



# Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors

Zhixiang Xie<sup>a,c</sup>, Yaochen Qin<sup>a,b,c,\*</sup>, Yang Li<sup>a,c</sup>, Wei Shen<sup>a</sup>, Zhicheng Zheng<sup>a</sup>, Shirui Liu<sup>a</sup>

<sup>a</sup> College of Environment and Planning, Henan University, Kaifeng 475004, China

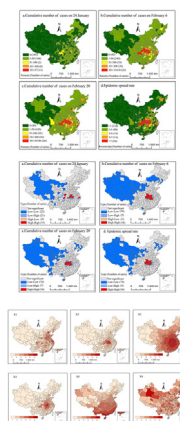
<sup>b</sup> Key Research Institute of Yellow River Civilization and Sustainable Development, Henan University, Kaifeng 475001, China

<sup>c</sup> Key Laboratory of Geospatial Technology for Middle and Lower Yellow River Region, Henan University, Kaifeng 475004, China

## HIGHLIGHTS

- Spatial and temporal differentiation of COVID-19 epidemic spread in China and its influencing factors are analyzed.
- The global and local spatial correlation characteristics of the epidemic distribution present a positive correlation.
- The population inflow from Wuhan and strength of economic connection are the main factors affecting the epidemic spread.
- The interaction influence of detection factors on the epidemic spread exceeds that of the single factor.
- When the average temperature in winter is maintained at 11–16 °C, the epidemic spread rate is higher.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

### Article history:

Received 13 May 2020

Received in revised form 4 July 2020

Accepted 11 July 2020

Available online 14 July 2020

Editor: Dr. Jay Gan

### Keywords:

COVID-19

Epidemic spread

Spatial-temporal differentiation

Spatial dependence

Geodetector

## ABSTRACT

This paper uses the exploratory spatial data analysis and the geodetector method to analyze the spatial and temporal differentiation characteristics and the influencing factors of the COVID-19 (corona virus disease 2019) epidemic spread in mainland China based on the cumulative confirmed cases, average temperature, and socio-economic data. The results show that: (1) the epidemic spread rapidly from January 24 to February 20, 2020, and the distribution of the epidemic areas tended to be stable over time. The epidemic spread rate in Hubei province, in its surrounding, and in some economically developed cities was higher, while that in western part of China and in remote areas of central and eastern China was lower. (2) The global and local spatial correlation characteristics of the epidemic distribution present a positive correlation. Specifically, the global spatial correlation characteristics experienced a change process from agglomeration to decentralization. The local spatial correlation characteristics were mainly composed of the 'high-high' and 'low-low' clustering types, and the situation of the contiguous layout was very significant. (3) The population inflow from Wuhan and the strength of economic connection were the main factors affecting the epidemic spread, together with the population distribution, transport accessibility, average temperature, and medical facilities, which affected the epidemic spread to varying degrees. (4) The detection factors interacted mainly through the mutual enhancement and nonlinear enhancement, and their influence on the epidemic spread rate exceeded that of single factors. Besides, each detection factor has an interval range that is conducive to the epidemic spread.

© 2020 Elsevier B.V. All rights reserved.

\* Corresponding author at: College of Environment and Planning, Henan University, Kaifeng 475004, China.

E-mail address: [qinyu@henu.edu.cn](mailto:qinyu@henu.edu.cn) (Y. Qin).

## 1. Introduction

On December 30, 2019, the Wuhan Municipal Health Commission issued an urgent notice on the treatment of pneumonia due to unknown causes, claiming that in the previous days some medical institutions have seen patients with unknown etiology, and requiring medical institutions to make statistics and report on the treatment on time (<http://wjw.wuhan.gov.cn/>). On January 8, 2020, the expert group of China Health Commission initially determined the cause of the outbreak as a novel coronavirus, which was identified by gene sequencing and named as COVID-19. This virus has a similar transmission route with the SARS (Severe Acute Respiratory Syndrome) virus, and may have been originally circulating in wild animals (Zhao et al., 2020). On January 12, the World Health Organization (WHO) officially named the novel coronavirus “2019-ncov” (2019 novel coronavirus), and estimated its incubation period to be about 2 to 10 days. Since then, the novel coronavirus gradually entered into the public view, and became the hot topic of the whole society. According to the novel coronavirus pneumonia protocol (4th edition) published by the National Health Commission of the People's Republic of China, the symptoms of COVID-19 include mainly fever, fatigue and dry cough, accompanied by nasal congestion, runny nose, and diarrhea in some patients. By the end of the March 13, a total of 116 countries or regions in the world appeared to be hit by the COVID-19 epidemic, and more than 130,000 people have been diagnosed. With the continuous spread of the COVID-19 epidemic, several countries or regions of the world have been forced to take emergency measures such as closing cities, stopping production, suspending school classes, and restricting population movement, causing great harm to economic development and residents' health (An and Jia, 2020). Therefore, it has become an urgent scientific problem to grasp the spatial and temporal changes of the COVID-19 epidemic spread, and clarify the driving mechanism.

Since the outbreak of COVID-19 epidemic, scholars have carried out abundant studies on the epidemic spread and achieved fruitful research results, which are of great guiding significance for the prevention and control of the epidemic. Joseph et al. (2020) estimated the size of the epidemic by using a mathematical model based on the data of confirmed cases of COVID-19 and residents' travel (including via trains, planes, and roads), and concluded that about 75,815 people were infected in the Wuhan city during the early outbreak stage of the epidemic. David et al. (2020) compared COVID-19 with other viruses, claiming that a sustained epidemic would pose a serious threat to global health, and proposing that the goal of sustainable development could be achieved by building a human-environment-animal health alliance. Liu et al. (2020a) used the exponential growth and maximum likelihood estimation method to determine the transmission dynamics of COVID-19 in Wuhan, and found that the average incubation period of the virus was 4.8 days, and the basic regeneration index reached 2.90 (95% Confidence Interval (CI): 2.32–3.63) and 2.92 (95% CI: 2.28–3.67). Ai et al. (2020) used the statistical analysis method to investigate the impact of lockdown measures in Wuhan (January 23, 2020) on the COVID-19 epidemic spread in other parts of China. They claim that if the closure measures were implemented 2 days in advance, it could have been possible to effectively prevent from being infected of 1420 people, if the city was closed 2 days later, there would have been 1462 more infections. Bai et al. (2020) used the transmission dynamics model to describe the evolution rule of the epidemic based on the data of confirmed COVID-19 cases in the Shaanxi province, revealed that the high incidence areas were mainly located in Xi'an, Ankang, and Hanzhong, and that the outbreak peak period was in early February 2020, with the basic regeneration index of the epidemic spread reaching 2.95. Wang et al. (2020a) used the Spearman correlation analysis method to find the relationship between the incidence of COVID-19 and the Baidu migration index in Guangdong province, and found that there was a positive correlation between the daily incidence and the 3-day migration index. Wang et al. (2020b) used the complex network model to explore

the impact of resuming work in surrounding cities on the epidemic situation in Hubei province on February 17, February 24, and March 2, and came to a conclusion that resuming work on March 2 would not cause a second outbreak of the epidemic. Yan et al. (2020) predicted the trend of the COVID-19 epidemic by building a time-delay dynamics model, and claimed that the epidemic could be controlled in the short period if the prevention and control efforts were kept unchanged. Chen and Cao (2020) made an epidemiological analysis of the daily confirmed cases in China, affirming that the situation of epidemic prevention and control in China was severe, and that targeted control measures should be formulated for the returning of enterprises and personnel in the future. Liu et al. (2020b) analyzed the spatial and temporal characteristics of the epidemic spread in Guangdong province, and found that the prevention and control measures adopted were effective, and high-risk areas were located in economically developed areas. Liu et al. (2020c) used the statistical analysis method to analyze the temporal and spatial characteristics and the transmission path of the COVID-19 epidemic in Zhuhai, revealing that the input from the epidemic area and family gatherings were the causes of epidemic spread. The research report published by the Yellow River Civilization and Sustainable Development Research Center of the Henan University presented detailed statistics on single case in Henan province, summarizing the epidemic spread into four modes, and providing control strategies (<http://skc.henu.edu.cn/info/1047/4673.htm>).

