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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The spatial clustering analysis of COVID-19 and its associated factors in mainland China at the prefecture level

Mengyang Liu 1, Mengmeng Liu 1, Zhiwei Li, Yingxuan Zhu, Yue Liu, Xiaonan Wang, Lixin Tao, Xiuhua Guo *

Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, Beijing, PR China
Beijing Municipal Key Laboratory of Clinical Epidemiology, Capital Medical University, Beijing, PR China

HIGHLIGHTS

• Population outflow from Wuhan city significantly increased the risk of COVID-19.
• Relative humidity and precipitation were positively associated with COVID-19 incidence.
• Average wind speed was negatively associated with COVID-19 incidence.
• NO2 was positively associated and O3 was negatively associated with COVID-19 incidence.

GRAPHICAL ABSTRACT

ARTICLE INFO

Article history:
Received 14 July 2020
Received in revised form 15 February 2021
Accepted 15 February 2021
Available online 20 February 2021
Editor: Scott Sheridan

Keywords:
COVID-19 incidence
Spatial clustering
Associated factors
Mainland China
Prefecture level

ABSTRACT

Coronavirus disease 2019 (COVID-19) has become a worldwide public health threat. Many associated factors including population movement, meteorological parameters, air quality and socioeconomic conditions can affect COVID-19 transmission. However, no study has combined these various factors in a comprehensive analysis. We collected data on COVID-19 cases and the factors of interest in 340 prefectures of mainland China from 1 December 2019 to 30 April 2020. Moran’s I statistic, Getis-Ord Gi* statistic and Kulldorff’s space-time scan statistics were used to identify spatial clusters of COVID-19, and the geographically weighted regression (GWR) model was applied to investigate the effects of the associated factors on COVID-19 incidence. A total of 67,449 laboratory-confirmed cases were reported during the study period. Wuhan city as well as its surrounding areas were the cluster areas, and January 25 to February 21, 2020, was the clustering time of COVID-19. The population outflow from Wuhan played a significant role in COVID-19 transmission, with the local coefficients varying from 14.87 to 15.02 in the 340 prefectures. Among the meteorological parameters, relative humidity and precipitation were positively associated with COVID-19 incidence, while the average wind speed showed a negative correlation, but the relationship of average temperature with COVID-19 incidence inconsistent between northern and southern China. NO2 was positively associated, and O3 was negatively associated, with COVID-19 incidence. Environment with high levels of inbound migration or travel, poor ventilation, high humidity or heavy rainfall, low temperature, and high air pollution may be favorable for the growth, reproduction and spread of SARS-CoV-2. Therefore, applying appropriate lockdown measures and travel restrictions, strengthening the ventilation of living and working environments, controlling air pollution and making sufficient preparations for a possible second wave in the relatively cold autumn and winter months may be helpful for the control and prevention of COVID-19.

© 2021 Published by Elsevier B.V.

* Corresponding author at: Department of Epidemiology and Health Statistics, School of Public Health, Capital Medical University, 10 Xi-Tou-Tiao, You-An-Men Street, Fengtai District, Beijing 100069, PR China.
E-mail address: statguo@ccmu.edu.cn (X. Guo).
1 Mengyang Liu and Mengmeng Liu contributed equally to this work.

https://doi.org/10.1016/j.scitotenv.2021.145992
0048-9697/© 2021 Published by Elsevier B.V.
1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has become a worldwide public health threat. We have identified the human-to-human transmission of SARS-CoV-2 based on clinical and epidemiological evidence (Shereen et al., 2020), and the infectivity of SARS-CoV-2 is greater than that of the previous two subtypes of coronavirus diseases, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) (Petrosillo et al., 2020). The WHO characterized COVID-19 as a pandemic on 11 March 2020 (WHO, 2020b), and there were 10,185,374 cases and 503,862 deaths across more than 200 countries by 30 June 2020 (WHO, 2020a). However, this pandemic is much more than a health crisis. The United Nations defined it as a human, economic and social crisis (UN, 2020). It is predicted that global trade could also fall by 13% to 32%, depending on the depth and extent of the global economic downturn (CRS, 2020). Although COVID-19 is still spreading in many countries around the world, China and some other Asian countries have essentially controlled the epidemic. Studying the epidemic of COVID-19 in China is of great importance, because it can give the public with an overall and comprehensive understanding of the disease.

Since COVID-19 is an emerging and serious infectious disease, it is critical to implement spatio-temporal surveillance, which can help prioritize locations for targeted interventions, rapid testing, and resource allocation. A study by Yang et al. explored the spatio-temporal patterns of the COVID-19 epidemic in Hubei Province and found that the spatial clusters were concentrated in Wuhan city (Yang et al., 2020). Other studies conducted at both the prefecture and county levels in Hubei Province and at the provincial level in mainland China also found similar results (Mo et al., 2020; Xiong et al., 2020). Several studies have been conducted to explore the spatial and temporal distribution of COVID-19 in mainland China, but their results were limited to the provincial level or to Hubei Province alone. Few studies have been launched at a more detailed spatial scale with appropriate statistical methods.

