

# The Pattern of COVID-19 Transmission In China Based On The Data From 294 Prefecture-Level Cities

Zaipeng Duan (✉ [duanzaipeng@fzu.edu.cn](mailto:duanzaipeng@fzu.edu.cn))

Fuzhou University <https://orcid.org/0000-0002-4233-3335>

Renhua Ai

Fuzhou University

---

## Research Article

**Keywords:** COVID-19, China, Geographically Weighted Regression, Law of propagation

**Posted Date:** May 17th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-428750/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

This paper concerns an analysis about the impact of a total of 18 factors on the spread of COVID-19 in five aspects: population and economy, education, medical care, insurance, and migration. The research is based on the cumulative number of confirmed cases in 294 prefecture-level cities in China, and geographic weighted regression models. The main conclusions are as follows: ①Population and economic factors are positively correlated with the overall impact of the epidemic, but the case is just opposite in the economically developed southeast region, possibly due to the fact that the outbreak of the epidemic coincided with the Spring Festival, with a large number of people outflowing from the southeast, and economic activities decreasing, resulting in a small number of confirmed cases; ②Educational factors have some impact on the spread of the epidemic, but the characteristics of the impact are uncertain; ③Medical and insurance factors have a positive relation with the overall epidemic, different from the common sense and the vast majority of studies. This may be explained by the fact that at the initial stage of the outbreak, a large number of patients were admitted to the regular wards without adequate protection, resulting in an increase in the number of infectious patients in hospitals; ④ Migration factors have a negative relation with the overall suppression of the epidemic, indicating that during the outbreak period, migration should be strictly controlled.

## Introduction

Coronavirus disease (COVID-19) is caused by the SARS-CoV-2 virus (Fan et al., 2020), and the rapid spread of the disease has made it a global health issue. On December 31, 2019, the Chinese government first reported the outbreak of COVID-19 in Wuhan. Within a few weeks, the virus had spread rapidly throughout China. In order to get the spread of the epidemic under control, the Chinese authorities took drastic and effective measures and achieved remarkable results. On March 19, 2020, the number of new domestic cases was "zero" for the first time. By March 19, 2020, 81,252 cases of COVID-19 were officially confirmed in mainland China, including 3252 deaths. At the same time, the spread of the epidemic is still continuing in many other countries and regions, and the situation of epidemic prevention and control remains severe. Therefore, an in-depth analysis of the spread and control of COVID-19 in China is of great significance for international epidemic prevention and control.

At the early stage of the outbreak, most of the researches on COVID-19 were medical ones. Chan (Chan et al., 2020) observed the infected patients and confirmed the presence of polymerase and surface spike protein in the COVID-19 virus, indicating that this is a new type of coronavirus close to SARS. Guan (Guan et al., 2020) summarized the clinical characteristics of patients by observing confirmed 1,099 patients two months before the outbreak. Zhou (F. Zhou et al., 2020) used Logistic regression to prove that senior patients and patients with higher sequential organ failure assessment scores have a higher risk of death. Ai (Ai et al., n.d.) discussed the diagnostic value of chest CT using RT-PCR as a reference standard in COVID-19 and confirmed that chest CT is highly sensitive to the diagnosis of COVID-19.

COVID-19 is a new highly contagious infectious disease. At the early stage of the outbreak of the epidemic, there were very few studies on its transmission rules, except for the studies on the transmission rules of other infectious diseases. The spread of infectious diseases needs to meet the three links at the same time: source of infection, route of transmission, and susceptible population. Therefore, the effective prevention of the spread of an epidemic lies in controlling the source of infection and cutting off the route of transmission. In terms of sources of infection, large-scale population movements are an important factor leading to the rapid spread of an epidemic. In 2005, Woolhouse (Woolhouse and Gowtage-Sequeria, 2005) pointed out that repeated outbreaks of infectious diseases were closely related to the surge of tourist traffic. Population mobility was often due to economic trade, so frequent economic trade was considered to be an important driving force for the spread of infectious diseases (Weiss and McMichael, 2004). Cutting off the route of transmission is an effective means to control the spread of an epidemic. It is crucial to improve the residents' consciousness of infectious disease prevention and control. Various factors such as income and expenditure, education level and other factors will significantly affect residents' consciousness of the prevention and control. Therefore, the level of education was closely related to the spread of the epidemic (Zhong et al., 2020). Whether there is insurance after the outbreak of the epidemic will directly affect income and

expenditure. The spread of infectious diseases was closely related to insurance, especially medical insurance (Ridde et al., 2020). Areas with better public medical and health security can detect the epidemic in time and take measures to stop the spread of the epidemic. The vast majority of studies showed that the medical and health system has an important inhibitory effect on the spread of epidemics (Dalziel et al., 2018). With the further development of the epidemic prevention and control situation, researches on the transmission pattern of COVID-19 are constantly on the rise. By studying 425 confirmed cases in Wuhan, Li (Li et al., 2020) pointed out that the average incubation period of COVID-19 is 5.2 days and the epidemic situation doubles every 7.4 days at the early stage. Wang (J. Wang et al., 2020) pointed out that environmental conditions such as humidity and temperature would affect the spread of COVID-19. Research by Anderson (Anderson et al., 2020) showed that isolating infected people could curb the epidemic. Bruinen (Bruinen de Bruin et al., 2020) analyzed the measures taken by countries around the world to confront the COVID-19 epidemic and sorted them out.