By combing the relevant literature, it was found that current studies on the COVID-19 epidemic spread have the following deficiencies: (1) The majority of existing studies explore the changes of the virus regeneration index from the perspectives of pathology, epidemiology, clinical medicine, molecular biology, and mathematics, so as to determine the epidemic stage of the disease and predict the development trend of the epidemic (Ai et al., 2020; Wang et al., 2020b). Although some scholars try to reveal the epidemic spread rules from a geographical perspective, they mainly focus on the spatial and temporal evolution characteristics of the epidemic, and seldom discuss the driving causes of the epidemic spread (Liu et al., 2020c). (2) In terms of research methods, current studies employ mainly the correlation analyses and regression analyses method, while the application of modern information technology and spatial analysis method are relatively limited (Wang et al., 2020a). (3) In terms of research scale, scholars generally investigate the epidemic spread characteristics at the city or regional scale, and there are few studies at the national level (Liu et al., 2020b; Liu et al., 2020b). (4) In relation to data sources, the data of COVID-19 cases can be obtained very easily; however, there are great difficulties in obtaining environmental and socio-economic data related to the epidemic spread, which is why current researches lag behind in the driving mechanism of epidemic spread. In this paper, the number of confirmed COVID-19 cases in mainland China was taken as the measurement index, and the spatial and temporal differentiation of the epidemic spread were described by the exploratory spatial data analysis method. Then, the key factors affecting the COVID-19 epidemic spread were identified by using the geodetector method, so as to provide references for clarifying the epidemic spread rule, formulating some protection policies, and promoting the resumption of work and production.

## 2. Data and methods

### 2.1. Data sources

The basic research objects of this paper are the administrative units at prefecture-level and above in mainland China, excluding Hong Kong, Macao, and Taiwan, and including Jiuyan, Tianmen, Xiantao, Qianjiang, the Shennongjia forest district, counties, and cities directly administered by the Hainan province and Xinjiang Uygur autonomous region. In order to maintain the integrity of the data, 366 units (excluding Sansha) were finally selected. The cumulative number of COVID-19 cases on January 24, February 6,

and February 20, 2020 comes from the epidemic announcements issued by each provincial (province-level municipality or autonomous region) health commission, while the first confirmation of COVID-19 in each region is taken from news reports. The spatial and temporal distance of each geographical unit arriving in Wuhan comes from the travel navigation service provided by Baidu map (<http://lbsyun.baidu.com>). Through the data interface, developers can calculate the route planning distance and driving time under different travel modes according to the starting and destination points. This paper took Wuhan as the starting point and other regions as the destination point to calculate the OD (Origination Destination) driving distance and time in self-driving mode. The proportion of inflow population from Wuhan is taken from the migration big data of Baidu map (<http://qianxi.baidu.com/>). These data allow to analyze the location information of mobile phone users through an LBS (Location Based Services) open platform, and to map their migration trajectory, so as to observe the population migration status. The acquisition period of Baidu migration data in the paper is from January 11 to 23, 2020, specifically referring to the top 100 cities toward which people move out of Wuhan every day. The winter average temperature data for each unit are from the weather network (<https://www.tianqi.com>). In addition, since it is impossible to obtain data on the population, gross domestic product, and number of beds in medical institutions for each region during the COVID-19 epidemic period, this paper employs the corresponding data in 2018, which is derived from the 2019 provincial statistical yearbooks or the 2018 statistical bulletins.

## 2.2. Research methods

### 2.2.1. Formula of epidemic spread rate

Using the cumulative number of COVID-19 cases as an indicator to measure the epidemic spread rate is biased due to the large differences in base population for different regions of mainland China. Therefore, the cumulative number of COVID-19 cases was divided by the number of days to calculate the epidemic spread rate, using the following formula:

$$V_i = \frac{S_i}{M - N_i} \quad (1)$$

where  $V_i$  represents the epidemic spread rate in region  $i$ ;  $S_i$  represents the cumulative number of COVID-19 cases in region  $i$  by February 20;  $M$  represents February 20; and  $N_i$  represents the date of the first confirmed case in region  $i$ .

### 2.2.2. Exploratory spatial data analysis method

The exploratory spatial data analysis method was used to verify whether the observed value of a unit has spatial correlation with the observed values of its neighboring units (Li et al., 2018). The global Moran's  $I$  index is used to measure the global spatial correlation, while the local Moran's  $I$  index in LISA (local indicators of spatial association) was used to measure the local spatial correlation (Rong et al., 2016). Their formulas (Anselin, 1995; Gallo and Ertur, 2003) are as follows:

$$\begin{cases} I = \frac{\sum_{i=1}^K \sum_{j=1}^K (X_i - \bar{X})(X_j - \bar{X})}{S^2 \sum_{i=1}^K \sum_{j=1}^K W_{ij}} \\ I^* = \sum_{p \neq q}^m W_{pq} Z_p Z_q \end{cases} \quad (2)$$

where  $I$  is the global Moran's  $I$  index;  $X_i$  and  $X_j$  are the observed values of unit  $i$  and  $j$ ;  $W_{ij}$  is the spatial weight matrix (with 1 as adjacent, and 0 as non-adjacent);  $S^2$  represents the variance;  $K$  represents the number of observation units;  $I^*$  is the local Moran's  $I$  index;  $W_{pq}$  is the normalized

form of the spatial weight matrix; and  $Z_p, Z_q$  are the normalized forms of the observed values in unit  $p$  and  $q$ .

### 2.2.3. Geodetector method

Spatial differentiation is a basic characteristic of geographical phenomena, the Geodetector method can measure the degree of spatial stratified heterogeneity and test its significance, through the within-strata variance less than the between strata variance (Xu et al., 2018). The Geodetector method comprises four modules: factor detection, interaction detection, risk detection and ecological detection. The factor detection is expressed by  $q$  value (Wang and Xu, 2017), its formulas are as follows:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST} \quad (3)$$

$$SSW = \sum_{h=1}^L N_h \sigma_h^2, SST = N \sigma^2 \quad (4)$$

where  $q$  represents the explanatory power of detect factor  $X$  on the spatial distribution of detected factor  $Y$ , the value of  $q$  ranges from 0 to 1;  $h = 1, \dots, L$ , which represents the stratification of the detect factor  $X$  and detected factor  $Y$ ;  $N_h$  and  $N$  are the number of samples for the layer  $h$  and the whole study area;  $\sigma_h^2$  and  $\sigma^2$  are the variance of  $Y$  value for the layer  $h$  and the whole study area;  $SSW$  and  $SST$  are the sum of intra-layer variances and the total variance of the whole study area.

