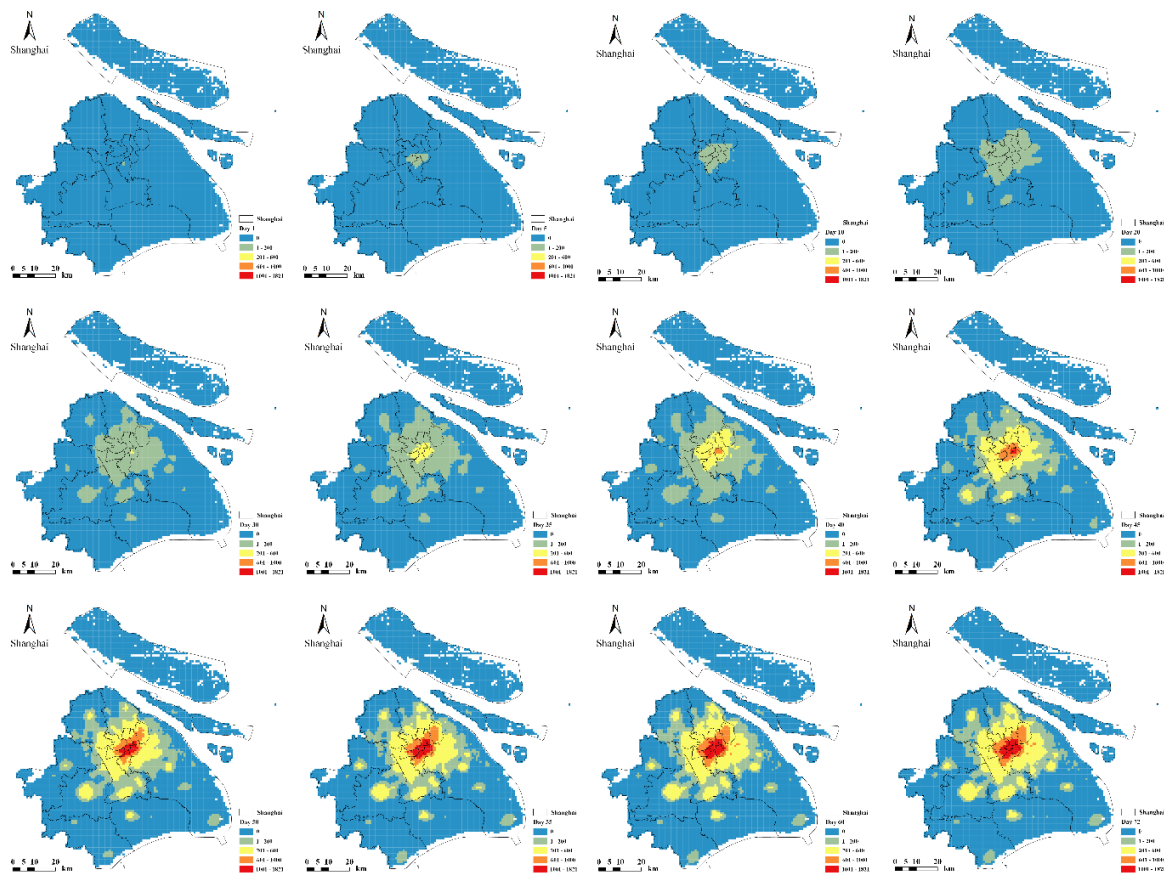


Figure 10. Spatial-temporal spread of the COVID-19 outbreak in Shanghai.