Facts have showed that the spread of the epidemic takes on obvious spatial characteristics. The epidemic in China was spreading outward from Wuhan, Hubei Province (Fan et al., 2020), but there are few specific studies. Understanding the spatial characteristics of the spread of the COVID-19 epidemic is essential for predicting local outbreaks of COVID-19 and formulating public health policies at an early stage. Up to now, researches on the spatial spread of COVID-19 in China have been very limited. This paper is based on multi-scale geographic weighted regression to study the spatial characteristics of the spread of COVID-19, shedding light on preventing and controlling the spread of COVID-19.

## Data And Methods

### 2.1 Data

#### (1) Epidemic data

Data on the cumulative number of confirmed COVID-19 cases in March were collected according to prefecture-level cities and dates. Before March, China's epidemic prevention and control period was relatively short, and the cumulative number of confirmed cases was small; after March, too many factors intervened and affected China's epidemic prevention and control. And on March 19, China cleared the number of new domestic cases for the first time, and the focus of epidemic prevention and control shifted to preventing imported cases from abroad.

#### (2) Related factor data

In this paper, the authors referred to the study on spatial transmission law of SARS (Fang et al., 2009) and the factors of infectious disease transmission. In order to study the impact of the source of infection, the control variables are controlled at the population and economic level and the migration traffic level. In order to study the impact of transmission routes, in this paper the control variables of higher education level, medical level and insurance level are controlled. The description of variables is shown in Table 1.

**Table 1 Description of variables**

Category	Name	Definitions
Constant	Intercept	Intercept term of model
	Pop	Total Population (10,000 people)
Economics	GovBudget	General Public Budget Expenditure (ten thousand yuan)
	GDP	GDP (yuan)
	NumUnemploy	Number of Unemployed (person)
	NumCollege	Number of Universities (institute)
Education	NumVocational	Number of Secondary Vocational Education Schools (institute)
	PopCollege	Number of General Undergraduate and Junior College students (person)
	PopVocational	Number of students in Secondary Vocational Education Schools (person)
	NumHos	Number of Hospitals (institute)
Medical	NumBed	Number of Hospital Beds (piece)
	NumDoc	Number of Licensed (Assistant) Physicians (person)
Insure	NumMeInsure	Number of urban Workers covered by basic Medical Insurance (person)
	NumUnInsure	Number of people insured by Unemployment Insurance (person)
Traffic	NumBus	Number of public steam (electric) vehicles in operation at the end of the year
	NumPassenger	Annual total number of bus (electric) passenger trips (ten thousand passengers)
	NumTaxi	Number of taxi operating vehicles at the end of the year (units)
	Out	Emigration scale index
	In	Immigration scale index

Except for the emigration scale index and immigration scale index from Baidu Migration (Ming et al., 2020), the other variables data were derived from China Urban Statistical Yearbook and China Statistical Yearbook in 2019. In order to eliminate the impact of magnitude and unit, the cumulative number of confirmed dependent variables was logarithmized (Add 1 to prefecture-level cities with 0 confirmed cases), and 18 independent variables were standardized.

## 2.2 Data analysis

### 2.2.1 Ordinary Least Squares (OLS)

The OLS (ordinary least squares) model is a multivariate linear function used to explain the relationship between the dependent variable and the independent variable . In this paper, the OLS model is used to analyze the linear relationship between the cumulative number of confirmed cases and various influencing factors (Mollalo et al. , 2020). The calculation formula of OLS is expressed as formula (1):

$$y_i = \beta_0 + \sum_i \beta_i x_i + \varepsilon_i \quad (1)$$

### 2.2.2 Geographically Weighted Regression (GWR)

General linear regression model is usually used to summarize the laws existing within the overall range, but the actual laws vary with geography. Therefore, geographically weighted regression model should be used to eliminate estimation errors caused by spatial heterogeneity. (Kang et al., 2020). The calculation formula of geographically weighted regression is expressed as formula (2):

$$y_i = \sum_{j=1}^m x_{ij} \beta_j(u_i, v_i) + \varepsilon_i \quad (2)$$

### 2.2.3 Multi-scale Geographically Weighted Regression (MGWR)

The bandwidth used by the classic GWR is fixed, making the data unable to be optimally used. The multi-scale geographically weighted regression model overcomes this drawback (Oshan et al., 2019). The calculation formula of multi-scale geographic weighted regression is expressed as formula (3):

$$y_i = \sum_{j=1}^n \beta_{bwj}(u_i, v_i) x_{ij} + \varepsilon_i \quad (3)$$

### 2.2.4 Model and software

In this study, ArcGIS was used to visualize the cumulative number of confirmed cases and draw the LISA(Local Indicator of Spatial Association) map of the cumulative number in prefecture-level cities. Geoda was used to analyze the OLS model and calculate the global autocorrelation index Moran's I of OLS model residuals. Finally, MGWR2.2 was used to set up and analyze GWR model and MGWR model, and ArcGIS was used for visual analysis of model results.