The interaction detection can identify the explanatory power of the detect factors  $X1$  and  $X2$  to the detected factor  $Y$ , whose operation steps are as follows: first, we calculate the  $q$  values of  $X1$  and  $X2$ , respectively. Second, a new layer  $X1 \cap X2$  can be obtained by stacking the layer  $X1$  and  $X2$ , on this basis, the  $q(X1 \cap X2)$  value can be calculated. Third, the interaction type between  $X1$  and  $X2$  can be determined by comparing  $q(X1)$ ,  $q(X2)$ , and  $q(X1 \cap X2)$  values.

The risk detection is used to determine whether there exists a significant difference in the mean value of an attribute between the two sub-regions, which is tested by the  $t$ -statistic. Its formula is as follows (Yang et al., 2018):

$$t_{\bar{Y}_{h=1} - \bar{Y}_{h=2}} = \frac{\bar{Y}_{h=1} - \bar{Y}_{h=2}}{\left[ \frac{\text{Var}(\bar{Y}_{h=1})}{N_{h=1}} + \frac{\text{Var}(\bar{Y}_{h=2})}{N_{h=2}} \right]^{1/2}} \quad (5)$$

where  $\bar{Y}_h$  represents the average value of epidemic spread rate in the layer  $h$ ,  $N_h$  is the number of samples in the layer  $h$ ,  $\text{Var}$  represents the variance.

We can compare whether there are significant differences in the influence of any detect factors  $X1$  and  $X2$  on the spatial distribution of the detected factor  $Y$  by using the ecological detection, which is measured by the  $F$ -statistic.

$$F = \frac{N_{X_1}(N_{X_2} - 1) \times SSW_{X_1}}{N_{X_2}(N_{X_1} - 1) \times SSW_{X_2}} \quad (6)$$

$$SSW_{X_1} = \sum_{h=1}^{L_1} N_h \sigma_h^2 \quad (7)$$

$$SSW_{X_2} = \sum_{h=1}^{L_2} N_h \sigma_h^2 \quad (8)$$

where  $N_{X_1}$  and  $N_{X_2}$  represent the sample sizes of the detect factors  $X1$  and  $X2$ ;  $SSW_{X_1}$  and  $SSW_{X_2}$  are the sum of the variances in the layers formed by  $X_1$  and  $X_2$ ; and  $L_1$  and  $L_2$  are the number of layers of  $X_1$  and  $X_2$ . The null hypothesis  $H_0$  is:  $SSW_{X_1} = SSW_{X_2}$ . If  $H_0$  is rejected at the significance level of  $\alpha$ , which indicates that  $X1$  and  $X2$  have significantly different effects on the spatial distribution of  $Y$ .



### 3. Results and analysis

#### 3.1. Spatial and temporal differentiation of epidemic spread

##### 3.1.1. Spatial distribution characteristics

The ArcGIS software was used to classify the cumulative number of COVID-19 cases in the following categories: 0; 1–50; 51–100; 101–300; and >300 persons. The epidemic spread rate was classified into the following categories: <1; 1–3; 3–5; 5–7; and >7 persons/day (Fig. 1).

As can be seen from Fig. 1, the number of administrative units with a cumulative number of COVID-19 cases in the range of 0, 1–50, 51–100, 101–300 and >300 persons on January 24 were 197, 166, 2, 0, and 1, respectively. This indicated that there were 169 regional units hit by the COVID-19 epidemic by that date in mainland China, accounting for 46.17% of the total number of research units, and which reflected the wide spread of the epidemic. More into detail, the cities or regions with a cumulative number of confirmed cases in the range of 1–50 persons accounted for 98.22% of the total of regional units affected, and the number of cities with a cumulative number of confirmed cases of more than 50 persons was only 3, indicating that as of January 24, the epidemic was in the primary development stage. Several infected people could not be identified in a timely and effective way, because the COVID-19 has the incubation period. In terms of the cumulative number of confirmed cases in the epidemic areas, the largest number of

confirmed cases (572) was recorded in Wuhan, while other regions recorded a number of confirmed cases of only 1 person, indicating that the COVID-19 epidemic had a significantly different spread across the various regions of China. In terms of spatial distribution, the areas with a high number of confirmed cases were scattered in a few cities such as Wuhan, Chongqing, and Huanggang. On February 6, there were 45 regional administrative units with 0 cases, 263 with 1–50 cases, 24 with 51–100 cases, 18 with 101–300 cases, and 16 with more than 300 cases. This indicates that a total of 321 regional units in mainland China were by COVID-19 epidemic by that date, accounting for 87.70%, with an increase of 41.53% compared with January 24, thereby reflecting a significant expansion of the geographical range of the epidemic. Among the 321 regional units with an epidemic situation, the cities or regions with a cumulative number of confirmed cases in the range of 1–50 persons accounted for 81.93%, in the range of 51–100 persons accounted for 7.48%, in the range of 101–300 persons accounted for 5.61%, and in the range of more than 300 persons accounted for 4.98%. Compared with the number of regional units included in each interval on January 24, it was found that the proportion of cities in the other intervals showed an increasing trend for all intervals except for the 1–50 persons interval, indicating that the epidemic reached the outbreak stage. The cumulative number of confirmed cases was still the highest in Wuhan city, reaching 11,618, while the lowest numbers of confirmed cases in some areas were still 1, indicating that the difference across