Many factors including population movement, meteorological parameters, air quality and socioeconomic conditions can affect COVID-19 transmission. A positive association was observed between population outflow from Wuhan and the number of COVID-19 cases (Liu et al., 2020b; Zhang et al., 2020). Many studies have explored the relationship between meteorological factors and COVID-19 transmission, especially temperature and humidity, but the results are contradictory. A multicity study in China identified a negative association such that each 1 °C increase in ambient temperature and each 1% increase in absolute humidity was related to a decline in daily confirmed case counts, with corresponding pooled RRs of 0.80 (95% CI: 0.75, 0.85) and 0.72 (95% CI: 0.59, 0.89), respectively (Liu et al., 2020a). Another study conducted in 166 countries worldwide reported similar results: a 1 °C increase in temperature or a 1% increase in relative humidity was associated with a 3.08% (95% CI: 1.53, 4.63%) or a 0.85% (95% CI: 0.51, 1.19%) reduction in daily new cases of COVID-19, respectively (Wu et al., 2020b). However, Xie et al. reported out the opposite result of the positive association in China (Xie and Zhu, 2020), while Pan et al. did not find significant relationships between meteorological factors and the R0 of COVID-19 in 202 locations of 8 countries (Pan et al., 2021). Air pollutants such as fine particulate matter (PM2.5) and inhalable particles (PM10) have also been shown to increase the risk of SARS-CoV-2 infection (Zhu et al., 2020). Moreover, increasing population density, construction land area proportion and aged population density were associated with an increased COVID-19 morbidity rate but gross domestic product (GDP) per unit of land area and hospital density were associated with a decreased COVID-19 morbidity rate in Wuhan, China (You et al., 2020). In summary, most previous studies only focused on one or a few limited influencing factors or limited regions (Hubei Province or Wuhan city), and their results were sometimes inconsistent. All factors can influence each other, and no associated factor can exist completely independently, but there is no study that has combined the various influencing factors together for analysis in mainland China.

This study aims to investigate the spatial clustering pattern of COVID-19 in mainland China at the prefecture level and explore the relationships between many factors, such as population movement, meteorological parameters, air quality factors and socioeconomic conditions, and the incidence of COVID-19.

2. Material and methods

2.1. Data collection

2.1.1. COVID-19 case data

The laboratory COVID-19 case data for 340 prefectures of mainland China (Fig. S1 and Table S1) were extracted from the R package “nCov2019” (Wu et al., 2020a). This package aggregated global COVID-19 data from four different sources and provides real-time and historical data on COVID-19 in every country worldwide, along with daily confirmed case counts. The four different data sources including China National Health Commission (CNHC, 2020), Tencent SARS-CoV-2 website (Tencent, 2020), a non-governmental organization Dingxiangyuan (DXY, 2020) and a public GitHub repository (GitHub Repo, 2020). Of course, these four data sources are based on the official website including the WHO, China National Health Commission, Chinese provincial and city health agencies, and public health agencies in other countries (Wu et al., 2020a). The authors of this package collect, check and synthesize the information from the above sources and determine the final database. The epidemic data in China mainly originated from the National Health Council and the provincial and city health committees from the late December 2019 to 30 April 2020. Cases imported from other countries were excluded from the study. In addition, the epidemiological investigation from Chinese CDC experts indicated that the onset date of the patients can be traced back to 8 December 2019 (Wuhan MHC, 2020). In order to consider the associated factors before the onset, we have incorporated influencing factors such as meteorology and air pollutants since 1 December 2019.

2.1.2. Meteorological data

Meteorological data in every prefecture were obtained from the China Meteorological Data Sharing Service System (https://data.cma.cn/en), which is the authoritative platform on which the China Meteorological Administration shares its meteorological data resources with the public. We obtained the daily data of average temperature (AT, °C), average wind speed (AWS, m/s), average relative humidity (ARH, %), average atmospheric pressure (AAP, hPa) and average precipitation (AP, mm) from 1 December 2019 and 30 April 2020, covering every prefecture in mainland China. The average concentrations of six pollutants as well as the air quality index (AQI) were used in this study, namely, PM2.5, PM10, carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and ozone (O3).

2.1.3. Air pollutant data

Air pollution data were obtained from more than 1600 fixed-site air quality monitoring stations from the China National Environmental Monitoring Centre (http://www.cnemc.cn/) between 1 December 2019 and 30 April 2020, covering every prefecture in mainland China. The average concentrations of six pollutants as well as the air quality index (AQI) were used in this study, namely, PM2.5, PM10, carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2) and ozone (O3).

2.1.4. Population movement data

We obtained the daily proportion of the population outflow from Wuhan to other cities (%) and population inflow from other cities to Wuhan (%) from 10 January to 15 March 2020, which covered the China Spring Festival travel period, from data released by Baidu Qianxi. Baidu Qianxi provides real-time dynamic information on regional population movements based on Baidu location-based services (LBS) and
Baidu Tianyan. Data from Baidu Qianxi during the Spring Festival are freely available to the public (https://qianxi.baidu.com).

### 2.1.5. Socioeconomic data

Socioeconomic factors were obtained from the China Statistical Yearbook 2020 released by the National Bureau of Statistics (http://data.stats.gov.cn/) for every province. The factors collected included population density, proportion of people aged 65 and above, GDP per capita, number of beds per 10,000 people, number of health technicians per 10,000 people and number of high schools and colleges per 100,000 people, which covered all aspects, such as population, health care, education and so on. It is worth noting that these data reflected the overall condition of each province in 2018 rather than 2020. In addition, the population data in every prefecture in mainland China were also collected from the China regional economic statistical yearbook to calculate the COVID-19 incidence.

