Spatial Characteristics of Coronavirus Disease 2019 and Their Possible Relationship With Environmental and Meteorological Factors in Hubei Province, China

Xiaochi Huang1, Han Zhou1,2, Xiaofeng Yang1, Wen Zhou1, Jiejun Huang1, and Yanbin Yuan1

1School of Resource and Environmental Engineering, Wuhan University of Technology, Wuhan, China, 2School of Energy and Environment, Guy Carpenter Asia–Pacific Climate Impact Centre, City University of Hong Kong, Hong Kong, China, 3Wuhan Regional Climate Center, Hubei Meteorological Service, Wuhan, China

Abstract As of July 27, 2020, COVID-19 has caused 640,000 deaths worldwide and has had a major impact on people’s productivity and lives. Analyzing the spatial distribution characteristics of COVID-19 cases and their relationships with meteorological and environmental factors might help enrich our knowledge of virus transmission and formulate reasonable epidemic prevention strategies. Taking the cumulative confirmed cases in Hubei province from January 23, 2020, to April 8, 2020, as an example, this study analyzed the spatial evolution characteristics of confirmed COVID-19 cases in Hubei province using exploratory spatial data analysis and explored the spatial relationship between the main environmental and meteorological factors and confirmed COVID-19 cases using a geographically weighted regression (GWR) model. Results show that there was no obvious spatial clustering of confirmed COVID-19 cases in Hubei province, while the decline and end of the newly confirmed cases revealed relatively obvious negative spatial correlations. Due to the lockdown in Hubei province, the main air quality indexes (e.g., AQI and PM2.5) decreased significantly and environmental quality was better than historical contemporaneous levels. Meanwhile, the results of the GWR model suggest that the impacts of environmental and meteorological factors on the development of COVID-19 were not significant. These findings indicate that measures such as social distancing and isolation played the primary role in controlling the development of the COVID-19 epidemic.

1. Introduction

Novel coronavirus-infected pneumonia (NCIP) is a new infectious disease caused by COVID-19. Many studies have shown that COVID-19 is more infectious than SARS-CoV (SARS) and other viruses. According to a WHO report, as of July 27, 2020, there have been more than 16 million confirmed cases worldwide and more than 640,000 people have died. According to data released by the National Health Commission, as of July 27, 2020, China has reported a total of 87,028 confirmed cases, including 4,659 deaths. Hubei province accounted for 78.29% of the country’s cumulative confirmed cases.

As of December 8, 2019, the first case of NCIP occurred in Wuhan, Hubei province. On December 27, the Wuhan Municipal Health Commission reported that the disease might be a new type of coronavirus. On January 7, 2020, the new coronavirus was isolated. On January 19, the government established a high-level group of experts, with academician Zhong Nanshan as the team leader. On January 20, COVID-19 was listed as a Class B infectious disease in China. The government adopted Class A infectious disease prevention and control measures to prevent NCIP. The number of confirmed COVID-19 cases grew rapidly. As of January 22, a total of 571 confirmed cases had been reported. The government decided to lock down the entire city to block the spread of the epidemic. As the most severely affected area in China, Wuhan city began lockdown on January 23. The cumulative confirmed cases in Wuhan continued to gradually increase after January 23. After March, the number of confirmed cases in Wuhan gradually stabilized. Other cities gradually achieved daily zero growth, and on April 5, Wuhan achieved this goal. China gradually prepared to start production and work.

