Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

International Journal of Infectious Diseases 105 (2021) 675–685

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Temporal and spatial analysis of COVID-19 transmission in China and its influencing factors

Qian Wang[a,c], Wen Dong[b,c,*], Kun Yang[b,c,*], Zhongda Ren[a,c], Dongqing Huang[a,c], Peng Zhang[a,c], Jie Wang[a,c]

[a] School of Information Science and Technology, Yunnan Normal University, Kunming, 650500, China
[b] Faculty Of Geography, Yunnan Normal University, Kunming, 650500, China
[c] GIS Technology Engineering Research Centre for West-China Resources and Environment of Educational Ministry, Yunnan Normal University, Kunming, 650500, China

ARTICLE INFO

Article history:
Received 14 December 2020
Received in revised form 26 February 2021
Accepted 4 March 2021

Keywords:
COVID-19
Spatio-temporal
Migration index
Environment temperature
Air pollution concentration
Government response strictness index

ABSTRACT

Objectives: The purpose of this study was to explore the temporal and spatial characteristics of COVID-19 transmission and its influencing factors in China, from January to October 2020.
Methods: About 81,000 COVID-19 confirmed case data, Baidu migration index data, air pollutants, meteorological data, and government response strictness index data were collected from 31 provincial-level regions (excluding Hong Kong, Macao, and Taiwan) and 337 prefecture-level cities. The spatio-temporal characteristics of COVID-19 were explored using spatial autocorrelation, hot spot, and spatio-temporal scanning statistics. At the same time, Spearman rank correlation analysis and multiple linear regression were used to explore the relationship between influencing factors and confirmed COVID-19 cases.
Results: The distribution of COVID-19 in China tends to be stable over time, with spatial correlation and prominent clustering regions. Spatio-temporal scanning analysis showed that most COVID-19 high-incidence months were from January to March at the beginning of the epidemic, and the area with the highest aggregation risk was Hubei Province (RR = 491.57) which was 491.57 times the aggregation risk of other regions. Among the meteorological variables, the daily average temperature, wind speed, precipitation, and new COVID-19 cases were negatively correlated. The air pollution concentration and migration index were positively correlated with new confirmed cases, and the government response strict index was strongly negatively correlated with confirmed COVID-19 cases.
Conclusions: Environmental temperature has a certain inhibitory effect on the transmission of COVID-19; the air pollution concentration and migration index have a certain promoting effect on the transmission of COVID-19. The strict government response index indicates that the greater the intensity of government intervention, the fewer COVID-19 cases will occur.

Introduction

At the end of 2019, a novel Coronavirus (SARS-COV-2) pneumonia outbreak appeared in Wuhan, China (Alberti and Faranda, 2020). On 7 February 2020, the National Health Commission, People's Republic of China (PRC) tentatively named it Novel Coronavirus Pneumonia; On 11 February, the World Health Organization (WHO) officially named it "Novel Coronavirus 2019" (COVID-19) with an expected incubation period of approximately 2–10 days (Li et al., 2020). Due to the widespread epidemic, the development of COVID-19 has drawn increasing global attention. According to the COVID-19 Programme of the National Health and Family Planning Commission (version 4), the symptoms of COVID-19 are fever, fatigue, dry cough, and, in some patients, are accompanied by nasal congestion, runny nose, and diarrhea. The early COVID-19 outbreak in China and the large population flow, could increase the spread of COVID-19 and pose a severe threat to human life. The Chinese governments quickly adopted emergency

* Corresponding authors at: GIS Technology Engineering Research Centre for West-China Resources and Environment of Educational Ministry, Yunnan Normal University, Kunming, 650500, China.
E-mail addresses: qian_wang829@163.com (Q. Wang), dong_wen131@163.com (W. Dong), kmdcynu@163.com (K. Yang), zd_ren@163.com (Z. Ren), dongqing_6312@163.com (D. Huang), zhangpeng_152@163.com (P. Zhang), withwangjie@163.com (J. Wang).

https://doi.org/10.1016/j.ijid.2021.03.014

measures to reduce and prevent the spread of the virus, and announced on 23 February 2020, in Wuhan to halt the city public transportation, and limit unprecedented personal mobility. One month later, the effect of isolation gradually emerged, proving that strict restrictions on population movement can play a positive role in curbing the spread of the epidemic (Manevski et al., 2020).

As of October 2020, there were more than 39.96 million cumulative confirmed cases of COVID-19 globally. However, the COVID-19 situation in China has been brought under control and significantly improved. With the rapid increase in confirmed COVID-19 cases globally, the COVID-19 epidemic shows no signs of slowing down. It is essential to support global cooperation and collaborative prevention and control of COVID-19 (Lazarus et al., 2020). Therefore, it has become an urgent scientific issue to identify the temporal and spatial changes of COVID-19 transmission and clarify its driving mechanisms.

Shortly after the outbreak of COVID-19, some scientists conducted extensive research on the epidemic from the aspects of pathogenesis, virology, biology, and clinical medicine and achieved fruitful results, providing a scientific basis for the prevention and control of COVID-19 (J et al., 2020). Xie and Zhu (2020) explored the nonlinear relationship between ambient temperature and confirmed COVID-19 cases, using a generalized additive model. The results indicated that higher temperature might not limit the transmission of this novel Coronavirus. Zhu et al. (2020a, b) explored the relationship between environmental air pollutants and newly confirmed cases of COVID-19 each day, proving that there was a statistically significant relationship between air pollution and COVID-19 infection. The results showed that short-term exposure to high concentrations of PM 2.5, PM 10, CO, NO 2, and O 3 was associated with an increased risk of COVID-19 infection (Zhu et al., 2020b). Sannigrahi et al. (2020) found a strong positive correlation between income/population and COVID-19 cases/deaths, suggesting that these two factors may be key control variables for estimating overall human casualties caused by COVID-19 in European countries (Sannigrahi et al., 2020). Wu et al. (2020), based on confirmed COVID-19 cases and residents' travel by train, plane, and road, published a mathematical model used to predict the trend of the epidemic; the results showed that about 75,815 people in Wuhan would be infected in the early stage of the epidemic (Wu et al., 2020). Chen et al. (2020) studied the correlation between migration index and the number of confirmed COVID-19 cases from 23 January 2020 to 12 February 2020. The results suggest that Wuhan may have played a positive role in controlling COVID-19 by blocking and activating the first-level emergency response to this significant public health crisis (Saqib, 2020). Although some scholars have discussed the transmission law of COVID-19 from the perspective of geography, only natural or social factors were considered when discussing the transmission factors of COVID-19, and the influencing factors of all aspects were not considered comprehensively. Therefore, they cannot have an in-depth understanding of the space-time pattern and influencing factors of the COVID-19 epidemic, which is of great significance for the prevention and control of the epidemic.

In this study, the number of newly confirmed COVID-19 cases per day in mainland China was used as the measurement index; spatial statistics and spatio-temporal scanning methods were used to describe the spatio-temporal distribution of epidemic transmission. Secondly, traditional statistical methods were used to identify the key factors affecting the spread of the COVID-19 epidemic from the two dimensions of social factors and natural factors, to provide a scientific basis for clarifying the spread law of the epidemic and formulating relevant prevention and control measures (Booth et al., 2020).

Data and methods

Data sources

The study used data sets from five different sources, including confirmed COVID-19 case data, Baidu Migration Index data, air quality and meteorological data, and government Response Strict Index data.

Data sets of 31 provinces (excluding Hong Kong, Macao, and Taiwan) and 337 prefecture-level cities were collected from the National Health Commission, PRC (http://www.nhc.gov.cn/) and (https://ncov.dxy.cn/) for the cumulative and daily confirmed cases of COVID-19 in China from January to October 2020, with about 81,000 entries. ArcGIS software was used to realize a visualization of the epidemic situation, in which the coordinate system is GCS_Beijing_1954, and the projection is LAMBERT_CONFORMAL_-CONIC.