## Results

### 3.1 Spatial analysis of the cumulative number of confirmed cases

Logarithmize and map the cumulative number of confirmed cases in each prefecture-level city, as shown in Figure 1. Use different colors to indicate the cumulative number of confirmed cases in different prefecture-level cities. The darker the color, the larger the number is. It can be seen that the cumulative number of confirmed cases in prefecture-level cities across the country has obvious clustering characteristics. The specific manifestation is that the levels of cumulative number of confirmed cases in cities across the country vary greatly and take on polarization. The closer a city is to Wuhan, the greater the cumulative number of confirmed cases, and the farther away a city is from Wuhan, the lower the cumulative number of confirmed cases. Most of the cumulative number of confirmed cases is concentrated in Central China, South China and East China.

Next, used local Moran's I (LISA) to explore the spatial autocorrelation model of the cumulative number of confirmed cases, as shown in Figure 2. From the LISA chart of the cumulative number of confirmed cases, it can be seen that it takes on an obvious positive spatial autocorrelation, and that Central China and East China are high agglomeration areas, while Northwest and Southwest China are low agglomeration areas.

### 3.2 Results of OLS model

In order to guarantee the overall validity of each factor, the OLS model was used for testing, and the results are shown in Table 2. It can be seen from Table 2 that at a significance level of 95% and below, there are altogether 6 independent variable factors having significant impacts on the cumulative number of confirmed cases. It can be seen from the absolute values of the coefficients that the descending order of importance is the In, the Out, the NumDoc, the NumTaxi, the NumHos, and the GDP. Positive are the coefficients of the GDP, the NumDoc, and the Out, while negative are the regression coefficients of the NumHos, the NumTaxi, and the In.

**Table 2 Statistical descriptions of regression coefficients**

Variable	OLS			GWR			MGWR			
	B	SE	t	Min	Median	Max	Min	Median	Max	Bandwidth
Intercept	3.17***	0.072	43.886	-0.599	0.244	0.761	-0.933	-0.058	1.251	43
Pop	-0.001	0.265	-0.005	-0.748	-0.313	0.658	-0.314	-0.309	-0.29	293
GovBudget	-0.201	0.304	-0.663	-1.304	-0.157	0.592	-0.144	-0.142	-0.125	293
GDP	0.229*	0.103	2.22	-0.145	0.046	0.893	-0.103	0.056	0.295	161
NumUnemploy	0.058	0.143	0.406	-0.256	0.094	0.26	0.079	0.084	0.094	293
NumCollege	0.717	0.44	1.629	-0.011	0.338	0.704	-0.023	-0.016	0.006	293
NumVocational	0.116	0.181	0.641	-0.302	0.064	0.551	-0.152	-0.137	-0.119	293
PopCollege	-0.006	0.408	-0.014	-0.424	0.118	0.818	0.319	0.326	0.346	293
PopVocational	-0.344	0.19	-1.808	-0.505	-0.066	0.162	0.091	0.111	0.118	292
NumHos	-0.537**	0.163	-3.287	-0.385	-0.184	0.029	-0.122	-0.114	-0.096	293
NumBed	0.317	0.403	0.788	-0.66	0.223	1.101	0.651	0.653	0.672	293
NumDoc	0.921*	0.413	2.23	-0.259	0.523	1.786	0.248	0.255	0.266	293
NumMeInsure	-0.082	0.115	-0.713	-0.759	0.276	2.461	-0.086	-0.081	-0.072	293
NumUnInsure	0.497	0.442	1.125	-2.097	0.288	1.206	0.429	0.434	0.443	293
NumBus	0.175	0.233	0.752	-0.787	-0.07	0.515	0.211	0.216	0.228	293
NumPassenger	-0.074	0.304	-0.244	-1.217	0.168	1.021	-0.692	-0.233	0.212	63
NumTaxi	-0.666*	0.31	-2.151	-1.156	-0.339	1.072	-0.201	-0.193	-0.175	293
Out	2.694***	0.456	5.904	0.374	2.025	3.817	1.503	1.531	1.576	291
In	2.958***	0.521	-5.673	-4.765	-2.6	-0.545	-3.4	-1.822	-1.088	43

Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

The diagnostic indicators of the OLS model were calculated, as shown in Table 3, where  $R^2$  is the determination coefficient, and AICc (corrected Akaike information criterion) is the modified AIC. According to the  $R^2$ , the model explains approximately 44.8% of the total variations of the dependent variables in 294 units. Calculation showed that the Moran's I index of the sample residual of the OLS model is 0.3740 ( $p \leq 0.001$ ), indicating that the residual has significant spatial autocorrelation and that the residual distribution shows obvious spatial agglomeration characteristics. Since the parameter estimation of the OLS model demonstrated larger deviation, GWR model was used for analysis.