The disease is spread to susceptible people mainly through close-range droplets, contact, and aerosols. At present, new cases nationwide have tended to be at a low level, but epidemic prevention and control...
measures cannot be relaxed. Many scholars have analyzed distribution characteristics and factors affecting COVID-19, SARS, AIDS, and other diseases by using spatial statistical methods (Q. Li et al., 2020; Su & Guo, 2020; Wang et al., 2005; Yuan et al., 2019). Sajadi et al. (2020) studied the role of temperature, humidity, and latitude in the survival and spread of seasonal respiratory viruses by comparing cities with and without epidemics. Ma et al. (2020) investigated the effects of temperature variation and humidity on deaths due to COVID-19 in Wuhan based on time series data. In addition, many studies have shown that the prevalence of hemorrhagic fever with renal syndrome (HFRS) also has typical population distribution, geographic distribution, and seasonal characteristics (Hansen et al., 2015; Zhang et al., 2017). The prevalence of HFRS in a specific area is related to temperature, daily precipitation, and daily humidity. At the same time, some studies have pointed out that there is a certain degree of spatial clustering in the incidence of malaria, and periodic changes are related to periodic climatic changes in the middle of the year (Kumar et al., 2014; Xia, 2015). Among these, the more significant factors are temperature and precipitation.

Therefore, studying the relationship between the temporal and spatial evolution of cases and environmental and meteorological elements is conducive to clarifying the external environment of virus transmission and enriching our understanding of the virus transmission process. This study uses mainly spatial statistical analysis to measure the evolution characteristics of the temporal and spatial patterns of confirmed COVID-19 cases. It also establishes a geographically weighted regression (GWR) model to explore the impact of meteorological conditions and environmental factors on the spread of COVID-19. This study helps to provide a reference for grasping the current situation and the development of the epidemic.

2. Data and Methods

2.1. Study Area and Data

Hubei province is located in central China, with dense river networks and extensive transportation infrastructure. It has a subtropical monsoon climate, with abundant annual precipitation of about 800–1,600 mm. Hubei province includes 12 cities, 3 municipalities, 1 autonomous prefecture, and 1 forest area (Figure 1). The cumulative confirmed cases from the coronavirus lockdown on January 23, 2020, until the lifting of the lockdown on April 8, 2020, are available from the Hubei Provincial Health Commission (http://wjw.hubei.gov.cn/). Various daily meteorological observations for 1981–2020, including ground temperature (GST), precipitation (PRE), atmospheric pressure (PRS), relative humidity (RHU), sunshine duration (SSD), temperature (TEM), and wind speed (WIN), can be obtained from the China Meteorological Administration.
GeoHealth 10.1029/2020GH000358

Hourly environmental indexes from January 23, 2015, to April 8, 2020, including AQI, CO, NO₂, O₃, PM₁₀, PM₂.₅, and SO₂, are obtained through the China Environmental Monitoring Center (http://www.cnemc.cn/). Data from 2015 to 2019 are used to analyze the change in air quality in Hubei province. The base period of 1981–2010 is used to obtain the climatic mean (Figure 2).

2.2. Methods

2.2.1. Exploratory Spatial Data Analysis

Exploratory spatial data analysis is often used to investigate the spatial correlation characteristics of objects or the properties of variables. Spatial autocorrelation describes the correlation between a variable at a location and its neighborhood (Anselin, 2013; Zhou et al., 2017). It often uses an index to measure the degree of spatial dependence of object attributes in the research area. It can be divided into negative correlation, positive correlation, and noncorrelation (i.e., randomness). It can also be divided into global autocorrelation and local autocorrelation from the perspective of the measuring method. Generally, changes in the overall spatial difference are not completely consistent with changes in local spatial differences. Therefore, this study applies the global spatial autocorrelation index (Moran’s I) to test whether there is a spatial global correlation (i.e., aggregation and dispersion) among the unit attribute values in the adjacent regions. It also uses the Local Moran’s I index (LISA) to analyze the local spatial association patterns and attribute variations.

The Moran’s I index can be expressed as follows:

\[ I = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (x_i - \bar{x})(x_j - \bar{x}) \]

where \( n \) is the number of spatial units, \( x_i \) and \( x_j \) are observational values of the spatial units \( i \) and \( j \), respectively, and \( W \) is the weight matrix, where \( W_{ij} = 1 \) if spatial units \( i \) and \( j \) share a border and \( W_{ij} = 0 \) otherwise.