The Migration Scale Index data was derived from Baidu migration data (http://qianxi.baidu.com/). This study used the population migration scale indicators of 339 prefecture-level cities in China, including intra-city travel intensity, emigration index, and immigration index.

Average daily air quality data were extracted from the online air quality monitoring and analysis platform (https://www.aqistudy.cn/) during the study period, including air quality index (AQI), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter with aerodynamic diameter <10 um (PM_{10}) and 2.5 um ($PM_{2.5}$), carbon monoxide (CO) and ozone (0_3). The meteorological data comes from the China Meteorological Data Service Center (http://data.cma.cn/en), including the average temperature, air pressure, precipitation, and wind speed.

Government Response Strictness Index (www.bsg.ox.ac.uk/covidtracker) is the Oxford COVID-19 government response tracking system (OxCGRT) provided by a system of measures across countries and across time to understand the government's response in the evolution of the epidemic spread (Sannigrahi et al., 2020). Typical measures include a series of standardized indicators such as school closures, travel restrictions, bans on public gatherings to track government policies and interventions to contain the spread of the virus, strengthen health systems, and manage the economic consequences of these actions (Kim and Castro, 2020).

Research methods

Exploratory spatial data analysis method

Spatial statistics and the modeling tool of ArcGIS 10.6 were used to test whether the confirmed provincial and municipal cases had significant global or local spatial autocorrelation. Spatial autocorrelation, which measures spatial autocorrelation based on feature locations and eigenvalues, can be divided into global spatial autocorrelation and local spatial autocorrelation (Eryando et al., 2020), using global Moran's I statistics to assess whether the cumulative number of confirmed COVID-19 cases in each region is spatially relevant. Moran's I statistics obey normal distribution and test significance based on a Monte Carlo simulation of a stochastic permutation process (Briz-Redon and Serrano-Aroca, 2020). I ranges from -1 (dissimilar value clustering) to +1 (similar value clustering), with 0 indicating that there is no spatial autocorrelation. The greater the absolute value of I, the stronger the spatial autocorrelation.

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij(x_i-\overline{x})(x_j-\overline{x})}}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

Where n is the total number of samples, regarding the x_i, i is confirmed cases, $\overline{x}$ is the average, W_{ij} is the spatial weight matrix between i and j, $\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij(x_i-\overline{x})(x_j-\overline{x})}$ is a covariance, that is to measure the overall error of the number of adjacent confirmed cases, so the value of I depends on the size of the i and j values for the mean deviation symbol; if the adjacent position, x_i, and x_j have the same number, then I is positive, or negative.

The expression of score Z_I is:

$$Z_I = \frac{I - E[I]}{\sqrt{E\left[I^2\right] - E[I]^2}}$$

Where, E[I] = -1/(n-1), the significance of the index is evaluated by calculating Moran's I value, score Z, and P-value. Statistical Z scores and P-values must be combined with Moran's I index to determine statistical significance (Briz-Redon and Serrano-Aroca, 2020). The global spatial autocorrelation technique was used to detect the spatial characteristics of the confirmed cases and analyze the overall spatial correlation within the entire study area, i.e., China.

Local spatial autocorrelation is usually characterized by local Moran's I and Getis-Ord Gi*. Local Moran's I is the decomposition of global Moran's I into sub-regional units. In this study, ArcGIS was used to detect the local spatial autocorrelation characteristics of COVID-19 cases and to identify areas with significant-high/low aggregation. In addition; the hot spot analysis (Getis-Ord Gi*) was used to identify hot spots and cold spots of COVID-19 cases at different spatial scales, with statistical significance (Islam et al., 2020).

A retrospective spatio-temporal scanning statistical method was used to include a time dimension in the analysis and determine when and where, the cluster occurred (and for how long) (Vadrevu et al., 2020). Because COVID-19 is transmitted from person to person, and more cases are likely to occur in densely populated areas, we selected the Poisson model in SaTScan to calculate each area's population. The spatio-temporal scanning statistics base is positioned around a possible center in the entire area, with the radius changing continuously from zero to a specified maximum (Vadrevu et al., 2020). Due to the virus's infectivity, the number of daily cases increases rapidly, and the time interval for this study is one day. The maximum spatial clustering scale was set to 15% of the risk population, and the maximum temporal clustering scale was set to 35%, in order to avoid super-large (and therefore meaningless) clustering.

Exploratory influencing factor analysis method

Secondly, the Spearman rank correlation analysis method was used to analyze the relationship between confirmed COVID-19 cases and migration index, air quality, meteorological data, and the government response strict index. Descriptive statistical results

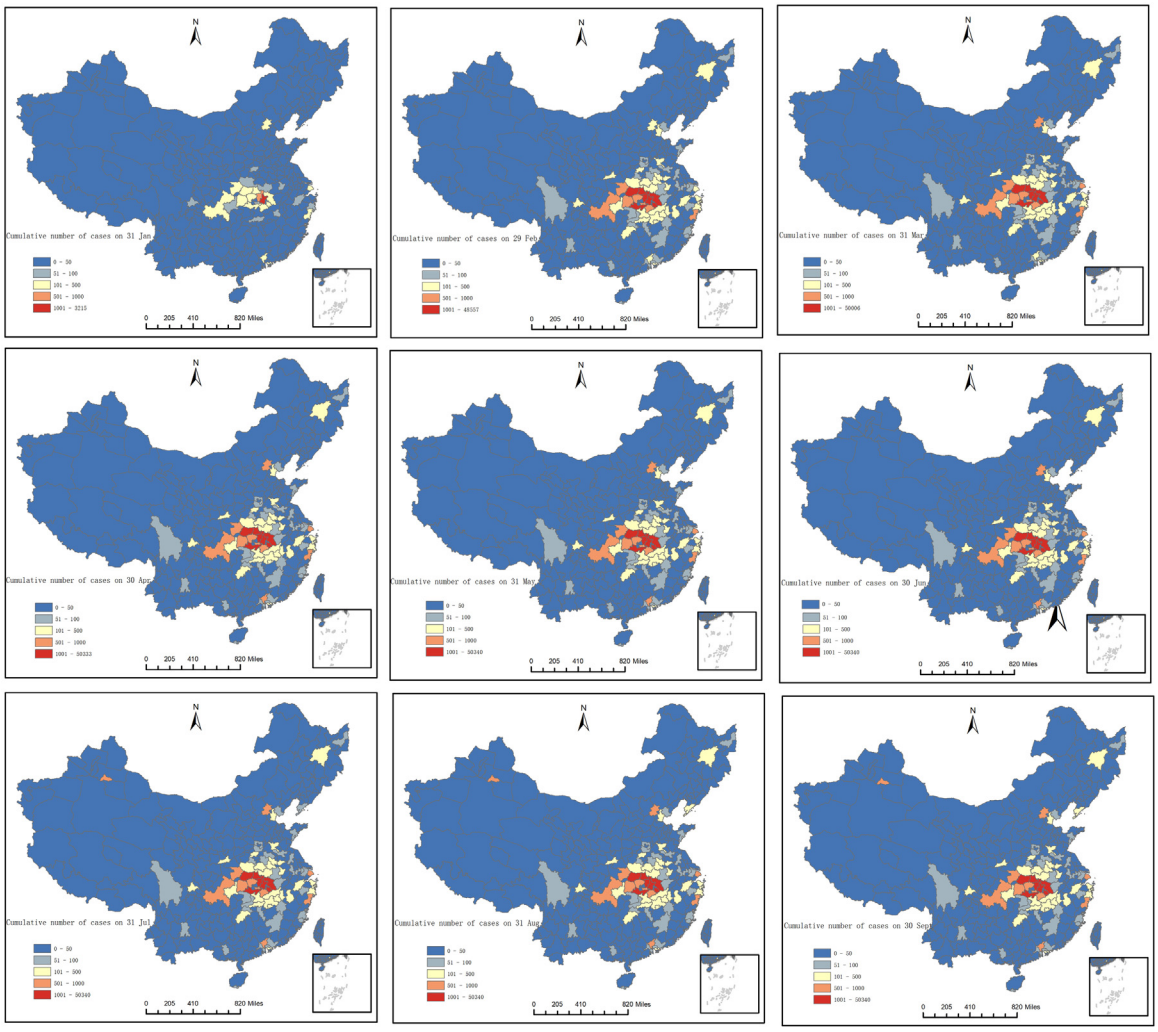

Figure 1. Spatial distribution of cumulative confirmed cases of COVID-19.