**Table 3** Diagnostic indicators

	$R^2$	Adjust $R^2$	AIC	AICc
Diagnostic indicators of OLS	0.448	0.412	697.646	702.723
Diagnostic indicators of GWR	0.741	0.673	560.445	593.542

### 3.3 Results of GWR model

Since the OLS fitting results had drawbacks, the GWR model was set and used for analysis (Oshan et al., 2019). The diagnostic indicators of the GWR model were obtained as shown in Table 3.

In this paper, the authors referred to the views of Fotheringham (Fotheringham et al., 1998). If the difference of AICc between GWR fitting results and OLS fitting results is greater than 3, it indicates that the GWR is superior to the OLS. As is shown in Table 3, the AICc values of the two models differ by 109.181, so the GWR model with a lower AICc value has better fitting result. Compared with the OLS model, the  $R^2$  and adjusted  $R^2$  of the GWR model are significantly improved, indicating that the residual is effectively reduced and the GWR fitting result is more ideal.

The results of the GWR model are shown in Table 2. The estimated coefficients of the 18 factors vary in different grid cells, and the regions with significant influence also have significant spatial heterogeneity. The regression coefficient ranges of impacting factors vary greatly, and the coefficients are either positive or negative. The regression coefficients obtained by OLS can only represent the overall average level. A stable coefficient relationship does not exist between the cumulative number of confirmed cases and the influencing factors, since the influencing factors show strong spatial instability. GIS map was used to show the spatial characteristics of the influence changes of each significant factor, as shown in Figure 3.

### 3.4 Results of MGWR model

The fitting result obtained from GWR model was superior to that from the OLS model. However, since the GWR uses fixed bandwidth and this time the bandwidth was calculated to be 173, the data could not be optimally utilized. Then the MGWR model had to be used for analysis (Oshan et al., 2019 ). It can be seen from Table 4 that the goodness of fit  $R^2$  of MGWR is higher than that of classic GWR, and the value of AICc is lower than that of classic GWR. Therefore, it could be determined that the result of MGWR is better than that of classic GWR. In terms of the number of valid parameters, the MGWR is smaller and the residual sum of squares is also much smaller, indicating that it can use fewer parameters to obtain a regression result closer to the true value. Therefore, the MGWR model in this case is better than the classic GWR model.

**Table 4** Indexes of classic GWR and MGWR models

Indicators of model	MGWR	Classic GWR
$R^2$	0.770	0.673
AICc	483.579	593.542
Number of valid parameters v1	56.585	60.409
Residual sum of squares	54.669	76.266

The statistical description of each coefficient of MGWR is shown in Table 2. It can be seen from Table 2 that MGWR can directly reflect the differential action scale of different variables, while the classic GWR can only reflect the average value of the action scale of each variable. The bandwidth of the classic GWR is 173, accounting for 59% of the total sample number. By calculating MGWR, it was found that the scales of action of different variables varied greatly. Among the MGWR results, significant were the regression coefficients of the 8 variables: Intercept, Pop, GDP, NumBed, NumBus, NumPassenger, Out and In, while others were insignificant.. The Intercept expressed the influence of different positions on the cumulative number of confirmed cases when other independent variables were determined. GIS map can be used to show the spatial characteristics of the influence variations of each significant factor, as shown in Figure 4.

### 3.5 Summary of regression coefficients of significant factors of each model

The regression coefficients of the significant influencing factors of the OLS model, GWR model and MGWR model were summarized, and the results are shown in Table 5.

**Table 5 Summary of regression coefficients of significant factors**

Categories	Variable	OLS	GWR	MGWR	Whether the basic consistent
Constant	Intercept	+	strong+↔weak-	equal strength	×
Economics	Pop		strong-↔weak+	strong-	✓
	GovBudget		strong-		
	GDP	+	strong+	strong+	✓
	NumUnemploy		strong+		
Education	NumCollege				
	NumVocational		strong+↔weak-		
	PopCollege		strong+		
	PopVocational		strong-		
Medical	NumHos	-	strong-		✓
	NumBed		strong+	strong+	✓
	NumDoc	+	strong+		✓
Insure	NumMeInsure		strong+		
	NumUnInsure		strong+		
Traffic	NumBus			strong+	
	NumPassenger		strong-↔weak+	strong-	✓
	NumTaxi	-	strong-↔weak+		
	Out	+	strong+	strong+	✓
	In	-	strong-	strong-	✓

**Notes:**↔In the third column: "+" means that the regression factor is positively significant in 95% of the confidence interval in the OLS model, "-" means negatively significant, and blank means not significant.