The LISA index is used to analyze the spatial heterogeneity between unit \( i \) and unit \( j \) in the study area. The calculation is as follows:

\[ Hi = (x_i - \bar{x}) \sum_{j=1}^{n} W_{ij} (x_j - \bar{x}) \]

where the standardized statistic \( Z \) test is used to measure the significance of the Moran’s I index. At a given significance level (\( \alpha = 0.05 \)), the index has a value range of \([-1, 1]\). A positive \( I \) value indicates that the confirmed cases of the epidemic in Hubei province have a spatial aggregation trend, and a negative \( I \) value indicates a negative correlation; an \( I \) value close to the expected value \(-1/(n-1)\) indicates that the confirmed cases of the epidemic are randomly distributed.

2.2.2. Geographically Weighted Regression Model

A GWR model is used to explore the impact of environmental and meteorological factors on the outbreak and spread of the epidemic (Fotheringham et al., 1998; X. Li et al., 2020; McMillen, 2004; Sajadi et al., 2020). The GWR model is different from global regression. It generates local regression results, which can reveal the spatial heterogeneity of parameters and can be used to explore the relationship between independent and dependent variables (i.e., cumulative confirmed cases) at different spatial locations. The specific model expression is:
\[
y_i = \beta_0(u_0, v_0) + \sum_{k=1}^{n} \beta_k(u, v_i) x_k + \epsilon_i
\]

(3)

where \((u_i, v_i)\) represents the coordinates of the observation site \(i\); \(\beta_k(u_i, v_i)\) represents the regression parameter of the \(k\)-th independent variable; \(\beta_0(u_0, v_0)\) represents the regression constant; and \(\epsilon_i\) represents the residual error of observation site \(i\).

3. Results and Discussion

3.1. Stages of Epidemic Development in Hubei Province

From January 23 to April 8, 2020, a total of 67,803 confirmed cases were reported in Hubei province, of which the cumulative confirmed cases in Wuhan were significantly higher than in other cities (Figure 1). The number of confirmed cases in Hubei province was almost the same as in Wuhan, and changes were highly consistent. The cumulative confirmed cases in Wuhan accounted for 73.75% of the total confirmed cases in Hubei province (Figure 3). Wuhan faced a severe situation during the infection process.

Outside Wuhan, Hubei province had different incidence levels. The cumulative confirmed cases in Xiaogan city and Huanggang city were also significantly greater than in other cities, accounting for 5.19% and 4.29% of the total confirmed cases, respectively. The prefectures and cities with a relatively small number of cumulative confirmed cases were concentrated mainly in the western part of Hubei province (Figure 1). Among these, the Shennongjia Forest District had the fewest cumulative confirmed cases (only 11). After February 20, the cumulative confirmed cases gradually stabilized (Figure 3), and the number of daily new cases fluctuated and started to decline until February 29. As of April 4, the number of newly confirmed cases in the whole province was zero.

On February 20, the number of newly confirmed cases in Wuhan dropped significantly (Figures 3a and 3b). On the one hand, this was due to the control of the epidemic; and on the other hand, it benefited from the significant increase in nucleic acid testing capabilities and the cancellation of clinically diagnosed cases after this time. According to the time series of the cumulative confirmed cases and newly confirmed cases, the development of the epidemic can be divided into four stages: the initial stage (January 23–30), the rapid development period (January 31 to February 19), the decline period (February 20–29), and the end period (March 1 to April 8).