### 2.2. Statistical analysis

#### 2.2.1. The spatial clustering analysis

First, the global spatial autocorrelation analysis method global Moran’s I index was used to check whether there was global spatial autocorrelation over the entire region regarding COVID-19 incidence and the associated factors. The value of Moran’s I ranges from −1 to 1. A higher positive Moran’s I indicates that values in neighboring positions tend to cluster, while a lower negative Moran’s I implies that higher and lower values are interspersed. When Moran’s I is near 0, there is no spatial clustering, meaning that the data are randomly distributed. The Z-score and p-value were calculated to evaluate the significance of the global Moran’s I index (Huo et al., 2012; Jiang and Zhao, 2021). Then, local spatial autocorrelation analysis methods, including Anselin Local Moran’s I statistic and Getis-Ord Gi* statistic, were used to analyse the clusters of COVID-19 incidence and to identify COVID-19 incidence hotspots in the study period. Similar to the global Moran’s I index, the value of Anselin Local Moran’s I statistic also range from −1 to 1, and the Z-score and p-value were used to test whether it was statistically significant. The results of Anselin Local Moran’s I statistic are presented as spatial clusters (high-high clusters and low-low clusters) and spatial outliers (high-low outliers and low-high outliers) in specific locations (Anselin, 1995). The Getis-Ord Gi* statistic can be used as a measure of the degree of spatial clustering of COVID-19 incidence in a study location. The degree of clustering and its statistical significance are estimated based on a confidence level according to the Z (Gi*)-score and P-value. Districts with Z-scores > 2.58, and Z-scores between 1.96 and 2.58 and 1.65–1.96 were considered to be significant at the 99% confidence level (P < 0.01), 95% confidence level (P < 0.05) and 90% confidence level (P < 0.10), respectively. A positive and significant Z score indicates a hotspot; on the contrary, a negative and significant Z score indicates a cold spot, and a Z (Gi*) score close to zero indicates no obvious spatial clustering (Lu et al., 2019; Peeters et al., 2015). Finally, Kulldorff’s space-time scan statistical analysis was used to explore the spatial and temporal clusters of COVID-19. The space–time scan statistics were defined by a potential cylindrical window in which the bottom of the cylinder corresponds to space and the height of the cylinder corresponds to time. Then, the cylindrical window was moved in dynamic space and time to find the potential geographical locations and possible time periods. The most likely cluster is the window with the maximum likelihood, and secondary clusters are also presented if they are statistically significant. The Monte Carlo method was used to estimate the P value with the number of replications limited to 999, and the relative risk was calculated as the ratio of the estimated risk within the cluster and outside the cluster (Liu et al., 2018).

#### 2.2.2. The geographically weighted regression (GWR) model

Before the regression analysis of associated factors and COVID-19 incidence, the Lasso dimension reduction method was first used to screen the essential factors from among all kinds of factors; this could also avoid the problem of collinearity of similar variables. Then, the geographically weighted regression (GWR) model was used to explore the spatial variation in the relationship between COVID-19 incidence and essential associated factors. Since the spatial positions of the data were taken into consideration in the GWR model,

### Table 1

The COVID-19 incidence and associated factors in 340 prefectures of mainland China from 1 December 2019 to 30 April 2020.

| Variable                        | Min    | P25   | P50   | P75   | Max    | Range  |
|---------------------------------|--------|-------|-------|-------|--------|--------|
| COVID-19 incidence (1/1,000,000) | 0.00   | 0.45  | 1.07  | 2.20  | 1513.16 | 1.74   |
| Population movement             |        |       |       |       |        |        |
| Population outflow from Wuhan to other cities (%) | 0.00  | 0.00  | 0.01  | 0.09  | 13.20  | 0.09   |
| Population inflow from other cities to Wuhan (%) | 0.00 | 0.00  | 0.01  | 0.06  | 100.00 | 0.06   |
| Meteorological factors          |        |       |       |       |        |        |
| Average temperature (°C)        | −13.54 | 1.19  | 8.21  | 11.66 | 22.44  | 10.47  |
| Average relative humidity (%)   | 30.36  | 57.21 | 67.87 | 75.87 | 85.25  | 18.66  |
| Average wind speed (m/s)        | 0.99   | 1.82  | 2.11  | 2.45  | 380.77 | 0.64   |
| Average atmospheric pressure (hPa) | 585.10 | 916.00 | 995.30 | 1013.60 | 1023.30 | 302.18 |
| Average precipitation (mm)      | 0.00   | 0.13  | 111.72| 302.31| 5162.20| 302.18 |
| Air pollutants                  |        |       |       |       |        |        |
| AQI                             | 22.99  | 50.28 | 66.34 | 85.79 | 211.24 | 35.51  |
| PM2.5 (μg/m³)                   | 5.64   | 30.91 | 42.88 | 57.45 | 153.92 | 28.54  |
| PM10 (μg/m³)                    | 12.32  | 49.90 | 66.38 | 90.33 | 355.68 | 40.47  |
| CO (mg/m³)                      | 0.29   | 0.69  | 0.80  | 0.96  | 2.06   | 0.26   |
| SO₂ (μg/m³)                     | 3.06   | 7.54  | 10.04 | 14.40 | 42.33  | 6.86   |
| NO₂ (μg/m³)                     | 5.83   | 19.12 | 25.38 | 33.08 | 54.21  | 13.96  |
| O₃ (μg/m³)                      | 27.46  | 48.60 | 55.44 | 62.06 | 88.31  | 13.46  |