showed that COVID-19 outbreaks and related data did not meet the Pearson correlation analysis preconditions and were mainly manifested as a non-Gaussian normal distribution, spatial autocorrelation, and possible nonlinear relations. In general, Spearman's rank correlation is an appropriate nonparametric estimator for estimating the correlation between two variables with unknown or non-Gaussian statistical distributions, and the relationship between these variables does not need to be linear (Rahman et al., 2020). It is usually measured in terms of Spearman's rank correlation coefficient ρ; the formula is as follows:

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

n is the total number of samples, x_i and y_i are the ranks of X_i and Y_i, respectively, ρ represents the Spearman rank correlation coefficient. This coefficient varies between -1 and + 1; the greater the absolute value of ρ, the stronger the relationship between the two variables. Like Pearson's coefficient, Spearman's absolute value of ρ in the range of 0.8–1 indicates a very strong correlation, 0.6–0.8 a strong correlation, 0.4–0.6 a medium correlation, 0.2–0.4 a weak correlation, and 0–0.2 an uncorrelation (Rahman et al., 2020).

Results

Spatial distribution characteristics

ArcGIS software was used to classify the cumulative number of confirmed COVID-19 cases in China into the following five categories: 1–50; 51–100;101–500; 501–1000; > 1000 (Figure 1).

As shown in Figure 1, as of 31 January, the cumulative number of COVID-19 cases at the beginning of the outbreak was 15, 2, and 1, respectively, in the 101–500, 501–1000, and >1000 ranges. Hubei province (7037 cases) accounted for 61.43% of the total cases, and Wuhan city (3215 cases) had the most confirmed cases, accounting for 45.69% of the total cases in Hubei province. In terms of spatial distribution, the regions with a high number of confirmed cases are mainly located around Hubei Province, indicating that the COVID-19 epidemic has been confirmed. On 29 February, the cumulative number of cases in 101–500 (29 cases), 501–1000 (seven cases), and > 1000 (eight cases), have risen sharply, reflecting the fact that the epidemic's geographical scope has expanded significantly. Except for the 1–50-people interval it was found that all other interval percentages of the city are on the rise, indicating that the outbreak has reached the epidemic stage. In terms of spatial distribution, it has the highest number of confirmed cases accumulated; there is a clear trend of a continuous distribution, mainly concentrated in Wuhan and economically developed cities nearby (for example, Shanghai and Beijing). While the number of cases with less area coverage is relatively stable (for example, Qinghai, Tibet, and Xinjiang), the differences between them show that the spread is increasing. Due to the closure of cities in February, the epidemic has been effectively brought under control. From 31 March to 30 April, except for Beijing, Shanghai, and Guangdong, the cases in other regions remained basically unchanged, which means that the spread of the epidemic has been initially contained. As of 31 July, the rapid increase of cases in Urumqi city, Xinjiang province (552 cases) was effectively controlled by 31 August.

Spatial correlation characteristics

Global spatial correlation characteristics

In this paper, the cumulative number of confirmed COVID-19 cases was taken as a variable, the spatial weight matrix based on geographical adjacency was selected, and the global Moran's I index, P test value, and Z statistical score were used to determine the number of confirmed cases with different types of neighbors (Vadrevu et al., 2020). To investigate whether there is a spatial correlation between confirmed COVID-19 cases in China,

Figure 2 shows the Moran's I index and Z statistical scores of the cumulative confirmed COVID-19 cases in China's prefecture-level cities from 15 January to 1 October. On January 15, the global spatial autocorrelation ($p > 0.05$, $Z < 1.96$) was not significant, while on January 23, there was a significant global spatial autocorrelation ($p < 0.05$, $Z > 1.96$). On 31 January, solstice, and 1 October, the cumulative confirmed cases at the prefecture-level showed a significant global spatial autocorrelation ($p < 0.0001$, $Z > 9.58$), indicating that the cumulative confirmed cases at the prefecture-level showed a very significant spatial dependence. In Figure 2, the trend change characteristics of Moran's I index are presented in two stages: they first increase and then decrease from 23 January, indicating that this may be a turning point. This means that, although the degree of clustering is lower than before, the global spatial correlation is still dominated by clustering characteristics and tends to develop in a decentralized manner.

Local spatial correlation characteristics

Different from global spatial autocorrelation, local spatial autocorrelation analysis deals with heterogeneous regions. Clustering and outlier analysis was conducted on the data of cumulative confirmed COVID-19 cases in China's prefecture-level cities from 15 January to 1 October. The results are shown in Figure 3. High-high is represented in pink and represents the high-value aggregation class; High-Low is shown in red, representing high-value elements and is surrounded by Low-value elements. Low-high is shown in dark blue, representing low-value elements and surrounded by high-value elements; Low-low, shown in light blue, represents the low-value cluster classes.

According to Figure 3, most cities in Hubei Province, China belong to the "high-high" clustering region, in addition to the surrounding cities (Xinyang City, Henan Province; Lu an and Xuanzhou, Anhui Province; Changde city, Hunan Province; Yueyang and Changsha City; Nanchang, Yichun And Jiujiang City, Jiangxi Province). In contrast, some regions close to these prefecture-level cities have significant "low-high" clustering regions." High-low" clustering areas appeared in Kunming city, Yunnan Province, on 31 January and Urumqi city, Xinjiang Province on 31 July. There is a continuous distribution trend in "low-low" clustering regions, mainly located in Inner Mongolia, Gansu, Ningxia, Qinghai, Tibet, and Xinjiang. From the perspective of

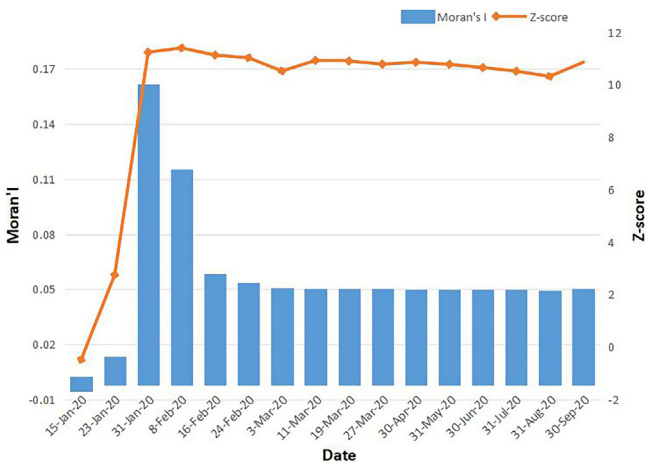

Figure 2. Global Moran's I index of the cumulative number of cases.