3.2. Spatial Distribution Pattern

A global spatial autocorrelation analysis of the newly confirmed COVID-19 cases in Hubei province was carried out according to the four periods of the epidemic development (Table 1). Although the overall number of newly confirmed cases tended to be distributed randomly in space, the distribution characteristics of the different stages were distinct. The global Moran’s I index shows a decreasing trend at first and then an increasing trend. During the decline period and the end of the epidemic, the confirmed cases presented obvious spatial clustering \((p < 0.05)\), showing a negative spatial correlation (i.e., high-low value clustering or low-high value clustering). There was no obvious spatial clustering in other periods.

| Period          | Moran’s I | Z value | p-value |
|-----------------|-----------|---------|---------|
| Initial         | 0.036     | 1.313   | 0.095   |
| Rapid development | −0.071   | −0.240  | 0.410   |
| Decline         | −0.110    | −1.727  | 0.014   |
| End             | −0.103    | −1.653  | 0.040   |
| Overall         | −0.068    | −0.302  | 0.410   |

Table 1

Spatial Autocorrelation Analysis of the COVID-19 Cases in Different Periods

Figure 3. (a) Daily cumulative confirmed COVID-19 cases in Hubei province and the three cities with the highest number (Wuhan City, Xiaogan City, and Huanggang City) from January 23 to April 8, 2020; (b) the same as (a) but for cumulative confirmed cases.
All air quality indicators in Hubei province showed obvious seasonal changes and decreasing trends, suggesting that air quality has been improving year by year (Figure 4). Air quality is better in summer and autumn. Moreover, the air quality in Hubei province in 2020 improved significantly compared with the past, which might be related to the lockdown of the province (Figure 2). For example, more than 3 million motor vehicles in Wuhan city stopped operation during the lockdown period, which contributed to the reduction of pollutants. Since the collinearity between O₃, SO₂, PM₁₀, GST, TEM, and the remaining factors is high (VIF > 7.5), these five influencing factors were removed in the next analysis. Unlike environmental variables, almost all meteorological variables showed increases compared to the same period except PRS (Figure 2). TEM had the largest change, with an average increase of 2.12°C (Figure 2). Among the various influencing factors, CO and SSD had an obvious spatial correlation. The Moran’s I index values of CO and SSD were 0.265 and 0.682, respectively. They passed the significance test (CO: $p = 0.024$, $Z = 2.062$; SSD: $p = 0.001$, $Z = 4.721$), indicating that these two elements had significant spatial aggregation characteristics.

The local spatial autocorrelation analysis of the confirmed cases in Hubei province from January 23 to April 8, 2020, is shown in Figure 5. The LISA cluster maps are given. The results show that there was only one high-low value cluster in Wuhan, and the results in other areas were insignificant. This might be because Wuhan accounted for the overwhelming majority of all confirmed cases in Hubei province (Figures 1 and 3), which was significantly higher than in other regions. The low-high value clusters in Huanggang city, Ezhou city, and Xianning city also confirmed this phenomenon.

### 3.3. Factors Affecting the Distribution of Newly Confirmed Cases in Hubei Province

Kriging interpolation was used to obtain centroid data for each city, and the GWR model was used to analyze how the anomalies of the environmental (AQI, CO, NO₂, and PM₂.₅) and meteorological (PRE, PRS, SSD, and RHU) factors impacted the COVID-19 outbreak and development (Cressie, 1993; Journel & Huijbregts, 1976). In the four periods, the correlation coefficient between CO and the number of COVID-19 cases was high ($r = 0.45$), and it was statistically significant ($p < 0.05$).

The GWR model was used to explore the factors influencing the newly confirmed cases from January 23 to April 8. Results of the overall analysis showed that the factors affecting the newly confirmed cases were AQI, CO, PM₂.₅, and PRE. Results of the analysis of the GWR model are given in Figure 6. The standardized residuals of the GWR model were all less than 2.5 times the standard deviation except Wuhan city (Figure 6a), and the local $R^2$ was basically in the range of 0.1581–0.1582 (Figure 6b), with a gradual increase from south to north. The GWR fitting results indicate that other areas in Hubei province were satisfactory except Wuhan. The environmental factors that affected the development of COVID-19 cases in different periods were the same (AQI, CO, and PM₂.₅), while the meteorological factors were different. The meteorological factors during the initial and rapid development periods were RHU and SSD, while they were PRE, PRS, and WIN during the decline and end periods, which is consistent with the results of Ma et al. (2020), who focused on the effects of meteorological factors on COVID-19 mortality.