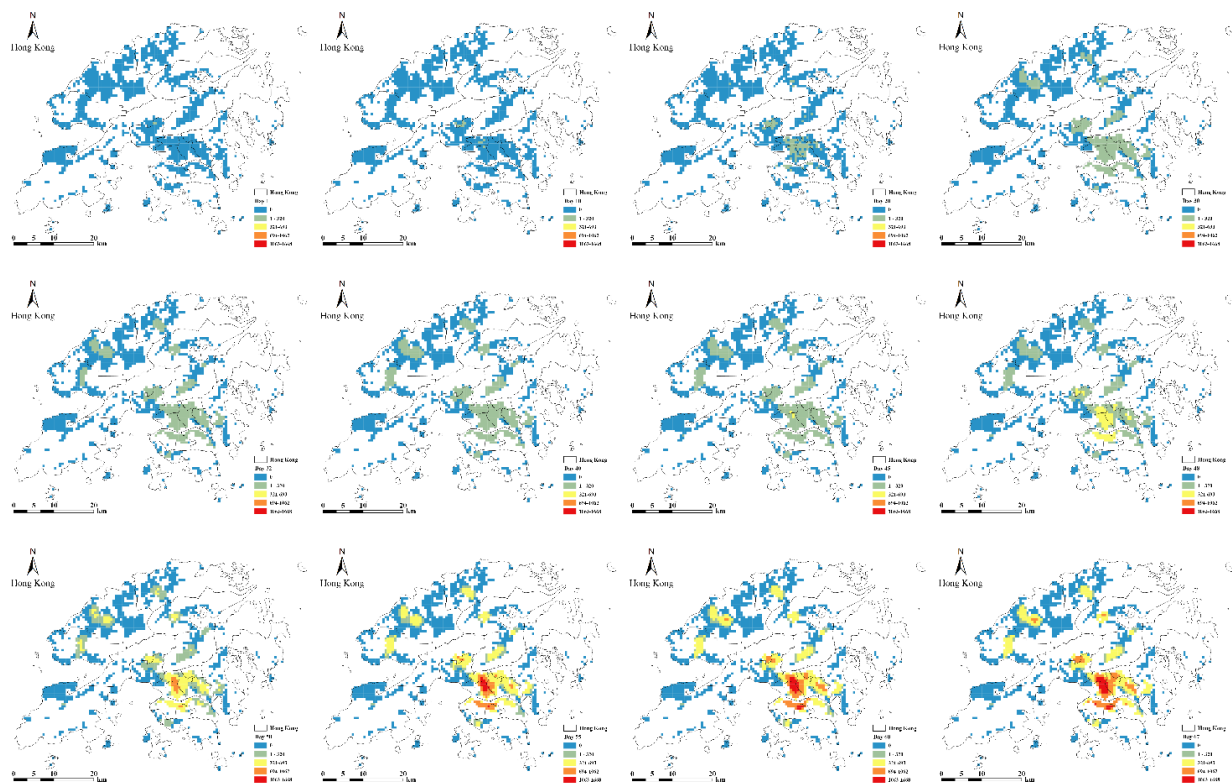


Figure 11. Spatial-temporal spread of the COVID-19 outbreak in Hong Kong.

3.3. Validation of model simulation results

To verify the soundness and validity of the spatial-temporal spread model of infectious diseases proposed in this study, the experimental results are validated in this section using the collected data on the spatial distribution of patients in Wuhan, Shanghai and Hong Kong. In this paper, we compare the validation results of Wuhan City with the results of Feng et al. [20]. To ensure the reasonableness of the comparison results, we keep the validation method (Estimated coverage of infected areas by actually infected communities) and validation area (Wuhan downtown area) of Wuhan city consistent with Feng et al's study, and the results are shown in Figure 12. Since Feng et al's validation method does not take into account the impact on the validation accuracy when the epidemic prediction results differ from the actual infection situation, i.e., when there is a rebound in the epidemic and the number of infected patients increases, there is a possibility that the estimated infection area is fully covered by patients. Therefore, in order to improve the rationality of the accuracy verification method, another accuracy verification method is designed in this paper. First, the collected spatial distribution data of patients are correlated with the simulation area in time and space. Second, according to the simulated daily new infection areas, the number of patients in these areas was counted. Finally, the proportion of these points to the total number of points for the day is calculated, and the proportion value is used as the validation accuracy value of the day's extrapolation results. This validation method was used to validate the accuracy of the epidemic simulation areas in Shanghai and Hong Kong, and the results are shown in Figure 13.

Figure 12 shows the accuracy verification of the spatial-temporal dispersion simulation results of this study's method compared with Feng et al's method for Wuhan City, from February 4, 2020 to

February 21, 2020. In terms of coverage accuracy, the method proposed in this study is better than Feng et al's method, and the average coverage accuracy of this method for infected cells in Wuhan is 91.03%, which is 18.31% higher than Feng et al's study (72.72%). Although Feng et al uses relatively fine-grained cell phone positioning data to reflect the status of population interactions in different spaces, it is limited by the total amount of data, which only accounts for 27% of the total population of Wuhan city, and is an estimation under small sample data. In terms of applicability, the simulation results of this paper are not limited by detailed observation data and can be simulated for areas outside the central city of Wuhan.

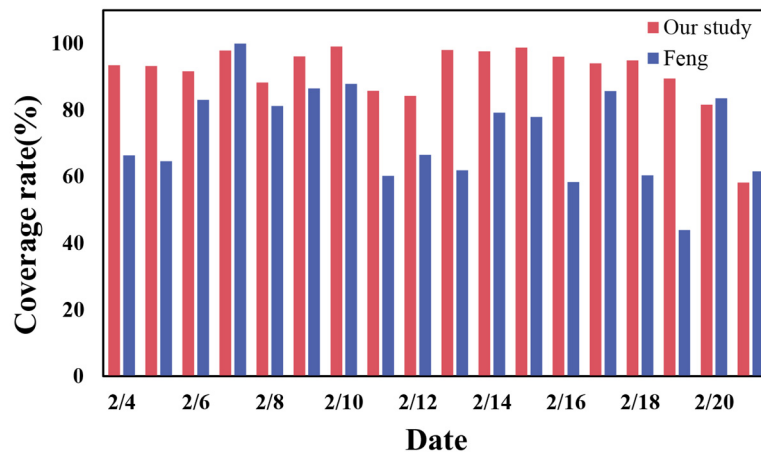


Figure 12. Comparative spatial and temporal diffusion results of COVID-19 in Wuhan.