| Socioeconomic factors           |        |       |       |       |        |        |
| Proportion of population aged 65 and above | 2.10  | 109.00| 247.00| 429.00| 3814.00| 320.00 |
| GDP per capita                   | 30,797 | 43,366 | 52,114 | 66,472 | 153,095 | 23,106 |
| Number of beds per 10,000 people | 43.75  | 55.84 | 61.71 | 66.51 | 72.14  | 10.67  |
| Number of health technicians per 10,000 people | 0.00 | 62.00 | 67.00 | 71.00 | 119.00 | 9.00   |
| Number of high schools and colleges per 100,000 people | 4124.00 | 5208.00 | 5311.00 | 5585.00 | 6633.00 | 377.00 |

Note:
- The COVID-19 incidence (1/1,000,000) was calculated from the late December 2019 to 30 April 2020.
- The “Population outflow from Wuhan to other cities (%)” was calculated from 10 January to 15 March 2020.
- The “Population inflow from other cities to Wuhan (%)” was calculated from 10 January to 15 March 2020.
- “Socioeconomic factors” in this study described the socioeconomic condition at the provincial level in 2018.
the local parameters changed with the spatial position. A conventional GWR model can be described by the following equation:

\[ y_j = \beta_0(\mu_j, \nu_j) + \sum_{i=1}^{p} \beta_i(\mu_j, \nu_j)X_{ij} + \epsilon_j \]

where \((\mu_j, \nu_j)\) is the spatial coordinate of sample point \(j\); \(\beta_0(\mu_j, \nu_j)\) and \(\beta_i(\mu_j, \nu_j)\) are the regression constant and the regression coefficient of sample point \(j\), respectively; and \(\epsilon_j\) is the random error of the independent distribution. Considering that the samples were not regularly spaced in 340 prefectures, the Golden section search method was used to identify the optimal bandwidth, and the adaptive Gaussian kernel was implemented to build the model (Brunsdon et al., 2010).

The global spatial autocorrelation analysis, Anselin Local Moran’s I statistic and the Getis-Ord Gi* statistic calculations were conducted by ArcGIS 10.2 software (ESRI Inc. Redlands, CA, USA) using the Spatial Autocorrelation packages. Kulldorff’s space-time scan statistical analysis was conducted with SaTScan 9.3 software. The GWR model was built with the GWR 4.0 software. All statistical tests were two-sided, and \(P < 0.05\) was considered statistically significant.

### 3. Results

#### 3.1. Characteristics of COVID-19 incidence and associated factors in 340 prefectures of mainland China

There were 67,449 laboratory confirmed COVID-19 cases included in the study from the late December 2019 to 30 April 2020. The incidence rate varied from 0 to 1513.16, with a median of 1.07 per million in 340 prefectures (Table 1, Fig. 1a). The distribution of associated factors including population movement, five meteorological factors, six air pollutants and AQI at the prefecture level, and six socioeconomic factors at the province level are also shown in Table 1.

#### 3.2. COVID-19 clustering analysis

The regions with the highest COVID-19 incidence were located in Wuhan and its surrounding cities, which are all located in Hubei Province (Fig. 1a). The global Moran's I statistic indicated that the COVID-19 cases were not randomly distributed in the 340 prefectures (\(z =\)
Fig. 2. The distribution of associated factors included in the GWR model in 340 prefectures of mainland China from 1 December 2019 to 30 April 2020. Note: The time periods of (a)–(b) are from 10 January to 15 March 2020; the time periods of (c)–(h) are from 1 December 2019 to 30 April 2020.
9.98, P < 0.05). The Anselin Local Moran’s I statistic and the Getis-Ord Gi* statistic showed similar results, namely, the Wuhan-centered areas were high-high cluster areas (Fig. 1b), or the hotspot areas (Fig. 1c), especially in Hubei Province, Anhui Province, the south of Henan Province, the north of Hunan and Jiangxi Province. Ten cluster areas were found by Kulldorff’s space-time scan statistical analysis. Aside from the most likely cluster in Wuhan city, it should be noted that some metropolises, such as Beijing, Chongqing, Shanghai and coastal cities, such as Shenzhen and Dongguan city in Guangdong Province were also scattered cluster regions (Fig. 1d). For temporal patterns, the clustering time of the most likely cluster ranged from 4 February 2020 to 18 February 2020 while nine secondary clusters occurred from 25 January 2020 to 21 February 2020 (Fig. S2).

### 3.3. The relationship between COVID-19 and its associated factors

Lasso regression analysis yielded seven associated factors including population outflow from Wuhan, population inflow to Wuhan, AWS, ARH, AP, NO2 and O3. Numerous studies have reported that temperature is also an important factor influencing of COVID-19 (Prata et al., 2020; Wu et al., 2020b; Xie and Zhu, 2020), so the seven previous parameters and average temperature were modeled as local variables in the GWR analysis (Fig. 2a–h), and the socioeconomic factors at the provincial level were modeled as global variables based on their spatial scale and global Moran’s I index (P > 0.05) (Table S2).