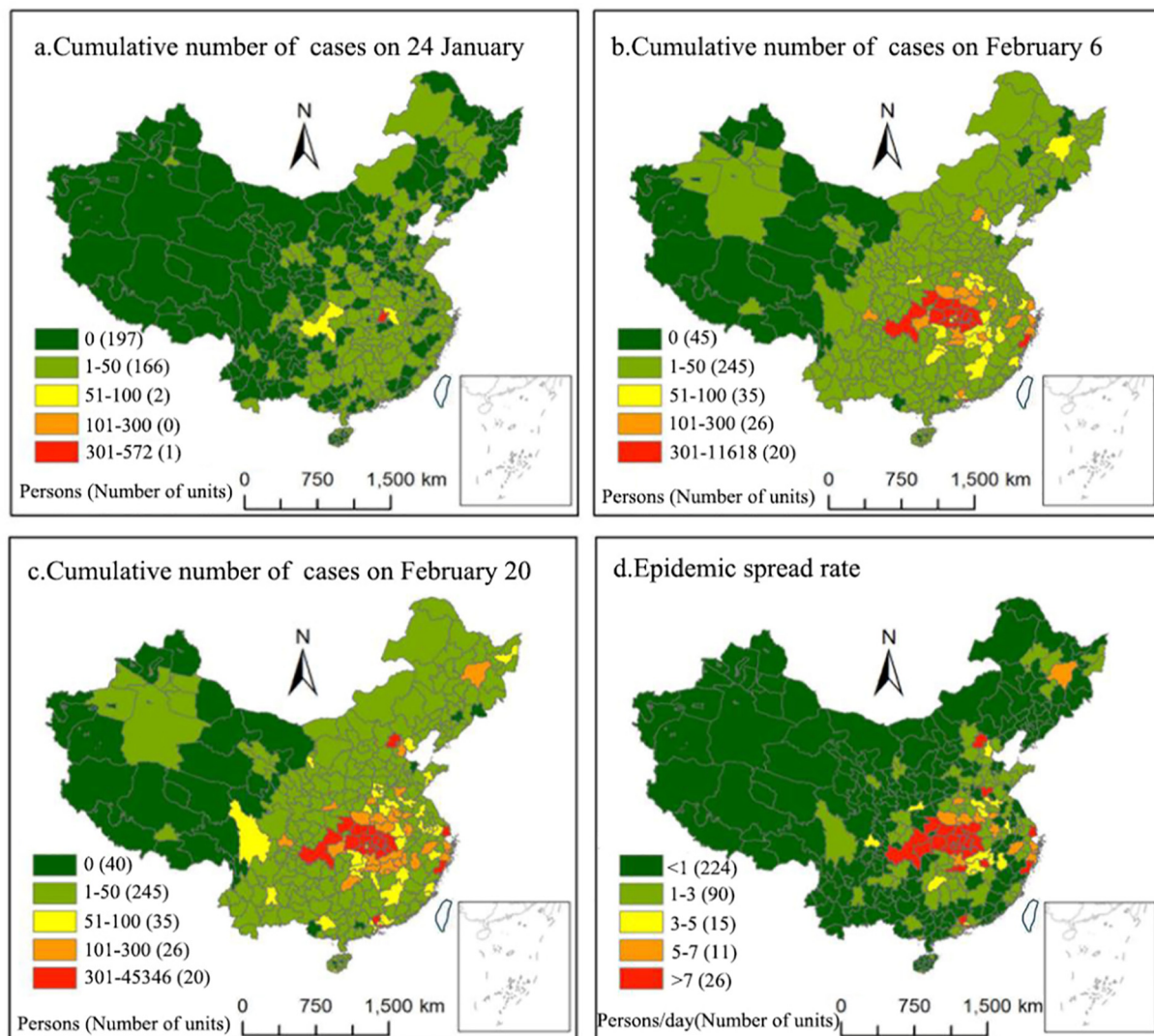


Fig. 1. Spatial distribution of the cumulative number of COVID-19 cases and epidemic spread rate.

regions was increasing. In terms of spatial distribution, there was an evident trend of continuous distribution in the areas with the highest cumulative number of confirmed patients, while the areas with a low number of patients continued to narrow, and were mainly located in some parts of Inner Mongolia, Qinghai, Tibet and Xinjiang Uygur. On February 20, the number of regions with a total cumulative number of COVID-19 cases in the range of 0, 1–50, 51–100, 101–300, and >300 persons were 40, 245, 35, 26 and 20, respectively, indicating that by that date there were 326 administrative units hit by the COVID-19 epidemic, accounting for 89.07% of the total number of research units. As such, there was an increase of only 1.37% compared with February 6, indicating a slowdown from previous increases, and implying that the diffusion trend of the epidemic spread has been initially contained. Among the 326 regional units with an epidemic situation, the cities or regions with a cumulative number of confirmed cases in the range of 1–50 persons accounted for 75.15%, in the range of 51–100 persons accounted for 10.74%, in the range of 101–300 persons accounted for 7.98%, and in the range of more than 300 persons accounted for 6.13%. The variation trend of the number of regional units in each interval was the same as that from January 24 to February 6; this means that the epidemic situation in a few cities has deteriorated rapidly during this period, while the areas affected by the epidemic were basically unchanged. In terms of spatial distribution, the coverage of the areas with a high cumulative number of confirmed cases continued to expand, being located mainly around Wuhan and economically developed cities, while the coverage of the areas with a low cumulative number of confirmed cases remained relatively stable. In terms of epidemic spread rate, there were 26 administrative units with an epidemic spread rate greater than 7 persons/day, 11 within the range of 5–7 persons/day, 15 within the range of 3–5 persons/day, and 90 within the range of 1–3 persons/day, while the remaining units had an epidemic spread rate lower than 1 person/day. In summary, the COVID-19 epidemic spread rate generally decreased toward the periphery with the city of Wuhan as the center, with high values in some economically developed cities, while the areas with lower values were located in the northwest, southwest, and northeast of China.

### 3.1.2. Spatial correlation characteristics

#### (1) Global spatial correlation characteristics

In this paper, the cumulative number of confirmed COVID-19 cases and the epidemic spread rate were taken as variables, the spatial weight matrix based on geographical adjacency was selected, and the global Moran's  $I$  index, the  $P$  test value and the  $Z$  statistic score of the cumulative number of confirmed COVID-19 cases and the epidemic spread rate were calculated by using the GeoDa software, so as to clarify the global spatial correlation characteristics (Table 1).

Table 1 shows that the global Moran's  $I$  indexes of the cumulative number of cases on January 24, February 6, and February 20 were 0.05, 0.18, and 0.08, respectively, passing the significance test at the 1% threshold level. Above values reflect the positive correlation, indicating that the spatial distribution of the cumulative number of confirmed COVID-19 cases in China at three time nodes shows an agglomeration trend. By observing the global Moran's  $I$  indexes at different time nodes, we can also find that this index first increased and then decreased. This indicates that the clustering development trend of the

cumulative number of confirmed COVID-19 cases increased from January 24 to February 6. The global Moran's  $I$  index decreased by 0.10 units on February 20 compared with February 6, meaning that the global spatial correlation during this period was still dominated by clustering characteristics, although the clustering degree was lower than before, and the spatial distribution of the cumulative number of confirmed COVID-19 cases showed a decentralized development trend. The global Moran's  $I$  index,  $P$  test value and  $Z$  statistic score of the epidemic spread rate were 0.15, 0.01, and 22.81 respectively, passing the significance test at the 1% level, implying that the spatial pattern of the epidemic spread rate was also characterized by a clustering distribution.

#### (2) Local spatial correlation characteristics

The Global Moran's  $I$  index has the defect of ignoring the instability of local spatial processes. Therefore, it is necessary to draw a LISA cluster map to analyze the local spatial correlation characteristics of COVID-19 epidemic (Fig. 2).