Figure 13 (blue) shows the graph of the validation results of the spatial and temporal spread accuracy of the epidemic in Hong Kong from January 28, 2022 to March 21, 2022. Since the official data on the spatial distribution of patients in Hong Kong were published as a list of buildings visited by patients in the previous 14 days, the 14-day cumulative infected patients and areas were used in the precision validation of Hong Kong. However, this will inevitably have an impact on the overall accuracy of the model.

Figure 13 (red) shows the graph of the validation results of the spatial and temporal spread accuracy of the epidemic in Shanghai from March 6, 2022 to May 11, 2022, from which it can be seen that the validation accuracy in Shanghai shows a trend of first increasing and then decreasing. From the beginning of the epidemic to March 12, 2022, the average precision was only 13.9% and grew slowly, mainly because in the pre-epidemic period, there were fewer infected patients and the randomness of distribution was high, making it difficult to model the infected area at this time.

After the emergence of the epidemic, Shanghai did not take strict epidemic prevention and control measures, which led to a large number of infected people at the social level, while the model in this paper is based on the spatial correlation of population activities between regions, which is more suitable for simulating the transmission process of infectious diseases with a long transmission time. Therefore, after this, the simulation accuracy of the model for infected regions gradually increased, and by the end of March, the model accuracy grew to more than 70%.

As we entered April, Shanghai gradually imposed controls on various regions, but the physical prevention and control measures were delayed in taking effect because of the late intervention of the

controls. The virus transmitted in Shanghai belongs to the Omicron BA.2 mutant strain [36], which is highly insidious, making a large number of latent people in the population, and because of the high number of infected patients, the medical burden is too great, and the population can only be tested for nucleic acid centrally in separate communities, which undoubtedly does not increase the possibility of cross-infection between communities. Therefore, due to the above-mentioned situation, the role of physical prevention and control measures is greatly weakened, and in a way, the higher number of infected patients in April in Shanghai can still be considered as being infected in a state of free transmission of the virus, which is in line with the diffusion conditions suitable for our model. Therefore, the simulation accuracy of this model for April in Shanghai has been slowly increasing.

After May, the number of infected persons at the social level in Shanghai gradually decreased, and physical prevention and control measures had come into play. At this time, the newly infected persons were no longer infected in the state of free transmission of the virus but were mostly infected within families or communities, and the present model was not applicable to this infection pattern, thus leading to a sharp decrease in the accuracy of the model. Among them, the predicted data of new patients after April 22 differed significantly from the actual ones, which also contributed to the decrease in the model accuracy.

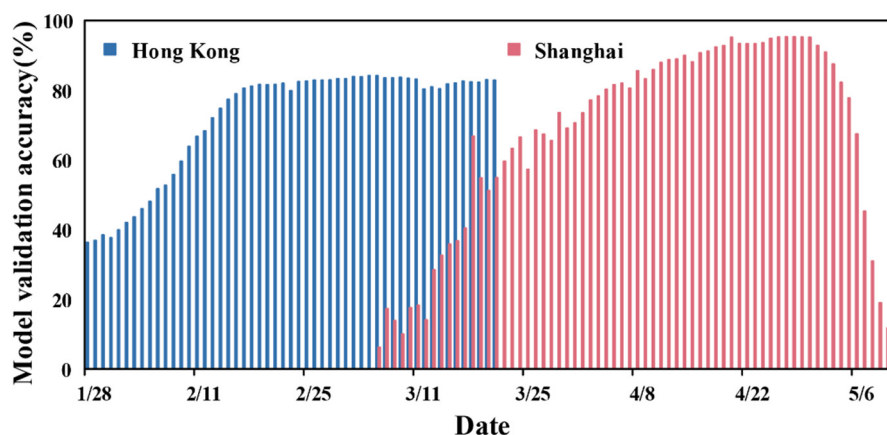


Figure 13. Accuracy of COVID-19 spatial-temporal dispersion results for Shanghai and Hong Kong (starting date of March 6, 2022 for Shanghai and January 28, 2022 for Hong Kong).

4. Discussion

The transmission process of infectious diseases has obvious spatial and temporal evolution patterns, which not only reflect the process of infected people recovering under the action of autoimmunity or external drugs, or withdrawing from transmission due to death from infection but also reflect the spatial interaction characteristics of susceptible and infected people and their spatial distribution information [37,38]. It is very important to master the transmission pattern of more infectious diseases, and simulate and reveal the spatial and temporal development trend of infectious diseases, which can identify the transmission pattern of infectious diseases and discover the high infection area.