The GWR model improved the fitting effect with adjusted R2 = 0.9224 and Akc = 3110.45, compared with the ordinary least squares corresponding global coefficients. The local GWR coefficients of population density, proportion of people aged 65 and older, GDP per capita, number of beds per 10,000 people, number of health technicians per 10,000 people and number of high schools and colleges per 100,000 people were −0.003, 0.027, 0.07093, −0.134, 0.018 and −0.000662, respectively.

### 4. Discussion

Our study not only explored the spatial distribution of COVID-19, but also examined the association between COVID-19 incidence and associated factors including population movement, meteorological parameters, air pollutants and socioeconomic conditions in 340 prefectures of mainland China from the late December 2019 to 30 April 2020. Our study used the GWR model, which gave a relatively full profile of the spatial distribution and impact of influencing factors on COVID-19 at the detailed prefecture level. In addition, Lasso regression was used to screen essential factors as well as to avoid collinearity of variables, which serves to provide more credible and robust results.

Several spatial analysis methods were used in this study int 340 prefectures in mainland China, and all the results indicated that the areas centered in Wuhan city or Hubei Province were COVID-19 aggregation areas. A case of pneumonia of unknown cause was detected in Wuhan at the end of 2019, and China reported this to the WHO Country Office in China on 31 December 2019. In fact, Wuhan city and Hubei Province were the most heavily impacted by the COVID-19 epidemic in China, respectively accounting for 52.33% (35,295/67449) and 79.36% (53,527/67449) of the total cases in the study period. Aside from Wuhan and its surrounding areas, several other metropolises were also COVID-19 cluster regions according to Kulldorff’s space-time scan analysis, which indicated that well-developed public transportation and large population movement may increase the infection risk (Liu et al., 2020b; Zhang et al., 2020). The clustering time of the most likely cluster as well as the other nine secondary clusters was from the end of January to the middle of February 2020. A national epidemiological analysis of COVID-19 also found that the epidemic reached the turning point in both Hubei and non-Hubei regions of China after on 12 February 2020, which indicated that China’s epidemic gradually came under control after mid-February (CCDC, 2020).

Our study found that population outflow from Wuhan may play an important role in COVID-19 transmission in China. Studies from You and Liu also provided similar results, which suggests that early implementation of lockdown measures could be more effective (Liu et al., 2020b; You et al., 2020). More specifically, Zhang et al.’s study pointed out that the frequencies of air flights and high-speed train (HST) services from Wuhan were significantly associated with the number of COVID-19 cases in destination cities after studying common travel patterns, and the presence of an airport or HST station in a city can accelerate the spread the COVID-19 pandemic, so air flights and HST are the main modes of transportation associated with COVID-19 transmission (Zhang et al., 2020). However, most of the previous studies only focused on population movement without considering the impact of other related factors, unlike our study.

For the meteorological factors, the relationship between temperature and COVID-19 incidence was complex and varied from region to region. In northern China, with a relatively low temperature, temperature was positively correlated with COVID-19 incidence, while in southern China, with a relatively high temperature, there was a negative correlation between temperature and COVID-19 incidence. We speculated that different temperature ranges have different effects on SARS-CoV-2 infection and that their relationship may not be linear but rather an inverted U shape. The SARS-CoV-2 infectivity increased with increasing temperature in the lower temperature range on the left side of the inverted U shape, while the infectivity decreased with increasing temperature in the higher temperature range on the right side of the inverted U shape. The left and right sides of the inverted U may correspond to the north and the south of China, respectively. Studies based on either epidemiological surveys or laboratory experiments also

### Table 2

The local GWR coefficient between reported COVID-19 incidence and influencing factors in 340 prefectures of mainland China from 1 December 2019 to 30 April 2020.

| Variable                                      | Min     | P25    | P50    | P75    | Max     | Range   |
|-----------------------------------------------|---------|--------|--------|--------|---------|---------|
| Intercept                                     | −16.71  | −11.38 | −6.05  | −0.26  | 3.48    | 20.19   |
| Population outflow from Wuhan to other cities (%)<sup>a</sup> | 14.87   | 14.90  | 14.92  | 14.96  | 15.02   | 0.14    |
| Population inflow from other cities to Wuhan (%)<sup>a</sup> | −11.27  | −10.04 | −9.30  | −7.77  | −5.19   | 6.08    |
| Average temperature (°C)                      | −0.16   | −0.09  | 0.02   | 0.12   | 0.20    | 0.35    |
| Average wind speed (m/s)                     | −0.05   | −0.05  | −0.04  | −0.04  | −0.03   | 0.02    |
| Average relative humidity (%)                | 0.02    | 0.06   | 0.12   | 0.18   | 0.21    | 0.19    |
| Average precipitation (mm)                   | 0.01    | 0.03   | 0.04   | 0.04   | 0.05    | 0.04    |
| NO2 (μg/m³)                                   | 0.09    | 0.13   | 0.16   | 0.19   | 0.21    | 0.13    |
| O3 (μg/m³)                                    | −0.19   | −0.17  | −0.12  | −0.07  | −0.02   | 0.17    |