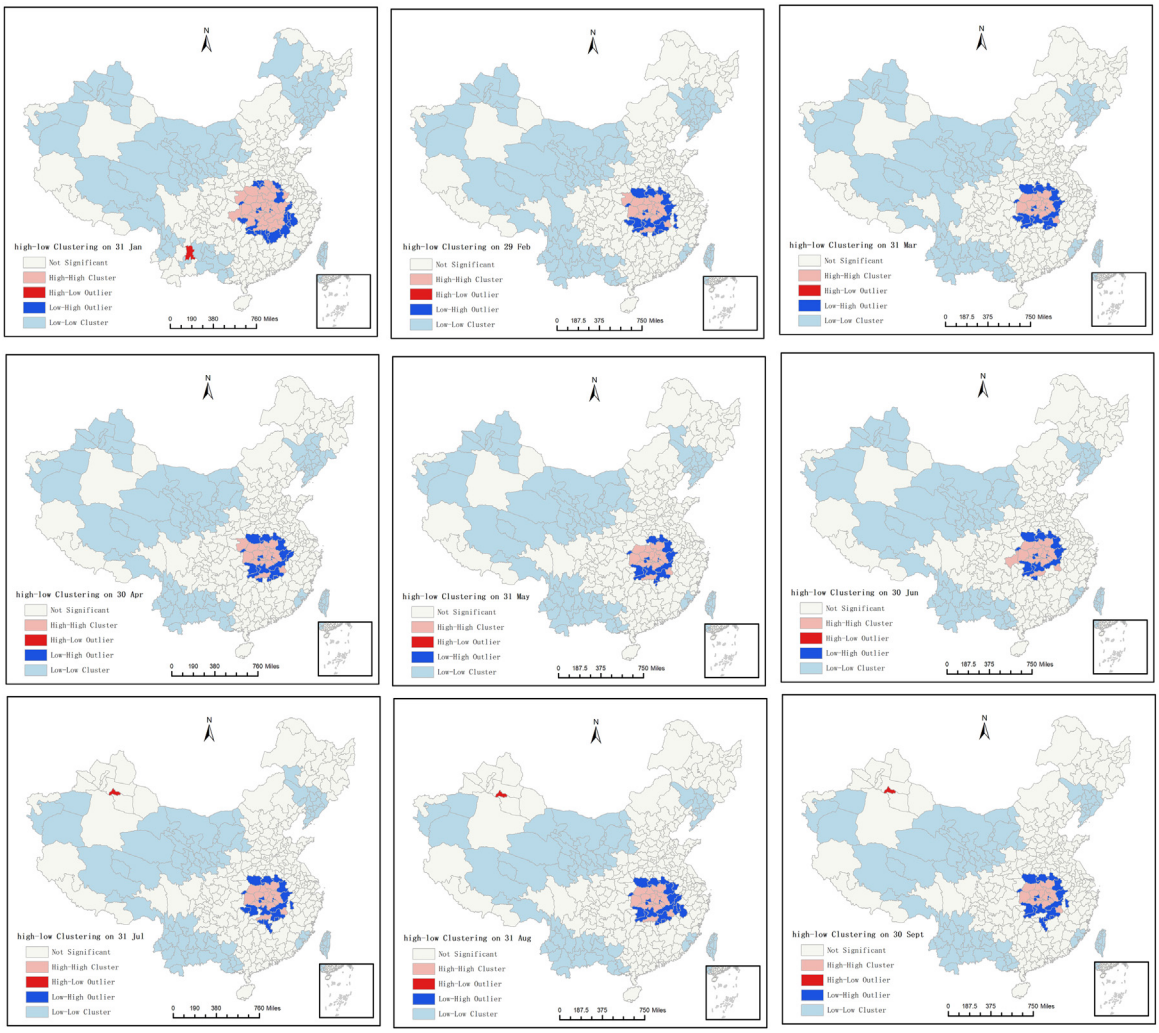

Figure 3. Local spatial correlation Characteristics of the cumulative number of cases.

quantity change, the number of units in all clustering areas has decreased. This suggests that, despite a weakening clustering trend, the local spatial correlation characteristics of confirmed COVID-19 cases are also predominantly positive.

Based on the hot spot analysis (Getis-Ord Gi*) results given on 15 January, the initial stage of the outbreak in Hubei province as the center of the area and the surrounding adjacent urban areas, a total of 32 cities present hot spots (90% Confidence). But by the end of January, hot spots quickly covered most of the city; as well, Henan, Hunan, Anhui, and Jiangxi provinces become significant hot spots (99% Confidence). By 29 February, the number of hot spots in Henan and Anhui had decreased, becoming an important hot spot (95% Confidence), while the remaining areas remained unchanged. On 31 March, Anqing city in Anhui Province became an extremely important hot spot area again, and the hot spot area remained unchanged after April (Figure 4).

Spatio-temporal scanning analysis

Spatio-temporal clustering of confirmed COVID-19 cases was explored using the spatio-temporal scanning analysis method, and a total of six clustering areas were detected through SaTScan software, including 17 provinces (Figure 5). It was found that the actual number of COVID-19 cases increased abnormally compared with the theoretical number during the period from 27 January 2020 to 1 March 2020, indicating a high incidence of COVID-19 aggregation. Daily scanning analysis showed that 6 April to 19 April was a high incidence month in Heilongjiang province, 27 July to 2 August was a high incidence month in Xinjiang Province, and the rest of the high incidence months were concentrated in the early phase of the epidemic from January to March. Also, the relative risk (RR) and likelihood ratio (LLR) of the first three aggregation regions were all >3, and all $P < 0.05$. The aggregation region with the largest RR (491.57) was located in Hubei Province. This region's aggregation risk was 491.57 times that of other regions, and its LLR value (280453.82) was also the highest (Table 1).

Analysis of influence factors

Generally, there are three links in the epidemic process of infectious diseases (source of infection, route of transmission, susceptible group) and two influencing factors (social factors and natural factors) (Zhu et al., 2020a). In this study, social factors (population migration index, government response severity index) and natural factors (air pollutants and climate factors) were analyzed and studied together with daily confirmed cases of COVID-19.

This study adopted the Chinese migration index to trace population flows, The intensity of its migration in (out) is the ratio of the number of individuals moving in (out) in a unit of time to the total number of individuals in the region.