4.1. Exploration of false information affecting outbreak point projection result

In this study, we propose an outbreak estimation method that takes into account the virus incubation period and population mobility. This method combines infectious disease dynamics models with modern spatial data analysis techniques, providing a novel and effective perspective for the formulation of public health and epidemic control strategies. However, any model and algorithm have certain limitations and uncertainties, whether constrained by data quality and integrity or influenced by the actual application environment. These factors may cause deviations in prediction results and increase the instability of the model. Therefore, in order to enhance the application value of the outbreak estimation method proposed in this paper in the future epidemic response, we deeply explore the false information that affects the estimation results of the method. Specifically, we will focus on analyzing the impacts possibly brought by four aspects: the prediction error of the prediction model, the difference in the virus incubation period, the influence of population mobility, and the quality of patient spatial distribution data.

In the outbreak point estimation method proposed in this paper, we chose the SEAIR model to predict the number of people infected with the virus and divided the number of infected people according to the 3sigma rule to determine the T_0 point. However, different accuracies of prediction results may lead to differences in the T_0 point, further affecting the accuracy of outbreak point estimation. In the experimental section of Subsection 3.1.1, we have already conducted a thorough analysis of this. The SEAIR model does indeed perform outstandingly in epidemiological prediction, with multiple parameters used in the model to simulate the actual transmission of infectious diseases. However, whether it's the SEAIR model or other infectious disease dynamic models, they all describe the complex virus transmission and evolution mechanism in the form of mathematical differential equations. This mathematical description method to some extent neglects the randomness and complexity in reality. For example, the prediction results of the model largely depend on the quality and representativeness of the training data. If the training data changes, such as an increase in training data, the prediction accuracy of the model will also change accordingly.

Secondly, in our method, we use the incubation period of the infectious disease as the time window for data selection. The size of the time window directly affects the range of selected data, thus influencing the spatial distribution pattern of the infected individuals, and further impacting the clustering results. In this study, we chose 14 days as the incubation period for COVID-19, aiming for a more conservative strategy when estimating the outbreak point. We subsequently performed experimental analyses of the outbreak points in different regions, which validated the reasonableness of our method. However, we must recognize that the incubation period of the virus is not constant. It may be affected by various factors, such as differences in strains, environmental changes, etc., which might cause variations in the incubation period of the virus in different regions. These differences could affect the predictive accuracy of our method when dealing with specific situations. Therefore, we need to consider this possibility when using this method and adjust the incubation period when necessary to adapt to the constantly changing real-world conditions.

On the other hand, population mobility also has an impact on the estimation results. Since the spread of the virus mainly depends on humans as carriers, and the transportation network of modern society is complex and efficient, the mobility of the population is greatly enhanced. This means that in the early stages of the outbreak, the spatial distribution of infected individuals might be quite scattered and may not easily exhibit significant aggregation. As the virus spreads further, the influence of

population mobility on the outbreak will become more complex and uncertain, which could lead to changes in the spatial aggregation of patients and consequently lead to a shift in the estimated location of the outbreak point. In our method, we assume that the epidemic curve follows a normal distribution, and the T_0 point is determined according to the 3sigma rule, which tries to minimize the impact of population mobility on the spread of the epidemic, in order to improve the accuracy of outbreak point estimation. However, we must clearly understand that the effects of population mobility are very complex and cannot be entirely eliminated through statistical methods.

Lastly, the quality of patient spatial distribution data directly influences the estimation results of the outbreak point. The spatial distribution data of patients reflect the life trajectory after virus infection, demonstrating the spatial distribution of patients. Through clustering analysis of this data, we can understand the spatial distribution pattern of the epidemic more clearly. However, the quality of the data will directly impact the accuracy of this analysis. If the data's accuracy and completeness are poor, such as the presence of a large number of false reports or omissions, our clustering results may have serious bias, leading to errors in the outbreak point estimation. In Subsection 3.1.2 of the paper, the estimated outbreak point in Xi'an was more accurate compared to the other two cities. This is largely attributed to the more detailed and accurate patient spatial distribution data in Xi'an compared to the other two cities.

In summary, there are numerous types of false information that can affect the estimation of outbreak points. Here, we have elucidated four main factors. Although our method may currently be influenced by these types of false information, we believe these issues can be resolved through continuous research and improvement. This will enhance the accuracy of outbreak point estimations, thereby providing better support for public health decision-making.

4.2. Delineation of outbreak prevention and control areas, with outbreak sites as the center

In the early days of COVID-19, due to the lack of scientific understanding of the virus, the entire city of Wuhan was designated as an epidemic area. This holistic approach to prevention and control was undoubtedly the most direct and effective, and was able to control the spread of the virus to a great extent. However, as knowledge of the virus increases, if a holistic prevention strategy is still adopted for cities with outbreaks, this will undoubtedly cause a more serious economic burden [39] and mental stress [40–42] to people.