According to Fig. 2, the total number of administrative units belonged to the high-high, high-low, low-high and low-low cluster areas on January 24 were 11, 4, 21, and 22. Specifically, the 'high-high' cluster areas were located in Wuhan, Shiyang, Xiaogan, Jingzhou, Huanggang, Enshi, Xiantao, Xinyang, Anqing, Ningbo, and Shaoxing; the 'high-low' cluster areas were located in Shijiazhuang, Yangjiang, Sanya, and Wanning; the 'low-high' cluster areas were located in Langfang, Jinhua, Lu'an, Jiujiang, Huangshi, Ezhou, Xianning, Qianjiang, Tianmen, Shennongjia forest area, Xiangxi, Dongguan, Zhongshan, Luzhou, Suining, Neijiang, Guang'an, Dazhou, Ziyang, and Tongren; and the 'low-low' cluster areas were located in Hohhot, Ordos, Alxa, Ganzhi, Kunming, Lijiang, Linzhi, Nagqu, Yan'an, Jiuquan, Longnan, Guoluo, Yushu, Haixi, Bayingoleng, Aksu, Yili, Tacheng, Alar, Tiemenguan, and Kirkdala. On February 6, the number of administrative units in four clusters was 16, 1, 5 and 74. The 'high-high' cluster areas were located in Wuhan, Huangshi, Yichang, Xiangyang, Ezhou, Jingmen, Xiaogan, Jingzhou, Huanggang, Xianning, Suizhou, Xiantao, Tianmen, Xinyang, Nanyang, and Jiujiang; the 'high-low' cluster area was only located in Chengdu; and the 'low-high' cluster areas were located in Anqing, Lu'an, Qianjiang, Shennongjia forest region, and Changde. The 'low-low' cluster areas were mainly located in Inner Mongolia, Gansu, Ningxia, Qinghai, Tibet and Xinjiang. On February 20, the total number of administrative units belonged to the each of the cluster areas reached 14, 0, 7, and 76. The 'high-high' cluster areas were located in Wuhan, Huangshi, Yichang, Xiangyang, Ezhou, Jingmen, Xiaogan, Jingzhou, Huanggang, Xianning, Suizhou, Xiantao, Tianmen and Xinyang; the 'low-high' cluster areas were located in Anqing, Lu'an, Jiujiang, Nanyang, Qianjiang, Shennongjia forest region, and Changde; and the range of 'low-low' cluster was basically consistent with that of February 6. In terms of quantity change, the number of units classified in the 'high-high' cluster first increased and then decreased; the number of units included in the 'high-low' cluster continued to decrease until they disappeared; that in the 'low-high' cluster experienced a process of initial decline and then rose again; and the number of units belonged to the 'low-low' cluster showed an increasing trend. Therefore, it is not difficult to see that the layout trend of the cumulative number of confirmed COVID-19 cases at the time nodes has not changed fundamentally, and was dominated by the 'high-high' and 'low-low' type. This indicates that the local spatial correlation characteristics of the confirmed COVID-19 cases were also dominated by a positive correlation, although the clustering trend was weakened. Overall, the 'high-high' cluster areas showed a layout trend from centralization to decentralization, which tended to be stable over time, especially for Wuhan and its surrounding areas. There was a contiguous layout trend of the 'low-low' cluster areas, which were mainly located in Inner Mongolia, Gansu, Ningxia, Qinghai, Tibet, and Xinjiang. As for the epidemic spread rate, there were 16 administrative units in the 'high-high' cluster, 0 in the 'high-low' cluster, 5 in the 'low-high' cluster, and 75 in the 'low-

**Table 1**  
Global Moran's  $I$  index of the cumulative number of cases and epidemic spread rate.

Variable	January 24	February 6	February 20	Spread rate
Moran's $I$	0.05	0.18	0.08	0.15
$P$ value	<0.01	<0.01	<0.01	<0.01
$Z$ -Score	7.57	21.08	19.61	22.81



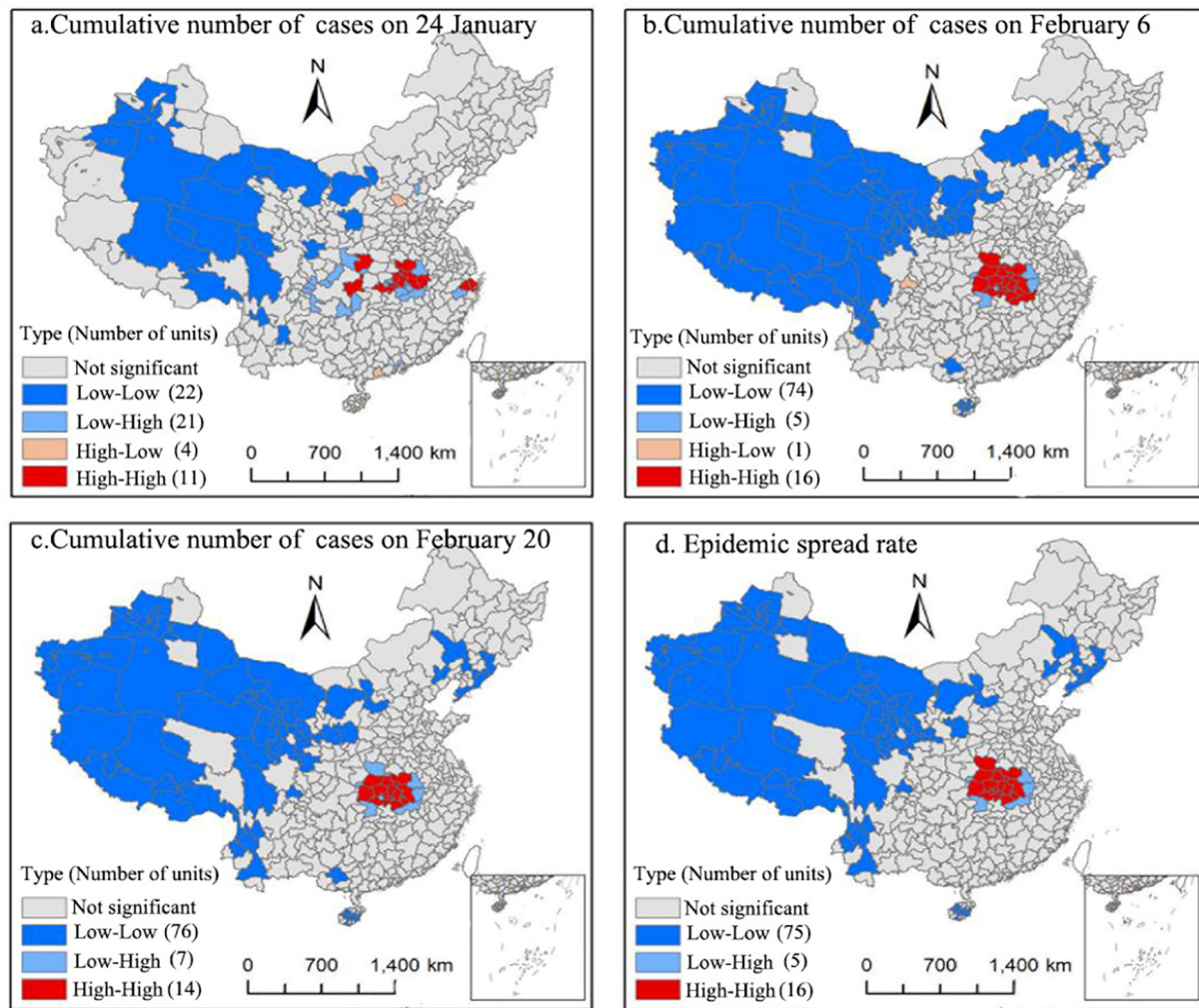


Fig. 2. Local spatial correlation characteristics of the cumulative number of cases and epidemic spread rate.

low' cluster. The high-high cluster areas were located in Wuhan, Huangshi, Yichang, Xiangyang, Ezhou, Jingmen, Xiaogan, Jingzhou, Huanggang, Xianning, Suizhou, Xiantao, Qianjiang, Tianmen, Nanyang, and Xinyang; the 'low-high' cluster areas were located in Anqing, Lu'an, Jiujiang, Shennongjia forest region, and Changde; and the 'low-low' cluster areas were located in western China.