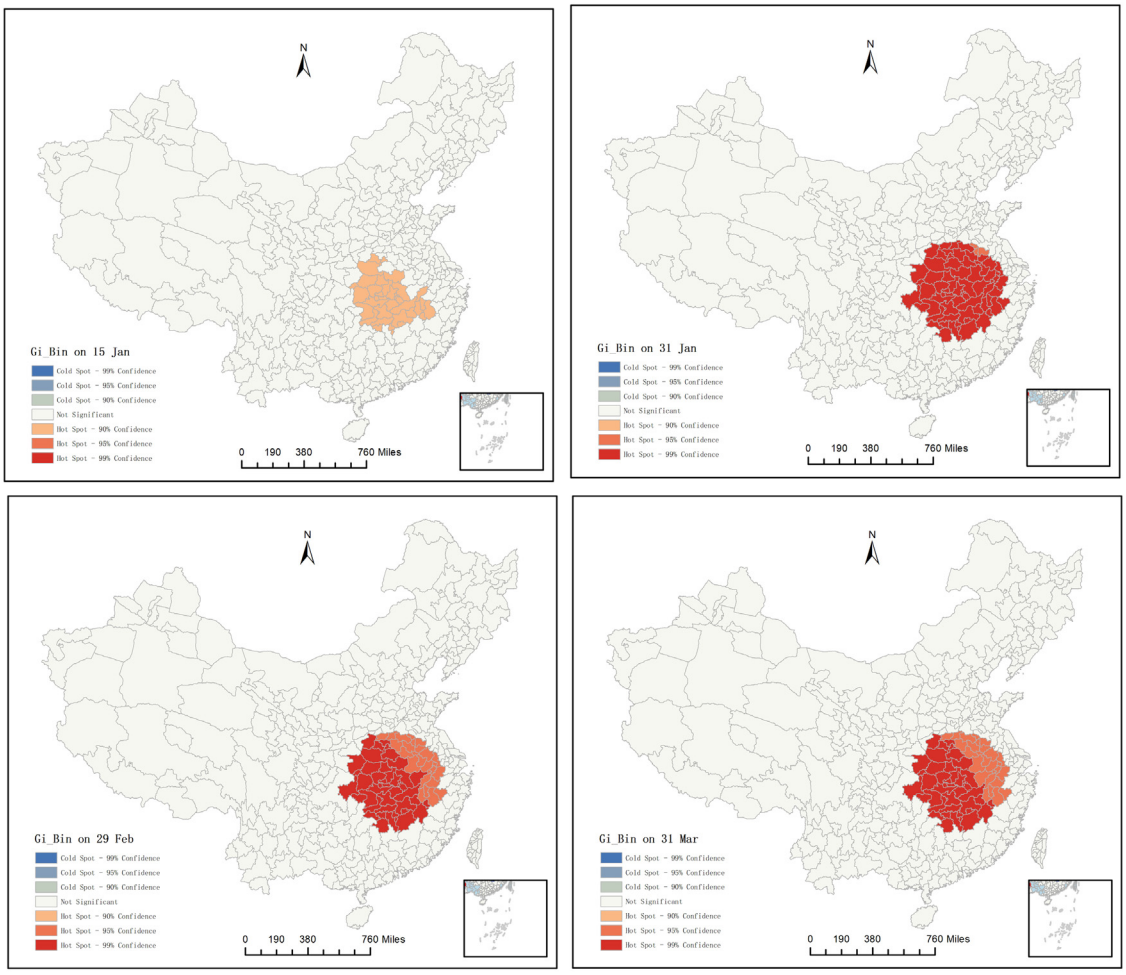

Figure 4. Hotspot (Getis-Ord Gi) analysis of the cumulative number of cases.

Table 1
Spatial-temporal scan statistics on the cumulative number of cases.

Number	Time	Concentrated areas	The actual number of cases	Number of theoretical cases	LLR	RR	P
1	2020/1/27–2020/3/1	Hubei	65,680	599.39	280453.82	491.57	<0.001
2	2020/7/27–2020/8/2	Xinjiang	455	50.38a	597.66	9.07	<0.001
3	2020/1/27–2020/2/9	Beijing	269	87.29	121.24	3.09	<0.001
4	2020/1/27–2020/2/9	Fujian, Zhejiang, Shanghai, Hunan Anhui, Guangdong	5145	1997.95	1780.53	2.68	<0.001
5	2020/4/6-2020/4/19	Heilongjiang	381	152.88	120.09	2.50	<0.001
6	2020/1/27–2020/2/9	Gansu, Ningxia Shaanxi, Sichuan Chongqing, Shanxi Henan	2090	1640.77	57.77	1.28	<0.001

First, ArcGIS software was used to map the migration intensity of the outflow population from Wuhan to cities across the country (Figure 6). It can be seen that the cities of Hubei Province are the popular migration destinations from Wuhan, followed by East China (including Shandong, Jiangsu, Anhui, Zhejiang, Fujian, and Shanghai), South China (including Guangdong, Guangxi, and Hainan), Central China (including Hunan, Henan, and Jiangxi), and North China (including Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia). The greater the intensity of migration, the greater the risk of disease in the region. Due to the closure of Wuhan city on 23 January and the closure of Hubei Province on 27 January (except for the Shennongjia area), the scale of population emigration in Wuhan decreased significantly in February, only the movement of supplies and epidemic prevention and control personnel are left. China lifted its last urban blockade of Wuhan on 7 April, after reopening other cities in March, so the migration index gradually rose in March and April.

Secondly, a visual diagram of average air quality (AQI), average precipitation (PRCP), average wind speed (WDSP), and temperature (TEMP) detection factors were drawn (Figure 7).

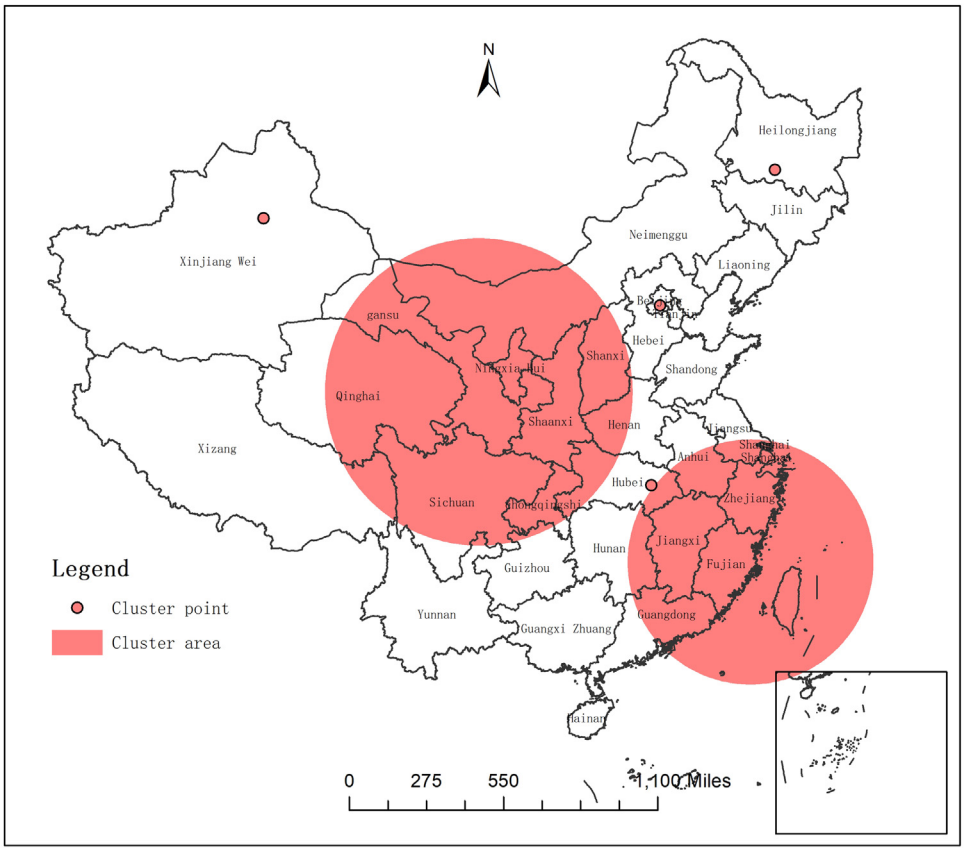

Figure 5. Spatial-temporal scan shows the cumulative number of cases.

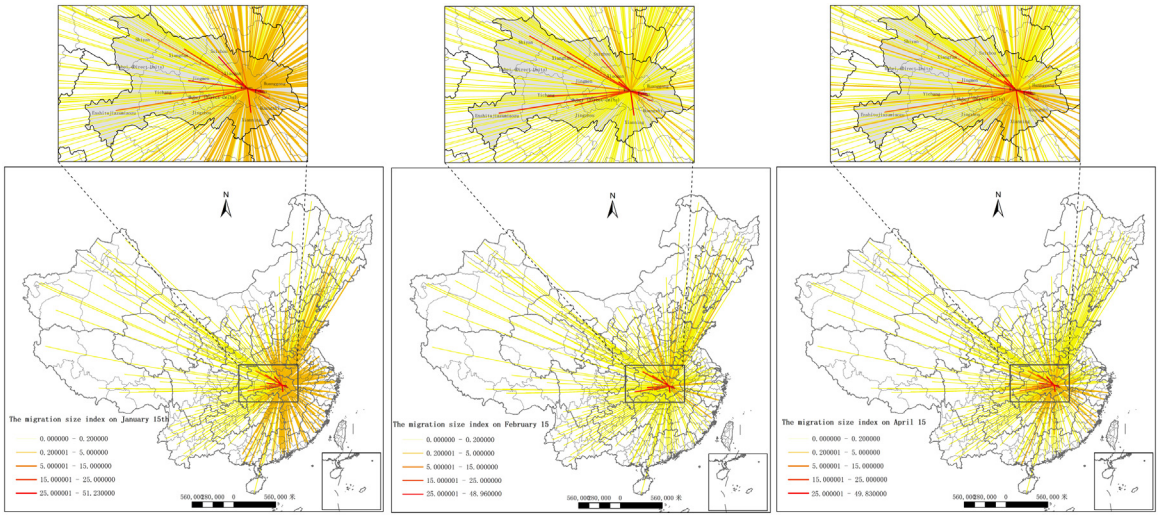

Figure 6. Population migration index of Wuhan and the whole country in January. February and April.