↔In columns 4 and 5: "Strong +, Weak -" means that the regression factors exist respectively in the GWR model and the MGWR model. In 95% of the confidence interval, both positive and negative significance exist, but mainly positive significance; "Strong- , Weak+" means that the majority is negatively significant; "equal strength" means that the difference between positive and negative is not big; "strong +" means that most of them are positively significant; "Strong-" means that most of them are negatively significant; and blank means not significant.

↔The "✓" in the sixth column indicates that the regression coefficients of two or three models are close in the "Whether basically consistent"; "×" indicates that the regression coefficients of two or three models are not close, and blank means that the factors are not significant in 95% of the confidence intervals of the three models, or are significant in only one model.

## Discussion

### 4.1 The impact of population and economic factors on the cumulative number of confirmed cases



Population and economic factors are positively related to the overall impact of the epidemic. The effect of the Pop in the GWR model was mainly reflected in the southeast side of China's "Hu Huanyong" line (D. Wang et al., 2020): in the northeast, the more the population, the greater the number of confirmed cases would be, while gradually to the north of China, the more the population, the smaller the number would be. The situation continued down to the south China region. In the MGWR model, they were limited to the South China region of China, and the effect of the "Hu Huanyong" line disappeared. The GDP in Southwest, Central and South China showed a significant role in promoting the spread of the epidemic. Take Wuhan, a city in central China, as an example. Wuhan, an economically developed city, was the area where the epidemic first broke out in China. Because COVID-19 is extremely contagious (Lai et al., 2020), the impact of economic activities on the cumulative number of confirmed cases had been magnified. South China is a region with better economy in China, and the overall epidemic prevention and control was better, with a small regression coefficient among significant regions. This might be explained by the fact that the outbreak coincided with the Spring Festival, and the region had a large population outflow during the Spring Festival and economic activities had decreased. Therefore, the cumulative number of confirmed cases was small. Southwest China had a relatively high regression coefficient due to its low GDP, which might be due to its relatively poor economic development and backward concept leading to the difficulty in epidemic prevention and control. NumUnemploy showed a significant role in promoting the spread of the epidemic in East China, possibly because of the increase in the number of unemployed people and the increased mobility of workers, which led to the spread of the epidemic.

## **4.2 The impact of educational factors on the cumulative number of confirmed cases**

The impact of educational factors on the characteristics of the epidemic is uncertain. The NumVocational had an impact on the spread of the epidemic in Southwest, Central and South China, but the characteristics of the impact were relatively uncertain. The negative correlation area was mainly concentrated in the north of the Qinling-Huaihe line (Liu et al., 2020), and the positive correlation area was mainly concentrated in the south of the Qinling-Huaihe line. The PopCollege in the northwestern region showed a significant role in promoting the spread of the epidemic, but the significant areas were too limited to show regularity. The PopVocational in the Northwest, Southwest and South China showed a significant inhibitory effect on the spread of the epidemic, and the correlation effect gradually decreased from south to north. Almost all students in secondary vocational education schools went to local schools, with poor mobility, which had an inhibitory effect on the spread of the epidemic, and they were highly educated people. Most studies showed that the higher the education level is, the easier it is to control the spread of infectious diseases (Zhong et al., 2020).

## **4.3 The impact of medical factors on the cumulative number of confirmed cases**

Medical factors are positively related to the overall epidemic. The NumHos in Central and East China showed a significant inhibitory effect on the spread of the epidemic, and its correlation effect gradually spread outwards from North China. Studies confirmed that the more the medical institutions, the better the prevention of the spread of the epidemic (Dalziel et al., 2018), so new infectious disease hospitals should be built in time. The Chinese government built the "Thunder God Mountain" and "Fire God Mountain" hospitals in Wuhan at an extremely high speed, which greatly restrained the spread of the epidemic (B. Zhou et al., 2020). The NumBed in Central and East China showed a significant role in promoting the spread of the epidemic. This result is contrary to common knowledge because the patients need special negative pressure beds (Thampi et al., 2020). Take Wuhan, a city in central China, as an example. At the early stage of the outbreak, a large number of patients were arranged to be treated in regular beds, resulting in an increase in the number of infected patients in the hospital. It is worth noting that Xinjiang has the largest regression coefficient among the areas with significant factors. Once the epidemic continues to develop, it will be more difficult to control. The NumDoc in Southwest, South China and Central China showed a significant role in promoting the spread of the epidemic. The positive correlation effect decreased from Southwest China to Central China, and disappeared in North and Northeast China. This result is contrary to common knowledge, caused by the extreme shortage of medical resources at the initial stage of the outbreak. The data used in this study was collected during

March 2020, a critical period for the prevention and control of the epidemic in China. The number of doctors was severely insufficient. There were many cases where a large number of domestic doctors rushed to help Hubei Province. In the case of a severe shortage of doctors, a large number of doctors can detect more cases, and because the medical conditions in cities in western China were relatively backward, the role of doctors became more obvious.