In our study, the prevention and control distance was calculated using the standard deviation method based on the spatial location data of the patients. Then draw a circle with the outbreak point as the center and the prevention and control distance as the radius, and the area covered by the circle as the prevention and control area. Finally, the prevention and control areas were divided between Xi'an and Shanghai, respectively. Among them, the prevention and control radius of Xi'an is 5.5 km, and the prevention and control area is about 95.03 km²; the prevention and control radius of Shanghai is 20 km, and the prevention and control area is about 1256.64 km², and the results are shown in Figure 14. From the area of prevention and control, it can be seen that the pressure of prevention and control in Xi'an is significantly lower than that in Shanghai, which can also be reflected in the number of infected people in the two cities. It can be seen that, in the absence of an effective vaccine, earlier intervention in human control measures can limit the spread of the virus and reduce the harm caused by infectious diseases, which is consistent with the view of some studies [43,44].

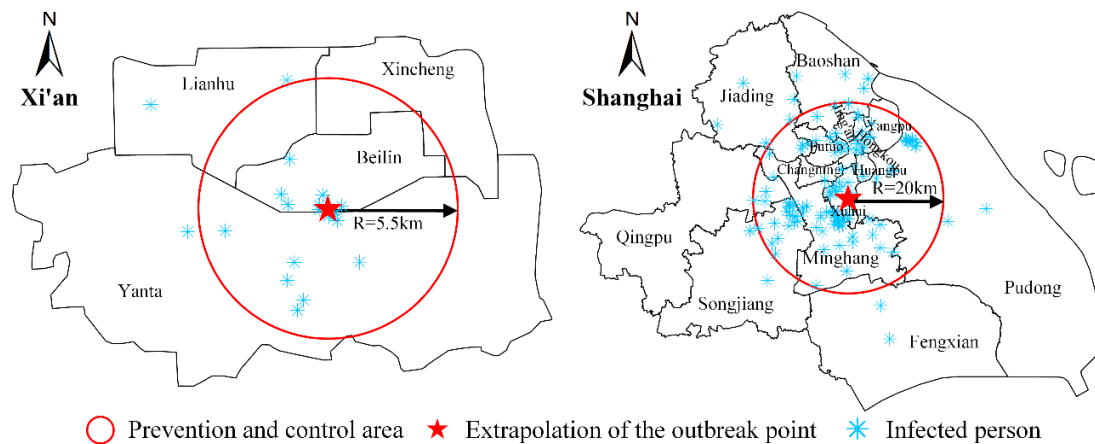


Figure 14. COVID-19 prevention and control area.

Nevertheless, this study has several limitations. First, the data collected on the spatial distribution of patients are not sufficiently granular; for example, Shanghai only published the places of residence of infected patients, hiding the impact of crowd movement on the location of outbreak sites. When an epidemic outbreak occurs, the relevant epidemic authorities have access to more accurate patient information that can solve the problem. Second, the location of the estimated outbreak point is likely to be located in a sparsely populated area, such as a woodland, field, etc., and then the accuracy of the location is questionable. Therefore, is it possible to add a grid weight constraint to correct the estimation results. Finally, the prevention and control area centered on the outbreak point may not completely cover the actual epidemic infection area. Whether to expand the prevention and control distance or to carry out prevention and control in multiple areas needs to be combined with the local epidemic situation.

4.3. Applicability of the spatial-temporal diffusion model

Our spatial-temporal diffusion model is based on the spatial correlation of population contact between different regions. When effective non-pharmaceutical interventions are applied to cities with outbreaks of infectious diseases, such as restricting crowd movements, the spatial interaction ability of crowds between sub-regions is greatly reduced. This means that the spatial aggregation effect of infectious diseases is smoothed out, and in the short-term diffusion process, the randomness of crowd travel activities is relatively large. In the absence of precise crowd travel data, it is difficult to describe the actual location of the patient's disease. Due to the incubation period of the virus, before the intervention takes effect, all patients can be considered to be infected during the free transmission phase of the virus. When the interventions take effect, the transmission routes of the virus are greatly reduced, and at this time, the newly infected patients in each sub-region are more often infected within the region, such as family infections and neighborhood infections. This also greatly increases the uncertainty of the spatial distribution of patients. Therefore, this model is more suitable for simulating the spatial-temporal diffusion process of infectious diseases that spread for at least one incubation period.

Although this model has achieved better simulation results in fine-grained granularity, it is still a generalization model from a macroscopic perspective, which is mainly reflected in the fact that the

diffusion pattern of the model is simulated according to the tree structure of people's daily life units, while the complex cell type and multi-core structure are more in line with the actual situation [34]. For a more realistic range of population activities, the incorporation of data, such as subway travel data and cab travel data, is still lacking. These data sources can further simulate the spatial patterns of preferential diffusion, convective diffusion, and jump diffusion of infectious diseases, and reduce the randomness of the diffusion scale.