Notes:
<sup>a</sup> The “Population outflow from Wuhan to other cities (%)” was calculated from 10 January to 15 March 2020.
<sup>b</sup> The “Population inflow from other cities to Wuhan (%)” was calculated from 10 January to 15 March 2020.
Fig. 3. The spatial mapping of local regression coefficients in 340 prefectures based on the GWR model from 1 December 2019 to 30 April 2020. Note: The time periods of (a)–(b) are from 10 January to 15 March 2020; the time periods of (c)–(h) are from 1 December 2019 to 30 April 2020.
supported this result (Casanova et al., 2010; Prata et al., 2020; Xie and Zhu, 2020). For example, a national study based on 122 Chinese cities observed that when the mean temperature was lower than 3 °C in the past two weeks, the daily confirmed COVID-19 cases increased by 4.86% with each 1 °C increase (Xie and Zhu, 2020), but in the tropical cities of Brazil, each 1 °C increase in AT was associated with a 4.88% decrease in the number of daily COVID-19 cases (Prata et al., 2020). Of course, the relationship between temperature and COVID-19 incidence is complex, and whether there is an inverted U-shaped relation and the optimum temperature of SARS-CoV-2 need further research for confirmation. Densely populated and poorly ventilated environments, whether indoors or outdoors, could also increase the risk of airborne diseases such as SRAR, tuberculosis and SARS-CoV-2 infection (Li et al., 2020b; Olsen et al., 2003; Zhang et al., 2019). Conversely, wind blowing can speed up air flow and virus dispersion, thereby reducing the COVID-19 incidence rate. Both ARH and AP were positively associated with COVID-19 incidence, perhaps because SARS-CoV-2 needs a relatively humid environment for growth and reproduction (Yao et al., 2020).

The first COVID-19 case was detected in the Huanan Seafood Wholesale Market of Wuhan and another wave of the epidemic was recently detected in the Xinfadi Wholesale Market of Beijing, which supports the theory that a wet environment is favorable for the transmission of COVID-19 (Li et al., 2020a; Wang et al., 2020). A climate-dependent epidemic model about the SARS-CoV-2 pandemic in different scenarios based on known coronavirus biology suggested that strong outbreaks were likely in more humid climates (Baker et al., 2020). Research on MERS and other forms of coronavirus also proved that the virus started to decay rapidly at a lower humidity level (Pyankov et al., 2018). Other studies presented the opposite results for ARH, but they only controlled for a limited number of meteorological factors (Qi et al., 2020); our analysis controlled for many confounding factors and should be more robust.

For the air pollutant parameters, higher O3 may decrease the SARS-CoV-2 infection risk, and this may be attributed to two aspects. First, many previous studies have found that O3 has a broad sterilization spectrum, and can deactivate various pathogens regardless of the virus, bacteria, fungus or protozoan (Elvis and Ekta, 2011). For example, a previous study reported that a high concentration of ozone in water of 27.73 mg/L can completely inactivate the SARS virus within 4 min, and medium and low concentrations could also deactivate the virus with different levels of efficacy (Zhang et al., 2004). Second, a moderate dose of O3 could trigger an entire cascade of immune reactions by producing a large amount of inflammatory cytokines including interferon, interleukin and tumor necrosis factor, so O3 was used to treat many kinds of diseases and may become a protective factor against COVID-19 (Elvis and Ekta, 2011). In addition, our research found that NO2 was positively related to COVID-19. Studies from the United States and Europe also observed a positive relationship between NO2 and the COVID-19 case-fatality rate and mortality rate (Liang et al., 2020; Ogen, 2020). In the US, each 4.6 ppb increase in NO2 was related to an increase of 7.1% (95% CI 1.2% to 13.4%) and 11.2% (95% CI 3.4% to 19.5%) in the COVID-19 case-fatality rate and mortality rate, respectively (Liang et al., 2020). As a highly reactive exogeneous oxidant, NO2 can enhance oxidative stress and induce inflammatory reactions, generating reactive nitrogen and oxygen species, which may eventually damage the immune system and make the human body more susceptible to SARS-CoV-2 infection (Bevelander et al., 2007). Of course, the correlation does not mean causality. Further studies should be conducted to prove the causal relationship between the influencing factors and COVID-19.

This study has several limitations. First, we could not obtain detailed socioeconomic information for each prefecture, only for the provincial level in 2019, so they were fitted into the model as global variables. Second, this was a cross-sectional study with all data averaged in the study period, so the time effects were ignored, and causality cannot be verified. Third, the population movement data were limited to the Spring Festival travel rush period, which was from 10 January to 15 March. This did not cover the whole study period, but it fully represented the population flow situation in the study period for the following reasons: 1) the Spring Festival in China is often accompanied by large-scale population movements and the travel rush in 2020 started on 10 January, so the number of population movement began to grow dramatically from this moment until the end of the holiday; 2) a lockdown measure was implemented in Wuhan city from 23 January 2020 until 8 April 2020, so the population outflows from Wuhan and population inflows from other cities to Wuhan were sufficiently small (except for some special personnel such as medical staff) to be ignored; and 3) we used the average proportion of population movements rather than the absolute number of people, which was relatively stable for a particular region.