### 3.2. Influencing factors of the epidemic spread

#### 3.2.1. Selection of evaluation indicators

The COVID-19 epidemic first occurred in Wuhan, and then spread to other parts of China. Therefore, people have been the carrier, the transportation network has been the channel, and the social and economic connections have been the internal driving force in the process of the epidemic spread. Thus, we selected the indicators reflecting the population distribution, population inflow from Wuhan, traffic accessibility, economic connection intensity, average temperature, and medical facilities conditions as the detection factors (Table 2), and the epidemic spread rate as the detected factor to assess the formation mechanism for the spatial pattern of COVID-19 epidemic.

Firstly, the classification method of natural discontinuities in ArcGIS10.2 software was used to divide detection factors into 6 categories, the classified maps of the detection factors were drawn (Fig. 3). According to formulas (3)–(8), the determination ability of detection

factors was calculated by using the geodetector software to analyze the influencing factors of epidemic spread.

#### 3.2.2. Analysis of the influencing factors

##### (1) Factor detection analysis

The  $q$  values of all the detection factors passed the significance test at the 5% level, indicating that these factors have a significant determination ability of the spatial distribution of the COVID-19 epidemic spread. Specifically, the  $q$  ( $p$ ) values of X1, X2, X3, X4, X5 and X6 were equal to 0.060 (0.003), 0.504 (0.000), 0.041 (0.000), 0.404 (0.000), 0.021 (0.028) and 0.078 (0.000), respectively. According to the size of  $q$  value, the inflow of population from Wuhan was the primary factor affecting the epidemic spread, and its explanatory power reached 50.4%. The economic connection intensity was the secondary determinant factor, and its explanatory power was 40.4%. The availability of medical facilities was the third determinant factor, which accounted for a 7.8% of explanation power. The determination ability of population distribution was 6%, while the traffic accessibility and average temperature were both relatively weak, below 5%. It is worth noting that the differentiation and factor detection analysis discussed only the determination ability of single factor on the epidemic spread rate, and did not consider the interaction effect of factors.

**Table 2**  
Detection indicator of the COVID-19 epidemic spread.

Factor code	Detection factor	Specific indicator	Treatment method
X1	Population distribution	Population density(persons/km <sup>2</sup> )	Total population/land area
X2	Population inflow from Wuhan	Proportion of incoming population in Wuhan (%)	Map migration data
X3	Traffic accessibility	Distance from Wuhan (km)	Determined by means of map navigation
X4	Economic connection intensity	Strength of economic connection with Wuhan	Using the gravity model
X5	Average temperature	Average temperature in winter (°C)	(mean maximum temperature in winter + mean minimum temperature in winter)/2
X6	Medical facilities conditions	Number of hospital beds per 1000 persons	Using statistical yearbooks or bulletins

Note: The gravity model was used to calculate the intensity of economic contact between each region and Wuhan, and the distance was the time reachable distance (Meng and Lu, 2011).

## (2) Interaction detection analysis

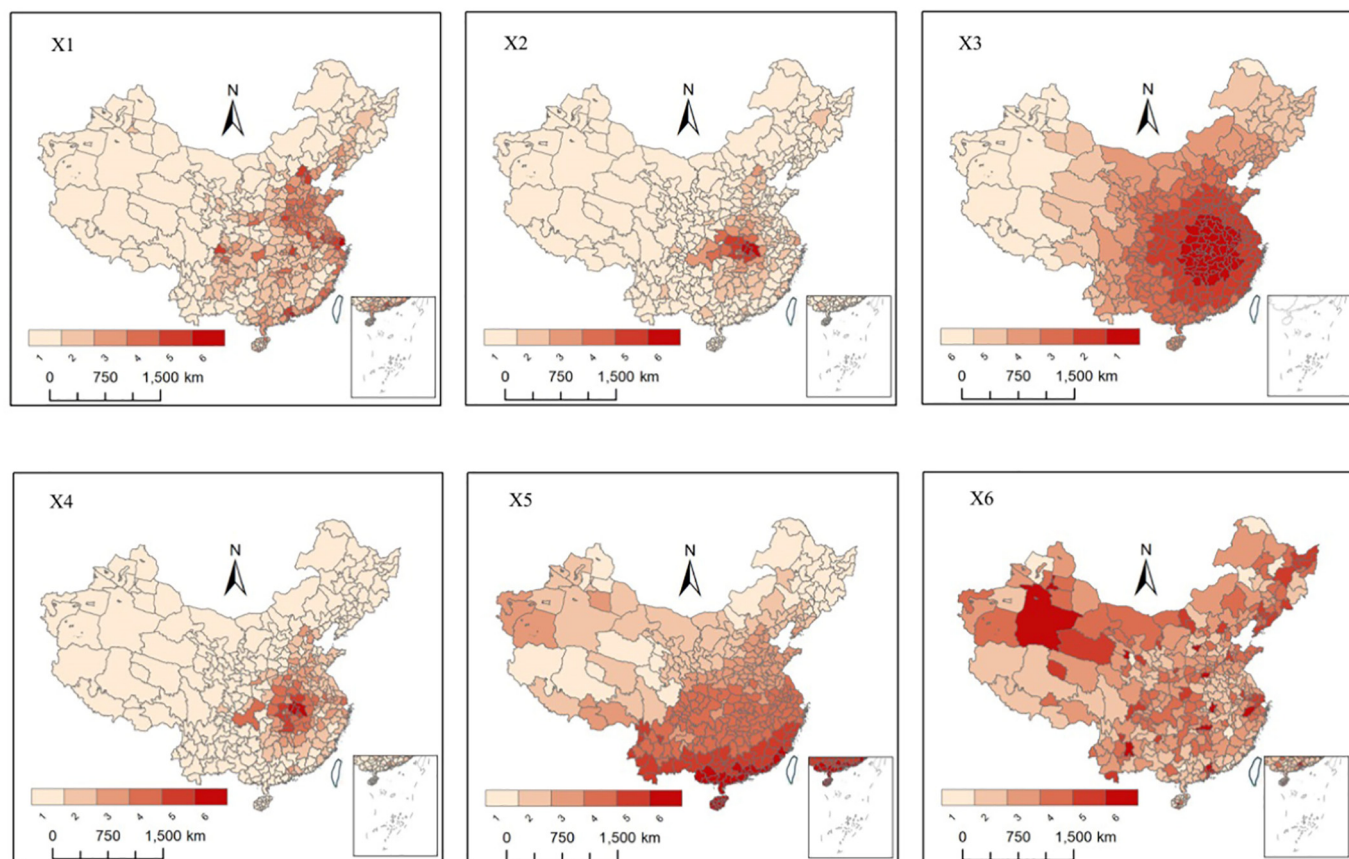
The interaction detector analysis is used to identify the interactions between any two factors. Table 3 shows the interactions detection results between factors.

It can be seen from Table 3 that the  $q$  values of  $X1 \cap X2$ ,  $X1 \cap X3$ ,  $X1 \cap X5$ ,  $X1 \cap X6$ ,  $X2 \cap X6$ ,  $X2 \cap X6$ ,  $X4 \cap X6$  and  $X5 \cap X6$  were greater than the  $q$  values of single factors, indicating that the interaction probe of detection factors had a significant impact on the epidemic spread rate. More detail, the population inflow from Wuhan and traffic accessibility, economic connection intensity, and average temperature interacted with each other, and produced mutually reinforcing effects. The traffic accessibility interacted with the economic connection intensity and average temperature to produce mutual reinforcing effects. The economic connection intensity interacted with the average temperature, which produced mutual reinforcing effects. The interaction between the

population distribution and the traffic accessibility, population inflow from Wuhan, economic connection intensity, average temperature and medical facilities conditions produced a nonlinear enhancement effect. The interaction between the population inflow from Wuhan and medical facility conditions produced a nonlinear enhancement effect. The interaction between the medical facilities conditions and the traffic accessibility, economic connection intensity and average temperature produced a nonlinear enhancement effect. In summary, the influence of detection factors on the epidemic spread rate was not independent, but showed a type of mutual or nonlinear enhancement.