## **4.4 The impact of insurance factors on the cumulative number of confirmed cases**

Insurance factors are positively related to the overall spread of the epidemic. The NumMeInsure showed a significant role in promoting the spread of the epidemic in the central and southeastern regions, and the correlation effect was gradually decreasing from west to east. This result is contrary to common knowledge, because in order to get the epidemic under control quickly, China adopted the policy of free medical care on the basis of medical insurance reimbursement, and the remaining part was subsidized by the national finance. Such practice encouraged more people to take the initiative to receive medical examination, leading to an increase of confirmed cases within a certain period of time, but it controlled the subsequent spread of the epidemic. The NumUnInsure showed a significant role in promoting the spread of the epidemic in East China, and the correlation effect gradually decreased from south to north. By comparing the distribution of the number of people participating in unemployment insurance, it could be found that in the regions with a large number of people covered by unemployment insurance, the cumulative number of confirmed cases rose slowly. This might be explained by the fact that unemployment insurance reduced unnecessary economic activities and mobility of the unemployed., but it was not enough to fully cover living expenses, so the cumulative number of confirmed cases would still increase.

## **4.5 The impact of migration and traffic factors on the cumulative number of confirmed cases**

Migration and traffic factors have a negative impact on the overall suppression of the epidemic. The NumBus in Xinjiang showed a significant role in promoting the spread of the epidemic, but only three prefecture-level cities were marked as regions with no regularity. The NumPassenger had a significant impact on the spread of the epidemic in North, Central and South China, but the impact characteristics were relatively uncertain. In North and Central China, where the epidemic first broke out, the more the NumPassenger, the greater the cumulative number of confirmed cases became. In southern China, the number of confirmed cases was reduced due to the lockdown of the cities and the banning of bus operations by the Chinese government after the outbreak. The NumTaxi showed a significant restraint on the epidemic in North, Central and East China. Except for Xinjiang, the more the NumTaxi, the fewer the cumulative number of confirmed cases were. The reason might be explained by the fact that taxis carry fewer people than public transport, lowering the cumulative number of confirmed cases. The Out showed a significant effect on the spread of the epidemic. The higher the emigration scale index, the higher the urban mobility, and the spread of the epidemic intensified with population movement. The positive correlation was increasing from coastal cities to central and western cities, because most of the labor force in China migrated from central and western cities to coastal cities. The In showed a significant inhibitory effect on the spread of the epidemic. Cities with high immigration coefficients were generally metropolises or tourist cities, and they took relatively more strict prevention measures at the initial stage of the epidemic. From the GWR model, it can be seen that the effect of negative correlation increased from coastal cities to central and western cities. This might be explained by the fact that compared with central and western cities, coastal cities are larger and more populated, so it was more difficult to prevent and control the spread of the disease, but their impact coefficient was still negative. From the MGWR model, it can be found that the area with weak negative correlation effect extended north from Beijing to the northeast, Inner Mongolia and Xinjiang, seemingly foretelling the further outbreak of the epidemic in Heilongjiang, Inner Mongolia, Xinjiang and other regions in May. It was worth noting that there were many significant areas in the emigration scale index and the immigration scale index, indicating that emigration and immigration are crucial to the spread of the epidemic. In order to better prevent and control the epidemic, tight control should be imposed on the emigration and immigration.

## Declarations

### Declaration of Ethics approval and consent to participate

This manuscript does not report on or involve the use of any animal or human data or tissue.

### Consent to Participate

This manuscript does not contain data from any individual person.

### Consent to Publish

All authors agree to the publication of this paper.

### Availability of data and materials

The cumulative number of confirmed cases of COVID-19 was obtained from online crawlers using Python, and can also be collected and downloaded from major websites. Except for the emigration scale index and immigration scale index from Baidu Migration (Ming et al., 2020), the other variables data were derived from China Urban Statistical Yearbook and China Statistical Yearbook in 2019. In order to eliminate the impact of magnitude and unit, the cumulative number of confirmed dependent variables was logarithmized (Add 1 to prefecture-level cities with 0 confirmed cases), and 18 independent variables were standardized.

### Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Funding

This work was supported by The National Social Science Fund of China (Grant No. 17CGL049).