5. Conclusions

In conclusion, our study found that a variety of associated factors including population movement, meteorological parameters, air pollutants and socioeconomic conditions can affect COVID-19 incidence, which can help policymakers conduct targeted interventions to mitigate further transmission especially in countries severely affected by the epidemic. A large number of in-migrants from the epidemic focus obviously increased the COVID-19 risk. Additionally, meteorological factors and air quality factors including lower temperature, poor ventilation, wet and rainy environments and heavy air pollution, especially NO2 pollution, could also increase the risk of SARS-CoV-2 infection. Accordingly, taking appropriate lockdown measures and travel restrictions, strengthening ventilation of living and working environments, and controlling air pollution may help prevent further COVID-19 outbreaks. Furthermore, sufficient preparations should be made in advance to prevent a possible second wave of COVID-19 in the relatively cold autumn and winter months.

Abbreviations

COVID-19 coronavirus disease 2019
SARS-CoV-2 severe acute respiratory syndrome coronavirus 2
SARS severe acute respiratory syndrome
MERS Middle East respiratory syndrome
PM2.5 fine particulate matter
PM10 inhalable particles
GDP gross domestic product
AT average temperature
AWS average wind speed
ARH average relative humidity
AAP average atmospheric pressure
AP average precipitation
AHI air quality index
CO carbon monoxide
SO2 sulfur dioxide
NO2 nitrogen dioxide
O3 ozone
LBS location-based services
GWR geographically weighted regression
OLS ordinary least squares regression
HST high-speed train

CRediT authorship contribution statement

Mengyang Liu: Writing original draft, Data collection and cleaning, Methodology, Visualization. Mengmeng Liu: Writing - review & editing, Data collection and cleaning, Methodology, Software. Zhiwei Li: Data collection, Methodology. Yingxuan Zhu: Data collection, Methodology. Yue Liu: Data collection, Methodology. Xiaonan Wang: Writing - review & editing, Supervision. Lixin Tao: Writing - review & editing,
Project administration, Supervision. Xiuhua Guo: Writing - review & editing, Conceptualization, Project administration, Supervision.

Funding source
No funder supported this work.

Ethical approval
Not applicable.

Declaration of competing interest
The authors declare no conflicts of interest.

Acknowledgements
COVID-19 data were obtained from the National Health Council and the provincial and city health committees of China. We thank all of the healthcare personnel who contributed to the detection, epidemiological investigation, and diagnosis of COVID-19 cases.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.145992.