The government response strictness index codes qualitative policies into numbers and then obtains the average value of these specific policies, such as school closures, business closures, cancellation of public events, and generalized codes to represent the scope of specific policies (Sannigrahi et al., 2020). The index is, therefore, a good indicator of the government's response to the current crisis. Table 2 summarizes the descriptive statistics for COVID-19 confirmed cases and government response strictness index variables.

Table 3 lists the statistics of daily new COVID-19 cases and the air pollution concentration, meteorological variables, and the migration index. From January to October, the study included more than 85,000 cases, the mean value was 338.34 cases per day. The daily average concentrations of $PM_{2.5}$, PM_{10}, SO_2, CO, NO_2 and O_3 were 29.32 μg/m^3, 53.95 μg/m^3, 9.43 μg/m^3, 0.68 μg/m^3, 21.02 μg/m^3 and 69.22 μg/m^3, respectively; the average values of temperature (Temp), wind speed (Wdsp), precipitation (Prcp) and gusts (Gust) were 62.43 °F = 16 °C, 4.79 knots = 2.46 m/s, 0.14 in. = 3.556 mm and 9.89 knots = 5.08 m/s, respectively. The average daily immigration index (IM), emigration index (EM), and inner-city travel intensity (Inner) were 0.83, 0.83, and 3.99, respectively.

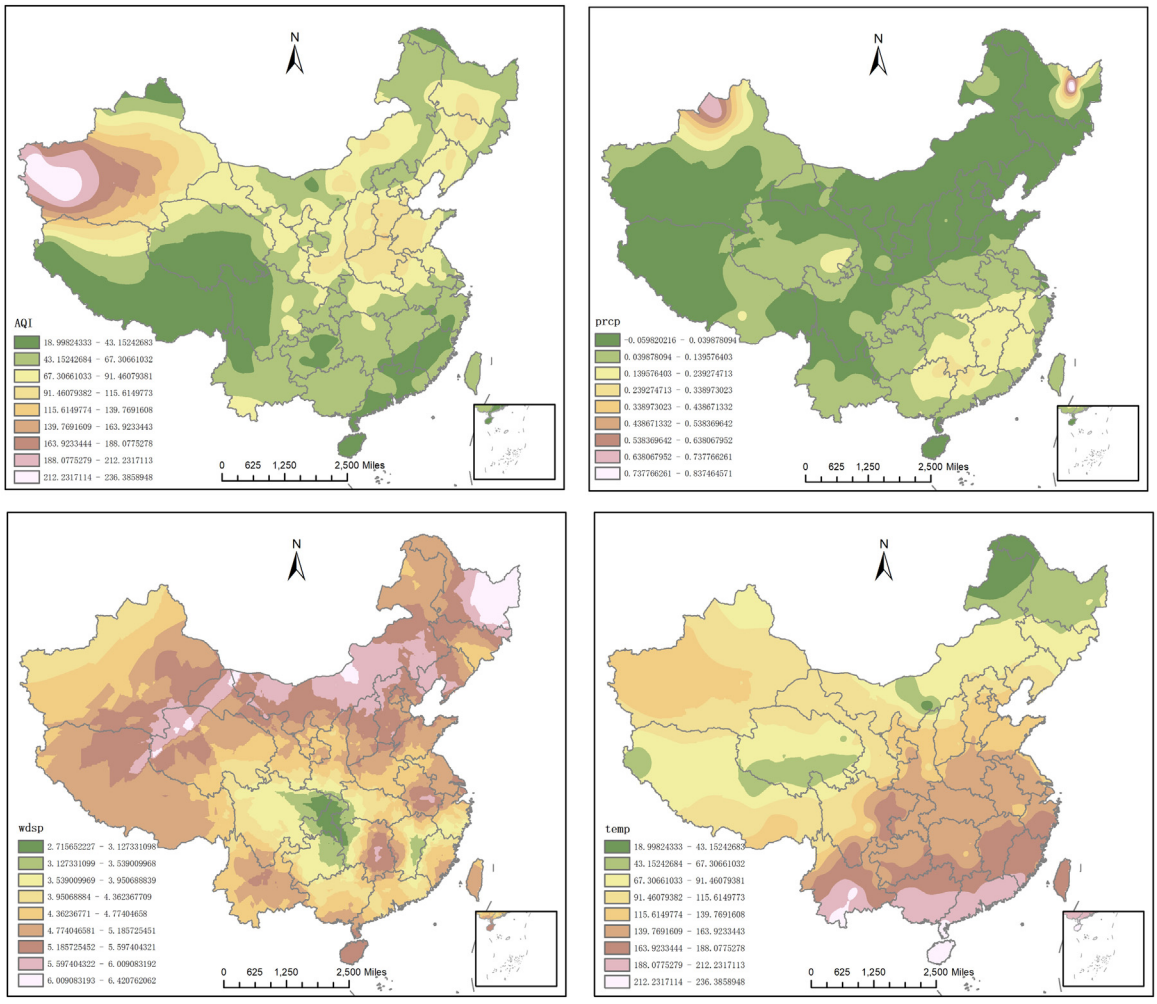

Figure 7. Spatial distribution of Average air Quality (AQI), average precipitation (PRCP), average wind speed (WDSP), and Temperature (TEMP) In China.

Table 2
Descriptive statistics of categorical variables.

Variable	Descriptive	Coding instructions
C1	School closing	0: No measures; 1: Recommended to close; 2: Request to close; 3: Ask to close all
C2	Workplace closing	0: No measures; 1: Recommended to close; 2: Request to close; 3: Ask to close all
C3	Cancel public events	0: No measures;1: Recommended to close; 2: Request to close
C4	Restrictions on gathering size	0: Unlimited; 1: Limit more than 1000 employees; 2. Limit within 101–1000; 3: Limit within 11–100; 4: Limitless than 10 people
C5	Close public transport	0: No measures; 1: Recommended to close; 2: Request to close
C6	Stay at home requirements	0: No measures; 1: Don't leave home; 2: Ask not to leave home except for "essential" travel; 3: Ask not to leave the premises, but with very few exceptions
C7	Restrictions on internal movement	0: Unlimited; 1: Screening arrival; 2: Some areas are isolated from arriving;
C8	International travel controls	0: Unlimited; 1: Screening arrival; 2: Some areas are isolated from arriving; 3: No access to certain areas; 4: Prohibition of all areas or complete closure of borders

As the variables are not normally distributed, the Spearman rank correlation test was performed. Table 4 shows the correlation coefficient between the newly increased COVID-19 cases per day and the air pollution concentration and meteorological variables. Among the meteorological variables, the daily average temperature, wind speed, and precipitation were negatively correlated with the daily new cases (Zhu et al., 2020a). Maximum wind speed was positively correlated with newly confirmed cases. Air pollution indexes (CO, $PM_{2.5}$, PM_{10}, and SO_2), air quality index (AQI), and migration index were positively correlated with newly confirmed cases. International travel control (C8) in the Government response stringency index was strongly negatively correlated with confirmed cases of COVID-19, with $P < 0.05$, which was statistically significant.

Table 5 shows the analysis of air pollution concentrations and weather information data from the date of the daily report of new COVID-19 cases (lag 0, lag 03, lag 07, and lag 014, i.e., the incubation period of COVID-19).