## (3) Ecological detection analysis

According to Table 4, it could be found that the differences among the detection factors were statistically significant. Specifically, the influence of the population distribution (X1) on the spatial distribution of the



**Fig. 3.** Categorized spatial distribution of the detection factors.

**Table 3**  
Interactive detection results.

A∩B	A+B	Interaction probes	A∩B	A+B	Interaction probes
X1∩X2 = 0.998	>0.564 = X1 + X2	↑	X2∩X6 = 0.999	>0.581 = X2 + X6	↑
X1∩X3 = 0.199	>0.101 = X1 + X3	↑	X3∩X4 = 0.406	<0.445 = X3 + X4	↑↑
X1∩X4 = 0.993	<0.464 = X1 + X4	↑	X3∩X5 = 0.054	<0.061 = X3 + X5	↑↑
X1∩X5 = 0.124	>0.080 = X1 + X5	↑	X3∩X6 = 0.332	<0.524 = X3 + X6	↑
X1∩X6 = 0.329	>0.138 = X1 + X6	↑	X4∩X5 = 0.406	<0.424 = X4 + X5	↑↑
X2∩X3 = 0.505	<0.541 = X2 + X3	↑↑	X4∩X6 = 0.993	>0.482 = X4 + X6	↑
X2∩X4 = 0.506	<0.908 = X2 + X4	↑↑	X5∩X6 = 0.247	>0.098 = X5 + X6	↑
X2∩X5 = 0.505	<0.524 = X2 + X5	↑↑			

Note: “↑” means that factors A and B reinforce each other; “↑↑” indicates the nonlinear enhancement of factors A and B.

**Table 4**  
Ecological detector results (95% confidence level).

Factor code	X1	X2	X3	X4	X5	X6
X1						
X2	Y					
X3	N	Y				
X4	Y	Y	Y			
X5	Y	Y	N	Y		
X6	N	Y	N	Y	N	

Note: Y means the difference of the influence of the two factors is significant with the confidence of 95%, while N means no significant difference.

epidemic spread rate was significantly different from the population inflow from Wuhan (X2), economic connection intensity (X4), and average temperature (X5), but not different from the traffic accessibility (X3) and medical facility conditions (X6). The influence of the population inflow from Wuhan (X2) was significantly different from that of the traffic accessibility (X3), economic connection intensity (X4), average temperature (X5), and medical facilities conditions (X6). There was a significant difference between the influence of traffic accessibility (X3) and that of economic connection intensity (X4), but there were no significant difference with the average temperature (X5) and medical facilities conditions (X6). The influence of economic connection intensity (X4) was different from that of the average temperature (X5) and medical facility condition (X6). There was no significant difference between the average temperature (X5) and the medical facilities conditions (X6). Generally speaking, the detection factors selected in this paper are reasonable, and the differences among them are statistically significant.

#### (4) Risk detection analysis

Table 5 showed that the epidemic had the fastest spread rate when the population density was 1162–2564 persons/km<sup>2</sup>. When the proportion of population inflow from Wuhan was maintained at 6.94–14.25%, the epidemic spread rate was fastest. When the economic contact intensity with Wuhan was kept in the range of 598,158.64–1,524,023.05, the epidemic spread rate was fastest. When the geographical distance from Wuhan was 68.38–540.98 km, the spread rate was fastest. When the average temperature in winter was maintained at 11–16 °C, the epidemic spread rate was higher. The epidemic spread rate was higher when there were between 9.58 and 14.49 beds for 1000 persons. It can also be found that the population distribution, population inflow from Wuhan, economic connection intensity, medical facilities, and the epidemic spread

rate were significantly positively correlated, while the traffic accessibility was negatively correlated with the epidemic spread rate.

## 4. Discussions

This paper studied the spatial and temporal variation and the influencing factors of the COVID-19 epidemic spread in mainland China, which can provide references for formulating the public health policies and promoting the resumption of production. However, there exist the following problems. In terms of data sources, although many countries or regions have published the epidemic announcements of COVID-19 in real time, and the epidemic data was very convenient, virtually most of countries or regions had more people infected than registered, which could affect the accuracy of the evaluation results. Then, the population density data in 2018 was used as a replacement due to a fact that the population density for the each administrative unit in mainland China during the epidemic period was unavailable. The treatment method hid the drastic changes in the data because the COVID-19 epidemic happened during the Chinese Spring Festival period, which had a characteristic that the scale and frequency of population movements were intensified. It's worth noting that the population density is an important indicator to explain the epidemic spread rate, so the alternative data inevitably weakened the explanatory power of current research from this perspective. In addition, the factors affecting the epidemic spread were complex, and involved both the quantitative and the non-quantitative indicators. This paper constructed an indicator system of the multiple factors influencing the epidemic spread based on the principle of data availability; the other non-quantitative indicators might be ignored, which increased the uncertainty of evaluation results.

For the research method, the formula of epidemic spread rate was applicable to compare the epidemic spread rate of different administrative units at three time nodes, which actually did not conform to the exponential growth rule of infectious diseases (such as the COVID-19, SARS, and MERS (Middle East Respiratory Syndrome)) in the exposed population. How to accurately measure the actual spread rate of the epidemic in each region was the direction of future research. Second, the exploratory spatial data analysis method investigated the spatial clustering characteristics of COVID-19 epidemic in administrative units at prefectural level and above, and did not consider the agglomeration development situation at a finer spatial scale, which inevitably weakened the application value of the research results. Third, the geodetector method was adopted to obtain the most favorable range of the COVID-19 epidemic spread in this paper, which was developed from the perspective of statistics. The sample data directly affected the final

**Table 5**  
Most favorable range of epidemic spread (95% confidence level).

Factor code	Favorable range	Factor code	Favorable range
X1	1162–2564(persons/km <sup>2</sup> )	X4	598,158.64–1,524,023.05
X2	6.94–14.25(%)	X5	11–16(°C)
X3	68.38–540.98(km)	X6	9.58–14.49(beds/1000 persons)



evaluation result, and no epidemiological investigation on the residents' health status was implemented, so the conclusions drawn from current research were uncertain to some extent. Finally, there might have multicollinearity between the strength of economic connection economic and other factors in this paper, and the geodetector method was not used to deal with it, which would weaken the persuasiveness of the research results.