5. Conclusions

In summary, this paper addresses the basic problem of infectious diseases by proposing a method to estimate the outbreak point considering the incubation period of the virus and population mobility. This helps address the uncertainty of the early outbreak point of infectious diseases. Additionally, a spatial and temporal diffusion model of infectious diseases under the discrete grid is established, taking into account the living habits of the population, to address the spatial diffusion process of infectious diseases.

The results show that the outbreak point estimation method can better estimate the area where the actual outbreak point is located, and the spatial-temporal diffusion model can also well simulate the infectious disease transmission process in a city. Although the current mortality rate caused by COVID-19 has decreased [45], every appearance of infectious diseases causes serious harm to human life, making scientific and efficient outbreak prevention and control measures essential.

Our research can significantly help governments and those working in epidemic prevention and control by providing insights into the location of infectious disease outbreak points and the process of transmission. This allows for the hierarchical deployment of control measures, while reducing the consumption of human and material resources, and ultimately stifling the spread of infectious diseases as early as possible.

Acknowledgments

This work was supported by The Excellent Youth Foundation of Henan Municipal Natural Science Foundation (212300410096), Program of Song Shan Laboratory (Included in the Management of Major Science and Technology Program of Henan Province) under Grant number 221100211000-03, and The National Key R&D Plan of China (2018YFB0505304).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Sawada, L. Sumulong, Macroeconomic impact of COVID-19 in developing Asia, Social Science Electronic Publishing, 2021. <http://dx.doi.org/10.2139/ssrn.3912360>
2. S. M. Kassa, J. B. H. Njagarah, Y. A. Terefe, Analysis of the mitigation strategies for COVID-19: from mathematical modelling perspective, *Chaos, Solitons Fractals*, **138** (2020), 109968. <https://doi.org/10.1016/j.chaos.2020.109968>

3. J. Riou, C. L. Althaus, Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020, *Eurosurveillance*, **25** (2020), 7–11. <https://doi.org/10.2807/1560-7917.Es.2020.25.4.2000058>
4. H. Nishiura, S. M. Jung, N. M. Linton, R. Kinoshita, Y. C. Yang, K. Hayashi, et al., The extent of transmission of novel coronavirus in Wuhan, China, 2020, *J. Clin. Med.*, **9** (2020). <https://doi.org/10.3390/jcm9020330>
5. J. T. Wu, K. Leung, G. M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *Lancet*, **395** (2020), 689–697. [https://doi.org/10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9)
6. Q. Li, X. H. Guan, P. Wu, X. Y. Wang, L. Zhou, Y. Q. Tong, et al., Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, *N. Engl. J. Med.*, **382** (2020), 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>
7. H. S. Lee, K. Kim, K. Choi, S. Hong, H. Son, S. Ryu, Incubation period of the coronavirus disease 2019 (COVID-19) in Busan, South Korea, *J. Infect. Chemother.*, **26** (2020), 1011–1013. <https://doi.org/10.1016/j.jiac.2020.06.018>
8. A. Viguerie, G. Lorenzo, F. Auricchio, D. Baroli, T. J. R. Hughes, A. Patton, et al., Simulating the spread of COVID-19 via a spatially-resolved susceptible-exposed-infected-recovered-deceased (SEIRD) model with heterogeneous diffusion, *Appl. Math. Lett.*, **111** (2021). <https://doi.org/10.1016/j.aml.2020.106617>
9. B. F. Maier, D. Brockmann, Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China, *Science*, **368** (2020), 742–746. <https://doi.org/10.1126/science.abb4557>
10. N. Crokidakis, COVID-19 spreading in Rio de Janeiro, Brazil: do the policies of social isolation really work? *Chaos, Soliton Fractals*, **136** (2020). <https://doi.org/10.1016/j.chaos.2020.109930>
11. M. H. D. Ribeiro, R. G. da Silva, V. C. Mariani, L. D. Coelho, Short-term forecasting COVID-19 cumulative confirmed cases: perspectives for Brazil, *Chaos, Solitons Fractals*, **135** (2020). <https://doi.org/10.1016/j.chaos.2020.109853>
12. A. J. Kucharski, T. W. Russell, C. Diamond, Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *Lancet Infect. Dis.*, **20** (2020), 553–558. [https://doi.org/10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
13. M. J. Keeling, K. T. D. Eames, Networks and epidemic models, *J. R. Soc. Interface*, **2** (2005), 295–307. <https://doi.org/10.1098/rsif.2005.0051>
14. N. A. Christakis, J. H. Fowler, Social network sensors for early detection of contagious outbreaks, *PLoS One*, **5** (2010). <https://doi.org/10.1371/journal.pone.0012948>
15. H. J. Li, W. Z. Xu, S. P. Song, W. X. Wang, M. Perc, The dynamics of epidemic spreading on signed networks, *Chaos, Solitons Fractals*, **151** (2021). <https://doi.org/10.1016/j.chaos.2021.111294>
16. Z. S. Wang, C. Y. Xia, Z. Q. Chen, G. R. Chen, Epidemic propagation with positive and negative preventive information in multiplex networks, *IEEE Trans. Cybern.*, **51** (2021), 1454–1462. <https://doi.org/10.1109/Tcyb.2019.2960605>
17. J. A. Firth, J. Hellewell, P. Klepac, S. Kissler, A. J. Kucharski, L. G. Spurgin, et al., Using a real-world network to model localized COVID-19 control strategies, *Nat. Med.*, **26** (2020), 1616–1622. <https://doi.org/10.1038/s41591-020-1036-8>