The results of multiple linear logistic regression showed that temperature and control of international travel were strongly negatively correlated with the confirmed cases of COVID-19 among the remaining factors after removing the factors affected by

Table 3
Descriptive statistics of continuous variables.

	Minimum	Maximum	Mean ± SD.
New confirmed	−1	14,106	338.34 ± 1165.36
Meteorological factors			
Temp (°F)	32.94	80.42	62.43 ± 14.49
Wdsp(knots)	3.42	6.53	4.79 ± 0.62
Prcp(inches)	0.005	0.43	0.14 ± 0.09
Gust (knots)	4.25	22.63	9.89 ± 3.68
The concentration of air pollutants			
AQI	29.21	104.69	52.04 ± 16.36
$PM_{2.5}(\mu g/m^3)$	12.50	77.64	29.32 ± 13.36
$PM_{10}(\mu g/m^3)$	25.84	98.52	53.95 ± 17.60
$NO_2(\mu g/m^3)$	10.21	37.00	21.02 ± 4.75
$SO_2(\mu g/m^3)$	7.02	17.26	9.43 ± 1.71
$CO(mg/m^3)$	0.54	1.21	0.68 ± 0.13
O_3 $(\mu g/m^3)$	38.51	102.03	69.22 ± 11.34
migration index			
Im	0.24	2.29	0.83 ± 0.43
Em	0.24	2.20	0.83 ± 0.41
Inner	2.09	5.41	3.99 ± 1.02

Table 4
Association between confirmed cases of COVID-19 and influencing-factor variables.

variable	correlation coefficient	P value
CO	0.450**	<0.001
$PM_{2.5}$	0.354**	<0.001
PM_{10}	0.205**	<0.001
SO_2	0.286**	<0.001
AQI	0.424**	<0.001
wdsp	−0.137*	<0.05
temp	−0.449**	<0.001
prcp	−0.270**	<0.001
gust	0.453**	<0.001
C8	−0.628**	<0.001
Im	0.323**	<0.001
Em	0.323**	<0.001
Inner	0.471**	<0.001

Table 5
Variables associated with confirmed cases of COVID-19 in different days of lag.

variable	Lag 0	Lag 03	Lag 07	Lag 014
CO	0.450**	0.455**	0.423**	0.424**
NO_2	−0.146*	−0.158*	−0.186**	0.008
O_3	−0.552**	−0.530**	−0.551**	−0.485**
$PM_{2.5}$	0.354**	0.367**	0.363**	0.389**
PM_{10}	0.205**	0.207**	0.183**	0.266**
SO_2	0.286**	0.276**	0.240**	0.287**
AQI	0.424**	0.409**	0.376**	0.383**
wdsp	−0.137*	−0.124	0.087	−0.016
temp	−0.449**	−0.462**	−0.490**	−0.514**
prcp	−0.270**	−0.263**	−0.281**	−0.360**
gust	0.453**	0.424**	0.415**	0.428**

collinearity. The maximum wind speed has a weak negative correlation. The intensity of Inner city travel is positively correlated, and its regression equation is y=-0.749* TEMP-0.300 *Gust+0.434*Inner-0.774*C8 (Table 6).

Discussion

This paper studied the spatio-temporal variation of COVID-19 in China and the key influencing factors of its transmission, which should be of significance to contain the spread of the COVID-19 epidemic, to provide a reference for the formulation of public health policies, and to promote productive recovery.

The mining of geospatial information can not only reveal the spatio-temporal transmission and clustering characteristics of the epidemic but also uncover the spatial risk factors that have a significant influence on the spread of the epidemic and identify the hot spots with high transmission risk, which is helpful for the scientific prevention and control of the epidemic. This paper shows the spatial distribution of the cumulative confirmed cases of COVID-19, and it can be seen that the areas with more cases have an obvious continuous distribution trend, with Hubei Province as the center for diffusion, while the areas with fewer cases have a relatively stable coverage. Moran's I is usually an important index to measure spatial correlation, which can be divided into global Moran's I and local Moran's I. The global Moran's I can only indicate whether space is clustered or an outlier, not where it is located, while the local Moran's I will indicate where the outlier or cluster appears. Hafner (2020) fitted the spatial autoregressive model for the global COVID-19 propagation process and found that the estimated spatial correlation was highly significant, consistent with our results. As shown in Section 3.2, the global Moran's I index, p-test value, and Z-statistical score were used to determine a significant spatial association among the confirmed COVID-19 cases in China. Secondly, local Moran's I was helpful to effectively detect the significant clustering/outlier areas and significant hot spots at different spatial levels in China. Hot spot analysis was used to identify hot spots and cold spots of COVID-19 cases at different spatial scales. As far as China is concerned, the first-generation epidemic transmission pattern is from the South China seafood market to the whole city of Wuhan. In the second generation epidemic transmission mode, the epidemic spread from Wuhan to the counties and cities in Hubei Province, the key cities outside Hubei Province and abroad; as the epidemic spread into the critical cities outside Hubei Province, it further spread within them. The third-generation epidemic transmission model was imported from other countries to China. The results show that the closer to a high-risk area, the greater the risk, especially at the prefecture-level city scale. This finding is helpful for governments at all levels to take effective classified prevention and control measures.

Identifying the key factors affecting the spread of the COVID-19 epidemic is of great significance for containing the spread of the COVID-19 epidemic. Generally speaking, the factors that affect the

Table 6
Multiple linear logistic regression.

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std.Error	Beta		
(Constant)	3467.569	679.181		5.106	<0.001
temp	−62.550	8.204	−0.749	−7.624	<0.001
gust	−105.026	32.241	−0.300	−3.258	<0.001
Inner	19.815	2.810	0.434	7.050	<0.001
C8	−863.962	83.534	−0.774	−10.34	<0.001

outbreak and transmission of an epidemic occur mainly through the influence on the source of infection, the route of transmission, and the susceptible population. The spread of COVID-19 may be closely related to some natural, social, and economic factors as well as prevention and control policies, such as controlling the source of infection and cutting off transmission routes. Recently, researchers have shown that meteorological parameters such as temperature and humidity are potential environmental factors affecting the outbreak of COVID-19. This is consistent with the research results of meteorological parameters in this paper. After controlling population flow for the first time, (Liu et al., 2020) compared the relationship between meteorological parameters and COVID-19, and the results showed that low weather conditions, moderate temperature range, and low humidity were conducive to the spread of COVID-19. Fattorini and Regoli (2020) studied the relationship between atmospheric pollutants (NO_2, O_3, $PM_{2.5}$, and PM_{10}) and the transmission of COVID-19 in Italy, showing that long-term exposure to atmospheric pollution may be a favorable environment for the transmission of the virus. This is also consistent with this paper's research results, which show that air pollution concentration plays a positive role in the transmission of COVID-19.

In our society, most of these factors need to function through the population, so the spread of the COVID-19 epidemic is highly correlated with population movements, which may exacerbate the spread of novel coronavirus and pose a severe threat to human life and public health. Based on the migration intensity of the outflow population in Wuhan, this study describes the population flow in January, February, and April. Due to the blockade, the population outflow scale index in Wuhan decreased significantly in February, but after the blockade was lifted in March and April, the migration scale index began to rise gradually. Ran et al. (2021) showed that large-scale social unrest in the United States may be related to the rebound of COVID-19's transmission power. However, previous studies only analyzed the influencing factors of the COVID-19 epidemic on a spatial scale. This study elaborates on the natural and social influencing factors of the epidemic change in prefecture-level cities in China. Surprisingly, due to the mutual restriction of multiple factors, the factors removing collinearity may lead to prefecture-level city differences, showing a certain spatial scale effect, thus removing many potential influencing factors. Therefore, the causes and mechanisms of the spatial scale effect are worth further study.