References
Anselin, L. 1995. Local indicators of spatial association—LISA. Geogr. Anal. 27, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x.
Baker, R.E., Yang, W., Vecchi, G.A., Metcalf, C.E.J., Grenfell, B.T. 2020. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. Science. 369, 315–319. https://doi.org/10.1126/science.abd6255.
Bevelander, M., Mattey, J., Whittaker, L.A., Paveglio, S.A., Jones, C.C., Robbins, J., Hemenway, D., Akira, S., Uematsu, S., Poynter, M.E., 2007. Nitrogen dioxide promotes allergic sensitization to inhaled antigen. J. Immunol. 179, 3680–3688. https://doi.org/10.4049/jimmunol.179.17.3680.
Bruunsd, C., Fotheringham, A.S., Charlton, M.E., 2010. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28, 281–298. https://doi.org/10.1111/j.1538-4632.1995.tb00308.x.
Casanova, L.M., Jeon, S., Rutala, W.A., Weber, D.J., Sobsey, M.D., 2010. Effects of air temperature and relative humidity on COVID-19 transmission in Wuhan and geographical expansion of the 2019-nCoV pneumonia in China in January 2020. Clin. Infect. Dis. https://doi.org/10.1093/infdis/jiaa422 (Published Ahead of Print).
Cu, Q., Wu, H., Ding, Z., Wu, C., Lin, J., 2019. Analysis of epidemiological characteristics of 2019-nCoV pneumonia in Zhejiang Province, China, 2020-01-22. Sci. Total Environ. 726, 138513. https://doi.org/10.1016/j.scitotenv.2020.138513.
Ku, A., Song, S., Zhu, G., Tian, F., Li, H., Gao, Y., Wu, Y., Zhang, Sh., Zou, L., Qin, Q., Lin, M., 2020. Impact of meteorological factors on COVID-19 transmission: a multi-city study in China. Sci. Total Environ. 726, 138650. https://doi.org/10.1016/j.scitotenv.2020.138650.
Mo, C., Tan, M., Bei, T., Cai, Q., Jia, P., Wang, Z., Zhang, Z., 2020. An analysis of spatiotemporal pattern for COVID-19 in China based on space-time cube. J. Med. Virol. 92, 1587–1595. https://doi.org/10.1002/jmv.25834.
Ogen, Y., 2020. Assessing nitrogen dioxide (NO(_2)) levels as a contributing factor to coronavirus (COVID-19) fatality. Sci. Total Environ. 726, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605.
Olsen, S.J., Chang, H.L., Cheung, Y.T., Tang, A.F., Fisk, T.L., Ooi, S.P., Kuo, H.W., Jiang, D.D., Chen, K.T., Landis, J., Hsu, R.H., Chen, T.J., Dowell, S.F., 2003. Transmission of the severe acute respiratory syndrome coronavirus to healthcare workers. N. Engl. J. Med. 349, 2416–2422. https://doi.org/10.1056/NEJMoa031349.
Pan, J., Yao, Y., Li, Z., Zhang, X., Guo, Y., Qi, Y., Wang, W., Zhang, L., Wang, W., Han, H., 2021. Weather pattern unlikely to reduce the COVID-19 transmission: an ecological study in 202 locations in 8 countries. Sci. Total Environ. 753, 142272. https://doi.org/10.1016/j.scitotenv.2020.142272.
Peeters, A., Zade, M., Kalthun, J., Ilihi, M., Karber, R., Hetzel, M., Gebbers, R., Ben-Gal, A., 2015. Gatis-Ors’s hot- and cold-spot statistics as a basis for multivariable spatial clustering of orchard tree data. Comput. Electron. Agric. 111, 140–150. https://doi.org/10.1016/j.compag.2014.12.011.
Petrosillo, N., Viogelang, G., Ergonul, O., Petersen, E., 2020. COVID-19, SARS and MERS: are they closely related? Clin. Microbiol. Infect. 26, 729–734. https://doi.org/10.1016/j.cmi.2020.03.026.
Prata, D.N., Rodrigues, W., Bermejo, P.H., 2020. Temperature signifi cantly changes COVID-19 transmission in (sub) tropical cities of Brazil. Sci. Total Environ. 729, 138862. https://doi.org/10.1016/j.scitotenv.2020.138862.
Pandian, A.O.V., Bodew, S.A., Parys, O.C., Ogene, Y., 2018. Survival of aerosolized coronavirus in the ambient air. J. Aerosol Sci. 115, 158–163. https://doi.org/10.1016/j.jaerosci.2017.09.009.
Qi, H., Xiao, S., Shi, R., Ward, M.P., Chen, Y., Tu, W., Su, Q., Wang, W., Wang, X., Zhang, Z., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. Sci. Total Environ. 728, 138778. https://doi.org/10.1016/j.scitotenv.2020.138778.
Shereen, M.A., Khan, S., Kazmi, A., Bashir, N., Siddique, R., 2020. COVID-19 infection: origins, transmission, and MERS: are they closely related? F1000Res. 9, e91–98. https://doi.org/10.12688/f1000research.20033.01.
Tencent, 2020. Real-time tracking of the coronavirus infection. (In Chinese). https://news.qq.com/zt2020/page/feiyan.htm. (Accessed 15 February 2021).
UN (the United Nations). 2020. The social impact of COVID-19. https://www.un.org/development/desa/dsd/psd/2020/04/social-impact-of-covid-19. (Accessed 15 February 2021).
Wang, Q., Zhao, X., Yue, Z., Bai, X., Rao, H., Han, N., Wang, T., Wang, H., Jiang, B., 2020. Increasing SARS-CoV-2 nucleic acid testing capacity during the COVID-19 epidemic in Beijing: experience from a general hospital. Emerg. Microbes Infect. 9, 2358–2360. https://doi.org/10.1016/j.eoor.2020.03.005.
WHO (World Health Organization), 2020b. COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis. Sci. Total Environ. 728, 138778. https://doi.org/10.1016/j.scitotenv.2020.138778.
Wu, T., Shen, X., Hu, Y., Taylor, S., Leung, K., Li, Y., Wang, C., 2020. Evidence for enhanced SARS-CoV-2 transmission of SARS-CoV-2 in a poorly ventilated restaurant. medRxiv https://doi.org/10.1101/2020.04.12.20046726.
Xie, J., Zhu, Y., 2020. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci. Total Environ. 724, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201.
Xiong, Y., Yang, W., Chen, F., Zhuo, M. 2020. Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. Int. J. Environ. Res. Public Health 17, 5903. https://doi.org/10.3390/ijerph17115903.
Yang, W., Deng, M., Li, C., Huang, J., 2020. Spatio-temporal patterns of the 2019-nCoV epidemic at the county level in Hubei Province, China. Int. J. Environ. Res. Public Health 17, 2563. https://doi.org/10.3390/ijerph17072563.

Yao, M., Zhang, L., Ma, J., Zhou, L., 2020. On airborne transmission and control of SARS-CoV-2. Sci. Total Environ. 731, 139178. https://doi.org/10.1016/j.scitotenv.2020.139178.

You, H., Wu, X., Gao, X., 2020. Distribution of COVID-19 morbidity rate in association with social and economic factors in Wuhan, China: implications for urban development. Int. J. Environ. Res. Public Health 17, 3417. https://doi.org/10.3390/ijerph17103417.

Zhang, J.M., Zheng, C.Y., Xiao, G.F., Zhou, Y.Q., Gao, R., 2004. Examination of the efficacy of ozone solution disinfectant inactivating SARS virus. Chin. J. Disinfection 21, 27–28. https://doi.org/10.3969/j.issn.1001-7658.2004.01.009 (in Chinese).

Zhang, Y., Liu, M., Wu, S.S., Jiang, H., Zhang, J., Wang, S., Ma, W., Li, Q., Ma, Y., Liu, Y., Feng, W., Amsalu, E., Li, X., Wang, W., Li, W., Gao, X., 2019. Spatial distribution of tuberculosis and its association with meteorological factors in mainland China. BMC Infect. Dis. 19, 379. https://doi.org/10.1186/s12879-019-4008-1.

Zhang, Y., Zhang, A., Wang, J., 2020. Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. Transp. Policy (Oxf) 94, 34–42. https://doi.org/10.1016/j.tranpol.2020.05.012.

Zhu, Y., Xie, J., Huang, F., Cao, L., 2020. Association between short-term exposure to air pollution and COVID-19 infection: evidence from China. Sci. Total Environ. 727, 138704. https://doi.org/10.1016/j.scitotenv.2020.138704.