18. I. N. Lympieropoulos, #stayhome to contain COVID-19: neuro-SIR - neurodynamical epidemic modeling of infection patterns in social networks, *Expert Syst. Appl.*, **165** (2021). <https://doi.org/10.1016/j.eswa.2020.113970>
19. A. Bouchnita, A. Jebrane, A hybrid multi-scale model of COVID-19 transmission dynamics to assess the potential of non-pharmaceutical interventions, *Chaos, Solitons Fractals*, **138** (2020). <https://doi.org/10.1016/j.chaos.2020.109941>
20. M. Feng, Z. Fang, X. Lu, Z. Xie, S. Xiong, M. Zheng, et al., Traffic analysis zone-based epidemic estimation approach of COVID-19 based on mobile phone data: an example of Wuhan (in Chinese), *J. Wuhan Univ., Inf. Sci. Ed.*, **45** (2020), 651–657. <https://doi.org/10.13203/j.whugis20200141>
21. H. Y. Ren, L. Zhao, A. Zhang, L. Y. Song, Y. L. Liao, W. L. Lu, et al., Early forecasting of the potential risk zones of COVID-19 in China's megacities, *Sci. Total Environ.*, **729** (2020). <https://doi.org/10.1016/j.scitotenv.2020.138995>
22. J. Xia, Y. Zhou, Z. Li, L. Fan, Y. Yue, C. Tao, et al., COVID-19 risk assessment driven by urban spatiotemporal big data: a case study of Guangdong-Hong Kong-Macao Greater Bay Area, *Acta Geod. Cartographica Sin.*, **49** (2020), 671. <https://doi.org/10.11947/j.AGCS.2020.20200080>
23. M. M. Sugg, T. J. Spaulding, S. J. Lane, J. D. Runkle, S. R. Harden, A. Hege, et al., Mapping community-level determinants of COVID-19 transmission in nursing homes: a multi-scale approach, *Sci. Total Environ.*, **752** (2021). <https://doi.org/10.1016/j.scitotenv.2020.141946>
24. J. F. Ma, H. H. Zhu, P. Li, C. C. Liu, F. Li, Z. W. Luo, et al., Spatial patterns of the spread of COVID-19 in Singapore and the influencing factors, *ISPRS Int. J. Geo-Inf.*, **11** (2022). <https://doi.org/10.3390/ijgi11030152>
25. Y. Shi, D. Wang, Y. Chen, B. Chen, B. Zhao, M. Deng, An anomaly detection approach from spatio distributions of epidemic based on adjacency constraints in flow space, *Acta Geod. Cartographica Sin.*, **50** (2021), 777–788. <https://doi.org/10.11947/j.AGCS.2021.20200350>
26. X. Chen, J. Liu, L. Xu, J. Li, W. Zhang, H. Liu, Construction of the COVID-19 epidemic cases activity knowledge graph: a case study of Zhengzhou City (in Chinese), *J. Wuhan Univ., Inf. Sci. Ed.*, **45** (2020), 816–825. <https://doi.org/10.13203/j.whugis20200201>
27. B. Jiang, X. You, K. Li, X. Zhou, H. Wen, Interactive visual analysis of COVID-19 epidemic situation using geographic knowledge graph (in Chinese), *J. Wuhan Univ., Inf. Sci. Ed.*, **45** (2020), 836–845. <https://doi.org/10.13203/j.whugis20200153>
28. Q. Huang, Q. Liu, C. Song, X. Liu, H. Shu, X. Wang, et al., Urban spatial epidemic simulation model: a case study of the second COVID-19 outbreak in Beijing, China, *Trans. GIS*, **26** (2022), 297–316. <https://doi.org/10.1111/tgis.12850>
29. A. J. Tatem, Z. J. Huang, C. Narib, U. Kumar, D. Kandula, D. K. Pindolia, et al., Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning, *Malar. J.*, **13** (2014). <https://doi.org/10.1186/1475-2875-13-52>
30. A. Wesolowski, A. Winter, A. J. Tatem, T. Qureshi, K. Engo-Monsen, C. O. Buckee, et al., Measles outbreak risk in Pakistan: exploring the potential of combining vaccination coverage and incidence data with novel data-streams to strengthen control, *Epidemiol. Infect.*, **146** (2018), 1575–1583. <https://doi.org/10.1017/S0950268818001449>