The $R^2$ of the GWR model during the four periods was 0.137, 0.112, 0.080, and 0.419, respectively. Since the number of cases in many areas during the end of the epidemic was zero, the early stage of the epidemic was viewed as the best fitting period. The $R^2$ of the other periods except the end period showed a gradual decline, which indicates that with the development of the epidemic, the impact of environmental and meteorological factors gradually weakened. From January 23 to April 8, the standard regression coefficients of the model show that changes in some factors (CO, PM₂.₅, and PRE) had a weak negative correlation with COVID-19 diagnoses, while AQI had a weak positive correlation.

The analysis shows that environmental and meteorological factors had little impact on the development of the epidemic and were probably overwhelmed by the major impact of contact transmission, which is consistent with the findings of Xie and Zhu (2020), who reported that there was no evidence supporting a decline in COVID-19 case counts with warmer weather. The insufficient data samples (only 17 spatial units) and the impact of human activities such as the lockdown of Wuhan city, as well as comprehensive nucleic acid testing and other measures, might have contributed to the weak influences. At present, the epidemic has continued to grow in other countries and regions around the world. Meanwhile, the recent second
Figure 4. Changes in Hubei province’s environmental and meteorological factors in 2020 compared with historical contemporaneous values.
outbreaks in Hong Kong, Xinjiang province, and Dalian city in China also indicate that environmental and meteorological factors may contribute little to the development of the epidemic.

4. Summary and Conclusions

The number of confirmed COVID-19 cases in Hubei province increased day by day from January 23 to April 4, 2020. The number of new confirmed cases reached a peak on February 13 and then fluctuated and declined. The distribution of cases among cities was uneven. The development of the epidemic was divided into four stages. Except for the negative spatial correlation of confirmed cases during the decline and end periods, the entire period did not have a significant spatial correlation. That is, the number of confirmed cases generally tended to be distributed randomly in space. However, the number of cases in Wuhan was significantly higher than that in other parts of Hubei province, Xiaogan city, and Huanggang city, located...
in the northern part of Wuhan and adjacent to Wuhan, showing low-high cluster characteristics from the perspective of local spatial autocorrelation, while Wuhan showed a high-low spatial cluster.

A comparison of historical contemporaneous values of 14 environmental and meteorological factors demonstrates that many environmental elements showed decreases. The air quality in Hubei province improved over the same historical period. Almost all meteorological factors in Hubei province increased compared with the same historical period except atmospheric pressure, and temperature increased the most. Moreover, CO and sunshine duration presented positive spatial autocorrelation. Among all meteorological and environmental factors, anomalies of AQI, CO, PM$_{2.5}$, and precipitation had an impact on the increase of confirmed cases of COVID-19 for the whole study period, but this effect was not statistically significant ($R^2 = 0.158$, $p > 0.05$). At the same time, the impact changed with different stages of development, showing a gradual downward trend, especially in the later stage (end period). This analysis suggests that scientific measures such as lockdown, social distancing, and isolation played an important role in the containment of the COVID-19 epidemic.

**Conflict of Interest**

The authors declare no conflicts of interest relevant to this study.

**Data Availability Statement**

The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.4550136 (Huang et al., 2021).