## 5. Conclusions

- (1) The temporal changes of the COVID-19 epidemic in mainland China are clear, and the epidemic spread rate has an evident spatial variation. In terms of temporal change, the epidemic quickly spread to most regions from January 24 to February 6. The epidemic spread rate slowed down from February 6 to February 20, although the epidemic situation in some cities worsened sharply. The areas where the epidemic spread quickly were mainly located in the Hubei province, its surrounding areas, and some economically developed cities. The western part of China, as well as the remote areas of central and eastern China experienced a slow epidemic spread.
- (2) The global and local spatial correlation characteristics of the COVID-19 epidemic were dominated by clustering situations. Specifically, the global spatial correlation characteristics initially increased and then decreased, while the local spatial correlation characteristics tended to be stable with the passage of time, and were mainly composed of the 'high-high' and 'low-low' cluster types. The 'high-high' cluster areas were located in Wuhan, Huangshi, Yichang, Xiangyang, Ezhou, Jingmen, Xiaogan, Jingzhou, Huanggang, Xianning, Suizhou, Xiantao, Qianjiang, Tianmen, Nanyang and Xinyang. The 'low-low' cluster areas were located in parts of Inner Mongolia, Gansu, Ningxia, Qinghai, Tibet, and Xinjiang.
- (3) The population distribution, population inflow from Wuhan, traffic accessibility, economic connection intensity, average temperature and medical facilities conditions had significant effects on the epidemic spread rate. The population inflow from Wuhan was the primary factor affecting the epidemic spread, followed by the economic connection intensity, and the medical facilities conditions. The population distribution, traffic accessibility, and average temperature also had different degrees of influence on the epidemic spread. From the perspective of action direction, the population distribution, population inflow from Wuhan, economic connection intensity and medical facilities conditions played a positive role in the process of epidemic spread, while the traffic accessibility played a negative role.
- (4) Detection factors interacted through mutual enhancement and nonlinear enhancement, and their influence on the epidemic spread rate exceeded that of single detection factors. The interaction between the population inflow from Wuhan and medical facilities conditions, as well as that between the population distribution and population inflow from Wuhan, that between the population distribution and economic connection intensity, and that between the economic connection intensity and medical facilities conditions had a great influence on the epidemic spread. The interaction between the population distribution and traffic accessibility, as well as that between the population distribution and average temperature, that between the traffic accessibility and average temperature, and that between the average temperature and medical facilities conditions had little impact on the epidemic spread.

## CRediT authorship contribution statement

**Zhixiang Xie:** Writing - original draft. **Yaochen Qin:** Funding acquisition, Writing - review & editing. **Yang Li:** Visualization. **Wei Shen:** Methodology. **Zhicheng Zheng:** Data curation. **Shirui Liu:** Data curation.

## Declaration of competing interest

All the authors do not have any possible conflicts of interest.

## Acknowledgments

This paper is funded by the Natural Science Foundation of China (Grant No. 41671536), the Key Scientific Research Projects of Universities in Henan province (Grant No. 18A170002), and the Collaborative Projects of Henan Key Laboratory of Integrative Air Pollution Prevention and Ecological Security, China (Grant No. PAP201801).

## References

- Ai, S., Zhu, G., Tian, F., Li, H., Gao, Y., Wu, Y., Liu, Q., Lin, H., 2020. Population Movement, City Closure and Spatial Transmission of the 2019-nCoV Infection in China. *MedRxiv*. <https://doi.org/10.1101/2020.02.04.20020339>.
- An, G., Jia, F., 2020. Analysis of the economic impact of the NCP and countermeasure study. *Financ. Theor. Pract.* (3), 45–51.
- Anselin, L., 1995. Local indicators of spatial association-LISA. *Geogr. Anal.* 27 (2), 93–115.
- Bai, Y., Liu, K., Chen, Z., Chen, B., Shao, Z., 2020. Early transmission dynamics of novel coronavirus pneumonia epidemic in Shaanxi province. *Chin. J. Nosocomiology*. 30 (6), 834–838.
- Chen, Y., Cao, G., 2020. Incidence trend of novel coronavirus (SARS-CoV-2)-infected pneumonia in China Shanghai. *J. Prevent. Med.* 32 (2), 1–6.
- David, S.H., Esam, I.A., Tariq, A.M., Francine, N., Eskild, P., 2020. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* 91, 264–266.
- Gallo, J.L., Ertur, C., 2003. Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980–1995. *Pap. Reg. Sci.* 82 175–121.
- Joseph, T.W., Kathy, L., Gabriel, M.L., 2020. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. *Lancet* 395 (10225), 689–697.
- Li, C., Wang, X., Xie, Z., Qin, M., 2018. Evolution of patterns in the ratio of gender at birth in Henan province, China. *Probl. Ekorozw.* 13 (1), 59–67.
- Liu, T., Hu, J., Kang, M., Lin, L., Zhong, H., Xiao, J., He, G., Song, T., Huang, Q., Rong, Z., Deng, A., Zeng, W., Tan, X., Zeng, S., Zhu, Z., Li, J., Wan, D., Lu, J., Deng, H., He, J., Ma, W., 2020a. Transmission Dynamics of 2019 Novel Coronavirus (2019-nCoV). *BioRxiv*. <https://doi.org/10.1101/2020.01.25.919787>.
- Liu, Y., Li, Y., Li, Z., Han, F., 2020b. The diffusion characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in Guangdong province. *Trop. Geogr.* 40 (3), 367–374.
- Liu, Z., Ye, Y., Zhang, H., Guo, H., Yang, J., Wang, C., 2020c. Analysis of the spatio-temporal characteristics and transmission path of COVID-19 cluster cases in Zhuhai. *Trop. Geogr.* 40 (3), 422–431.
- Meng, D., Lu, Y., 2011. Impact of high-speed railway on accessibility and economic linkage of cities along the railway in Henan province, China. *Sci. Geogr. Sin.* 31 (5), 537–543.
- Rong, P., Zhang, L., Yang, Q., Qin, X., Qin, Y., Lu, H., 2016. Spatial differentiation patterns of carbon emissions from residential energy consumption in small and medium-sized cities: a case study of Kaifeng. *Geogr. Res.* 35 (8), 1495–1509.
- Wang, J., Xu, C., 2017. Geodetector: principle and prospective. *Acta Geograph. Sin.* 72 (1), 116–134.
- Wang, X., Liao, C., Li, Z., Hu, H., Cheng, X., Li, Q., Lu, J., 2020a. Preliminary analysis on the early epidemic and spatiotemporal distribution of new coronavirus pneumonia in Guangdong province. *J. Trop. Med.* 20 (4), 427–430.
- Wang, X., Tang, S., Chen, Y., Feng, X., Xiao, Y., Xu, Z., 2020b. When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemic? A data-driven network modeling analysis. *Sci. Sin. Math.* <https://doi.org/10.1360/SSM-2020-0037>.
- Xu, Q., Dong, Y., Wang, Y., Yang, R., Xu, C., 2018. Determinants and identification of the northern boundary of China's tropical zone. *J. Geogr. Sci.* 28 (1), 31–45.
- Yan, Y., Chen, Y., Liu, K., Luo, X., Xu, B., Jiang, Y., Cheng, J., 2020. Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system. *Sci. Sin. Math.* 50 (3), 385–392.
- Yang, D., Wang, X., Xu, J., Xu, C., Lu, D., Ye, C., Wang, Z., Bai, L., 2018. Quantifying the influence of nature and socioeconomic factors and their interactive impact on PM<sub>2.5</sub> pollution in China. *Environ. Pollut.* 241, 475–483.
- Zhao, X., Li, X., Nie, C., 2020. Backtracking transmission of COVID-19 in China based on big data source, and effect of strict pandemic control policy. *Bull. Chin. Acad. Sci.* 35 (3), 248–255.