31. J. R. Koo, A. R. Cook, M. Park, Y. X. H. Sun, H. Y. Sun, J. T. Lim, et al., Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study, *Lancet Infect. Dis.*, **20** (2020), 678–688. [https://doi.org/10.1016/S1473-3099\(20\)30162-6](https://doi.org/10.1016/S1473-3099(20)30162-6)
32. D. Okuonghae, A. Omame, Analysis of a mathematical model for COVID-19 population dynamics in Lagos, Nigeria, *Chaos, Solitons Fractals*, **139** (2020). <https://doi.org/10.1016/j.chaos.2020.110032>
33. B. Li, Y. Peng, H. He, M. Wang, T. Feng, Built environment and early infection of COVID-19 in urban districts: a case study of Huangzhou, *Sustainable Cities Soc.*, **66** (2021), 102685. <https://doi.org/10.1016/j.scs.2020.102685>
34. Q. Guo, *Research on the Public Service Facilities' Allocation of Basic Living Unit in Shenzhen*, Master's thesis, Harbin Institute of Technology, 2012. Available from: <http://kns.cnki.net.zzulib.vpn358.com/KCMS/detail/detail.aspx?dbname=CMFD201401&filename=1013038380.nh>.
35. Z. Cao, J. Wang, Y. Gao, W. Han, X. Feng, G. Zeng, Spatial pattern and heterogeneity risk of severe acute respiratory syndrome epidemic in Guangzhou, *Geogr. Res.*, **27** (2008), 1139–1149. <https://doi.org/10.11821/yj2008050017>
36. Xinjing News, More than 60 days in Shanghai against the epidemic: how a mega-city fought against Omicron [EB/OL] (in Chinese), 2022. Available from: <https://news.sina.com.cn/c/2022-05-01/doc-imcwiwst4954343.shtml>.
37. Z. Fang, Thinking and challenges of crowd dynamics observation from the perspectives of public health and public security (in Chinese), *J. Wuhan Univ., Inf. Sci. Ed.*, **45** (2020), 1847–1856. <https://doi.org/10.13203/j.whugis20200422>
38. D. Li, Z. Shao, W. Yu, X. Zhu, S. Zhou, Public epidemic prevention and control services based on big data of spatiotemporal location make cities more smart (in Chinese), *J. Wuhan Univ., Inf. Sci. Ed.*, **45** (2020), 475–487. <https://doi.org/10.13203/j.whugis20200145>
39. *World Trade Organization*, Trade set to plunge as COVID-19 pandemic upends global economy, 2020. Available from: https://www.wto.org/english/news_e/pres20_e/pr855_e.htm.
40. B. Pfefferbaum, C. S. North, Mental Health and the COVID-19 Pandemic, *N. Engl. J. Med.*, **383** (2020), 510–512. <https://doi.org/10.1056/NEJMp2008017>
41. J. Y. Qiu, B. Shen, M. Zhao, Z. Wang, B. Xie, Y. F. Xu, A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations, *Gen. Psychiatry*, **33** (2020). <https://doi.org/10.1136/gpsych-2020-100213>
42. Z. I. Santini, P. E. Jose, E. Y. Cornwell, A. Koyanagi, L. Nielsen, C. Hinrichsen, et al., Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis, *Lancet Public Health*, **5** (2020), E62–E70. [https://doi.org/10.1016/S2468-2667\(19\)30230-0](https://doi.org/10.1016/S2468-2667(19)30230-0)
43. A. Teslya, T. M. Pham, N. G. Godijk, M. E. Kretzschmar, M. C. J. Bootsma, G. Rozhnova, Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: a modelling study, *PLoS Med.*, **17** (2020). <https://doi.org/10.1371/journal.pmed.1003166>
44. W. Cao, H. Dai, X. Tong, F. Peng, C. Feng, Z. Wu, A model of artificial prevention and control measures for COVID-19 isolation and reception and cure based on discrete grids (in Chinese), *J. Wuhan Univ., Inf. Sci. Ed.*, **46** (2021), 167–176. <https://doi.org/10.13203/j.whugis20200343>

-
45. D. P. Peng, X. Y. Xing, Y. Wang, K. Y. Zhang, J. Q. Niu, Epidemiological and clinical characteristics of SARS-CoV-2 Omicron variant (in Chinese), *Chin. J. Viral Dis.*, **12** (2022), 385–389. <https://doi.org/10.16505/j.2095-0136.2022.0058>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)