**References**

Anselin, L. (2013). *Spatial econometrics: Methods and models* (Vol. 4). Springer Science & Business Media.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). *Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis*. *Environment and Planning A: Economy and Space*, 30(11), 1905–1927. https://doi.org/10.1068/a301905

Hansen, A., Cameron, S., Liu, Q., Sun, Y., Weinstein, P., Williams, C., et al. (2015). Transmission of hemorrhagic fever with renal syndrome in China and the role of climate factors: A review. *International Journal of Infectious Diseases*, 33, 212–218. https://doi.org/10.1016/j.ijid.2015.02.010

Huang, X., Zhou, H., Yang, X., Zhou, W., Huang, J., & Yuan, Y. (2021). Spatial characteristics of coronavirus disease 2019 and their possible relationship with environmental and meteorological factors in Hubei province, China [data set]. Zenodo. http://doi.org/10.5281/zenodo.4550136

Journel, A. G., & Huijbregts, C. J. (1976). *Mining geostatistics*. United Kingdom.

Kumar, V., Mangal, A., Panesar, S., Yadav, G., Talwar, R., Raut, D., & Singh, S. (2014). Forecasting malaria cases using climatic factors in Delhi, India: A time series analysis. *Malaria Research and Treatment*, 482851. https://doi.org/10.1155/2014/482851

Li, X., Zhou, L., Jia, T., Wu, H., Zou, Y., & Qin, K. (2020). Influence of urban factors on the COVID-19 epidemic: A case study of Wuhan city. *Geomatics and Information Science of Wuhan University*, 45(6), pp. 826–835. [In Chinese]. https://doi.org/10.13203/j.whu.whlg20200152

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., et al. (2020). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. *The Science of the Total Environment*, 724, 138226. https://doi.org/10.1016/j.scitotenv.2020.138226

McMillen, D. P. (2004). Geographically weighted regression: The analysis of spatially varying relationships. *American Journal of Agricultural Economics*, 86(2), 554–558. https://doi.org/10.1111/j.0002-9092.2004.600_2.x

Sajadi, M. M., Habibzadeh, P., Vintzileos, A., Shokouhi, S., Miralles-Wilhelm, F., & Amoroso, A. (2020). Temperature, humidity, and latitude analysis to estimate potential spread and seasonality of Coronavirus Disease 2019 (COVID-19). *Journal of Preventive Medicine and Hygiene*, 34(4), 51–58+65. (In Chinese). https://doi.org/10.3969/j.issn.1674-8425.2020.04.007

Su, L., & Guo, W. (2020). Spatial aggregation and spatial-temporal pattern of provincial cumulative confirmed count of Novel Coronavirus Pneumonia (COVID-19) in China. *Journal of Chongqing University of Technology (Natural Science)*, 34(6), pp. 826–835. [In Chinese]. https://doi.org/10.13406/j.cnki.jqyj.20200152

Wang, J., Meng, B., Zheng, X., Liu, J., Han, W., Wu, J., et al. (2005). Analysis on the multi-distribution and the major influencing factors on severe acute respiratory syndrome in Beijing. *Chinese Journal of Epidemiology*, 16–20. (In Chinese)

Xia, J. (2015). Research on spatio-temporal distribution and prediction for malaria epidemics in Hubei Province, China. (PhD dissertation). Huazhong University of Science and Technology. (In Chinese).

Xie, J., & Zhu, Y. (2020). Association between ambient temperature and COVID-19 infection in 122 cities from China. *The Science of the Total Environment*, 724, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201

Yuan, F., Liu, L., Liu, X., & Liu, L. (2019). Application of spatial statistical analysis in the field of AIDS research. *Journal of Preventive Medicine Information*, 38(12), 1450–1454. (In Chinese).
Zhang, Q., Yu, C., Yu, S., Jiang, L., & Wang, C. (2017). Influence of meteorological factors on hemorrhagic fever with renal syndrome in Anqiu city, Shandong Province. *Chinese Journal of Vector Biology and Control, 28*(1), 85–86. (In Chinese). https://doi.org/10.11853/j.issn.1003.8280.2017.01.025

Zhou, H., Huang, J., & Yuan, Y. (2017). Analysis of the spatial characteristics of the water usage patterns based on ESDA-GIS: An example of Hubei Province, China. *Water Resources Management, 31*(5), 1503–1516. https://doi.org/10.1007/s11269-017-1591-z