### -Authors Contributions

ARH analyzed and interpreted the data regarding the confirmed cases of COVID-19 in China. DZ was responsible for providing funding for the thesis, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

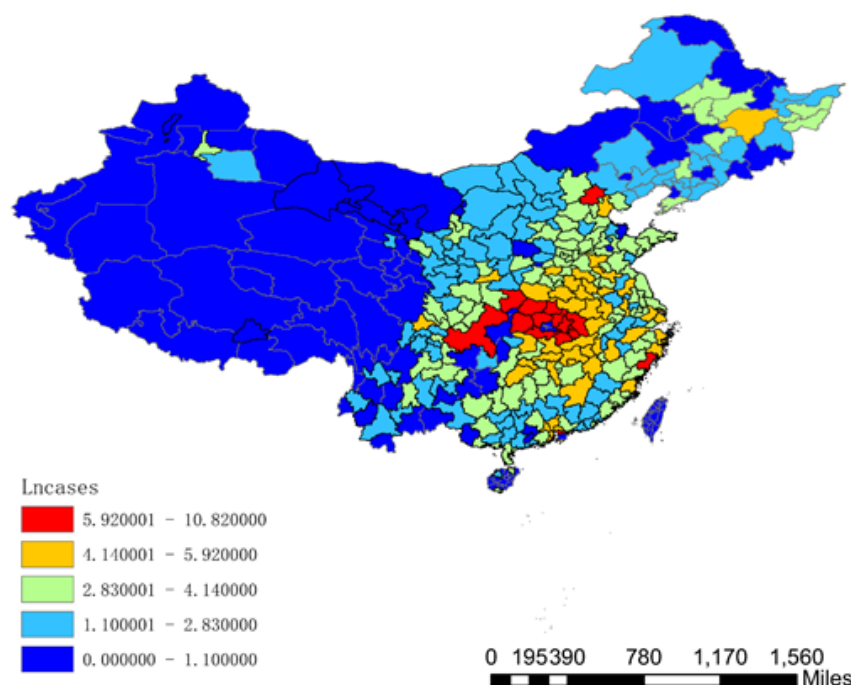
## References

1. Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L., n.d. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases 23.
2. Anderson, R.M., Heesterbeek, H., Klinkenberg, D., Hollingsworth, T.D., 2020. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet* 395, 931–934. [https://doi.org/10.1016/S0140-6736\(20\)30567-5](https://doi.org/10.1016/S0140-6736(20)30567-5)

3. Bruinen de Bruin, Y., Lequarre, A.-S., McCourt, J., Clevestig, P., Pigazzani, F., Zare Jeddi, M., Colosio, C., Goulart, M., 2020. Initial impacts of global risk mitigation measures taken during the combatting of the COVID-19 pandemic. *Safety Science* 128, 104773. <https://doi.org/10.1016/j.ssci.2020.104773>
4. Chan, J.F.-W., Yuan, S., Kok, K.-H., To, K.K.-W., Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C.-Y., Poon, R.W.-S., Tsoi, H.-W., Lo, S.K.-F., Chan, K.-H., Poon, V.K.-M., Chan, W.-M., Ip, J.D., Cai, J.-P., Cheng, V.C.-C., Chen, H., Hui, C.K.-M., Yuen, K.-Y., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* 395, 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9)
5. Dalziel, B.D., Kissler, S., Gog, J.R., Viboud, C., Bjørnstad, O.N., Metcalf, C.J.E., Grenfell, B.T., 2018. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science* 362, 75–79. <https://doi.org/10.1126/science.aat6030>
6. Fan, J., Liu, X., Pan, W., Douglas, M.W., Bao, S., 2020. Epidemiology of Coronavirus Disease in Gansu Province, China, 2020. *Emerg. Infect. Dis.* 26, 1257–1265. <https://doi.org/10.3201/eid2606.200251>
7. Fang, L.-Q., de Vlas, S.J., Feng, D., Liang, S., Xu, Y.-F., Zhou, J.-P., Richardus, J.H., Cao, W.-C., 2009. Geographical spread of SARS in mainland China. *Tropical Medicine & International Health* 14, 14–20. <https://doi.org/10.1111/j.1365-3156.2008.02189.x>
8. Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 1998. Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environ Plan A* 30, 1905–1927. <https://doi.org/10.1068/a301905>
9. Guan, W., Ni, Z., Hu, Yu, Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D.S.C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, Jin-lin, Liang, Z., Peng, Y., Wei, L., Liu, Y., Hu, Ya-hua, Peng, P., Wang, Jian-ming, Liu, J., Chen, Z., Li, G., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S., Zhong, N., 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 382, 1708–1720. <https://doi.org/10.1056/NEJMoa2002032>
10. Kang, D., Choi, H., Kim, J.-H., Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. *International Journal of Infectious Diseases* 94, 96–102. <https://doi.org/10.1016/j.ijid.2020.03.076>
11. Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., Hsueh, P.-R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *International Journal of Antimicrobial Agents* 55, 105924. <https://doi.org/10.1016/j.ijantimicag.2020.105924>
12. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med* 382, 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>
13. Liu, Jie, Yang, Q., Liu, Jian, Zhang, Y., Jiang, X., Yang, Y., 2020. Study on the Spatial Differentiation of the Populations on Both Sides of the “Qinling-Huaihe Line” in China. *Sustainability* 12, 4545. <https://doi.org/10.3390/su12114545>
14. Ming, W., Zhou, Z., Ai, H., Bi, H., Zhong, Y., 2020. COVID-19 and Air Quality: Evidence from China. *Emerging Markets Finance and Trade* 56, 2422–2442. <https://doi.org/10.1080/1540496X.2020.1790353>
15. Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment* 728, 138884. <https://doi.org/10.1016/j.scitotenv.2020.138884>
16. Oshan, T., Li, Z., Kang, W., Wolf, L., Fotheringham, A., 2019. mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *IJGI* 8, 269. <https://doi.org/10.3390/ijgi8060269>
17. Ridde, V., Aho, J., Ndao, E.M., Benoit, M., Hanley, J., Lagrange, S., Fillol, A., Raynault, M.-F., Cloos, P., 2020. Unmet healthcare needs among migrants without medical insurance in Montreal, Canada. *Global Public Health* 1–14. <https://doi.org/10.1080/17441692.2020.1771396>
18. Thampi, S., Lee, C.C.M., Ng, B.H.Z., Yap, A.P.A., 2020. Considerations for resuscitation and transfer of paediatric patients with COVID-19. *British Journal of Anaesthesia* 125, e183–e185. <https://doi.org/10.1016/j.bja.2020.04.063>