This paper also has some limitations. The factors that affect the spread of the epidemic are very complex. Based on the available data, this paper constructed an indicator system that affects the epidemic's multi-factors (Pei et al., 2020). Other non-quantitative indicators may have been ignored, which increases the inadequacy of the evaluation of the study results. Without access to detailed information (such as age, sex, medical history, and smoking status) about those patients diagnosed with COVID-19, it is impossible to determine how the underlying health problems may have contributed to the spread of the COVID-19 infection; this may have weakened the study results. In addition, regional differences in medical capacity and socio-economic status may also affect the number of COVID-19 patients. However, attempts to analyze possible external environmental impacts are essential for protecting healthcare professionals and the containment of the COVID-19 epidemic (Lian et al., 2020). Future studies may be helpful in considering the epidemiological parameters and social context of COVID-19 in more detail.

Conclusion

Based on the multi-scale open data of the COVID-19 epidemic in 337 prefecture-level cities in China, a new geographic database was established to study the COVID-19 epidemic. This paper used the ArcGIS spatial statistical method to analyze the spatial-temporal pattern of the COVID-19 epidemic and explore various influencing factors to analyze the outbreak and transmission of the epidemic in China from January to October 2020, and reached the following conclusions.

(1) Global spatial autocorrelation was used to confirm that there was a spatial correlation among the confirmed cases of COVID-19 in China, and the correlation characteristics were firstly increased and then decreased. However, considering local spatial autocorrelation, the correlation characteristics tend to be stable with the passage of time, and are mainly composed of high/low aggregation regions. The hot spots have also stabilized over time in the provinces surrounding Hubei (Henan, Hunan, Anhui, and Jiangxi).
(2) Spatial and temporal clustering, high incidence area, and time of confirmed cases of COVID-19 were explored by using a spatial and temporal scanning analysis method. Hubei Province (2020/1/27-2020/3/1) had the highest RR(491.57); that is, the clustering risk in this region was 491.57 times higher than that of other regions.
(3) Among social factors, the migration index was positively correlated with confirmed cases of COVID-19. It shows that in the early stage of the epidemic, with the expansion of the scale of migration, the probability of disease is greater. The government response strictness index is negatively correlated with international travel restrictions, indicating that under the current situation of effective control of the epidemic in China, it is necessary to pay close attention to population migration, especially international travel, thereby limiting human-to-human transmission and ultimately helping to contain the outbreak in China.
(4) For natural factors, the correlation test between the new cases of COVID-19 every day and the concentration of air pollution and meteorological variables showed that temperature was negatively correlated with the newly confirmed cases, indicating that the ambient temperature had a particular inhibitory effect on the transmission of COVID-19. Air pollution indicators (CO, $PM_{2.5}$, PM_{10}, SO_2) and air quality index (AQI) were positively correlated with newly confirmed cases. Meanwhile, in the lag effect, the average temperature increases with the increase of cumulative lag days.

In conclusion, the transmission rate of COVID-19 in China has evident spatial variation, and the spatio-temporal aggregation is also apparent. Research shows that population migration, air pollution concentration, and temperature on the spread of COVID - 19 played a positive role; government response strictness index, namely the related policy-making plays an inhibitory effect on COVID-19 spread.

Funding source

The study was supported by the National Natural Science Foundation of China (Grant No. 41661087).

Ethical approval

This study did not require ethical approval as the analysis was based on publicly available data.

Conflict of interest

All authors declare that there is no conflict of interest regarding this work.

CRediT authorship contribution statement

Qian Wang: Conceptualization, Formal analysis, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Wen Dong:** Funding acquisition, Project administration, Resources, Supervision, Validation, Writing - review & editing. **Kun Yang:** Funding acquisition, Project administration, Resources, Validation, Writing - review & editing. **Zhongda Ren:** Data curation, Investigation, Supervision, Visualization, Writing - review & editing. **Dongqing Huang:** Data curation, Formal analysis, Writing - review & editing. **Peng Zhang:** Formal analysis, Investigation, Writing - review & editing. **Jie Wang:** Formal analysis, Investigation, Writing - review & editing.

References

Alberti T, Faranda D. On the uncertainty of real-time predictions of epidemic growths: a COVID-19 case study for China and Italy. Commun Nonlinear Sci Numer Simul 2020;90.

Booth AL, Abels E, McCaffrey P. Development of a prognostic model for mortality in COVID-19 infection using machine learning. Modern Pathol 2020;.

Briz-Redon A, Serrano-Aroca A. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. Sci Total Environ 2020;728.

Chen H, Chen Y, Lian Z, Wen L, Sun B, Wang P, et al. Correlation between the migration scale index and the number of newly confirmed coronavirus disease 2019 cases in China. Epidemiol Infect 2020;148.

Eryando T, Sipahutar T, Rahardiantoro S. The risk distribution of COVID-19 in Indonesia: a spatial analysis. Asia-Pac J Public Health 2020;.

Fattorini D, Regoli F. Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy. Environ Pollut 2020;264:114732.

Hafner CM. The spread of the Covid-19 pandemic in time and space. Int J Environ Res Public Health 2020;17(11):3827.

Islam MS, Tusher TR, Roy S, Rahman M. Impacts of nationwide lockdown due to COVID-19 outbreak on air quality in Bangladesh: a spatiotemporal analysis. Air Qual Atmos Health 2020;.

Kim S, Castro MC. Spatiotemporal pattern of COVID-19 and government response in South Korea (as of 31 May 2020). Int J Infect Dis 2020;98:328–33.

Lazarus JV, Ratzan S, Palayew A, Billari FC, Binagwaho A, Kimball S, et al. COVID-SCORE: a global survey to assess public perceptions of government responses to COVID-19 (COVID-SCORE-10). PLoS One 2020;15(10).

Li AY, Hannah TC, Durbin JR, Dreher N, McAuley FM, Marayati NF, et al. Multivariate analysis of black race and environmental temperature on COVID-19 in the US. Am J Med Sci 2020;360(4):348–56.

Lian X, Huang J, Huang R, Liu C, Wang L, Zhang T. Impact of city lockdown on the air quality of COVID-19-hit of Wuhan city. Sci Total Environ 2020;742.

Liu J, Zhou J, Yao J, Zhang X, Li L, Xu X, et al. Impact of meteorological factors on the COVID-19 transmission: a multi-city study in China. Sci Total Environ 2020;726:138513.

Manevski D, Gorenjec NR, Kejzar N, Blagus R. Modeling COVID-19 pandemic using Bayesian analysis with application to Slovene data. Math Biosci 2020;329.

Pei Z, Han G, Ma X, Su H, Gong W. Response of major air pollutants to COVID-19 lockdowns in China. Sci Total Environ 2020;743.

Rahman M, Islam M, Shimanto MH, Ferdous J, Rahman AA-NS, Sagor PS, et al. A global analysis on the effect of temperature, socio-economic and environmental factors on the spread and mortality rate of the COVID-19 pandemic. Environ Dev Sustain 2020;.

Ran J, Zhao S, Han L, Chong MKC, Qiu Y, Yang Y, et al. The changing patterns of COVID-19 transmissibility during the social unrest in the United States: a nationwide ecological study with a before-and-after comparison. One Health 2021;12:100201.

Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. Sustain Cities Soc 2020;62.

Saqib M. Forecasting COVID-19 outbreak progression using hybrid polynomial-Bayesian ridge regression model. Appl Intell 2020;.

Vadrevu KP, Eaturu A, Biswas S, Lasko K, Sahu S, Garg JK, et al. Spatial and temporal variations of air pollution over 41 cities of India during the COVID-19 lockdown period. Sci Rep 2020;10(1).

Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;395(10225):689–97.

Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci Total Environ 2020;724.

Zhu L, Liu X, Huang H, Avellan-Llaguno RD, Lazo MML, Gaggero A, et al. Meteorological impact on the COVID-19 pandemic: a study across eight severely affected regions in South America. Sci Total Environ 2020a;744.

Zhu Y, Xie J, Huang F, Cao L. Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Sci Total Environ 2020b;727.