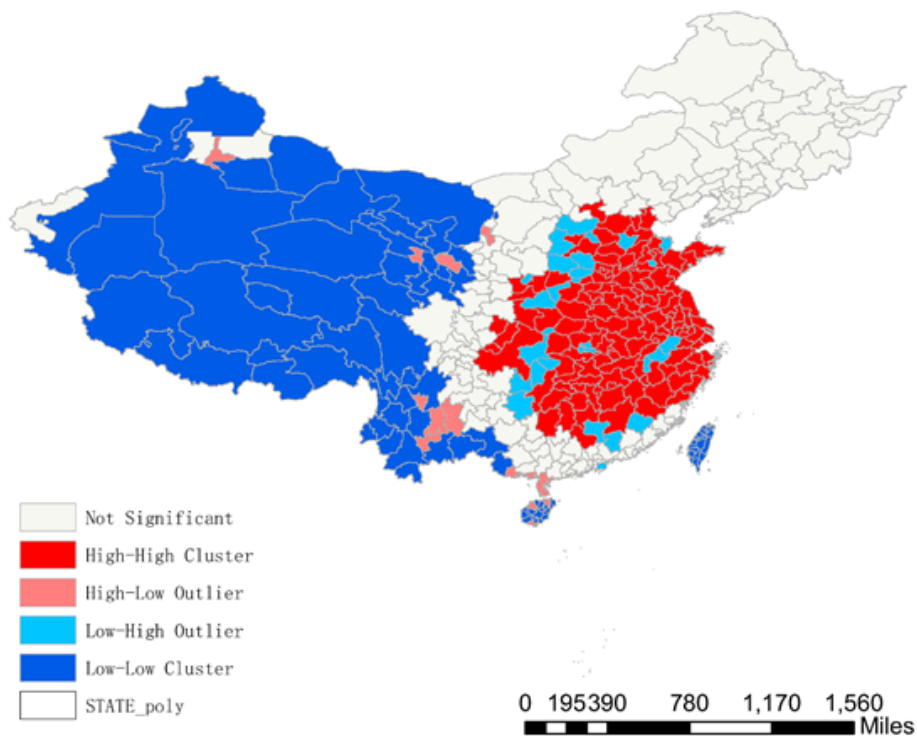
19. Wang, D., Jiang, D., Fu, J., Lin, G., Zhang, J., 2020. Comprehensive Assessment of Production–Living–Ecological Space Based on the Coupling Coordination Degree Model. Sustainability 12, 2009. <https://doi.org/10.3390/su12052009>
20. Wang, J., Tang, K., Feng, K., Lv, W., 2020. High Temperature and High Humidity Reduce the Transmission of COVID-19. SSRN Journal. <https://doi.org/10.2139/ssrn.3551767>
21. Weiss, R.A., McMichael, A.J., 2004. Social and environmental risk factors in the emergence of infectious diseases. Nat Med 10, S70–S76. <https://doi.org/10.1038/nm1150>
22. Woolhouse, M.E.J., Gowtage-Sequeria, S., 2005. Host Range and Emerging and Reemerging Pathogens. Emerg. Infect. Dis. 11, 1842–1847. <https://doi.org/10.3201/eid1112.050997>
23. Zhong, B.-L., Luo, W., Li, H.-M., Zhang, Q.-Q., Liu, X.-G., Li, W.-T., Li, Y., 2020. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int. J. Biol. Sci. 16, 1745–1752. <https://doi.org/10.7150/ijbs.45221>
24. Zhou, B., Wu, Q., Zhao, X., Zhang, W., Wu, W., Guo, Z., 2020. Construction of 5G all-wireless network and information system for cabin hospitals. Journal of the American Medical Informatics Association 27, 934–938. <https://doi.org/10.1093/jamia/ocaa045>
25. Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., Cao, B., 2020. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet 395, 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3)

## Figures



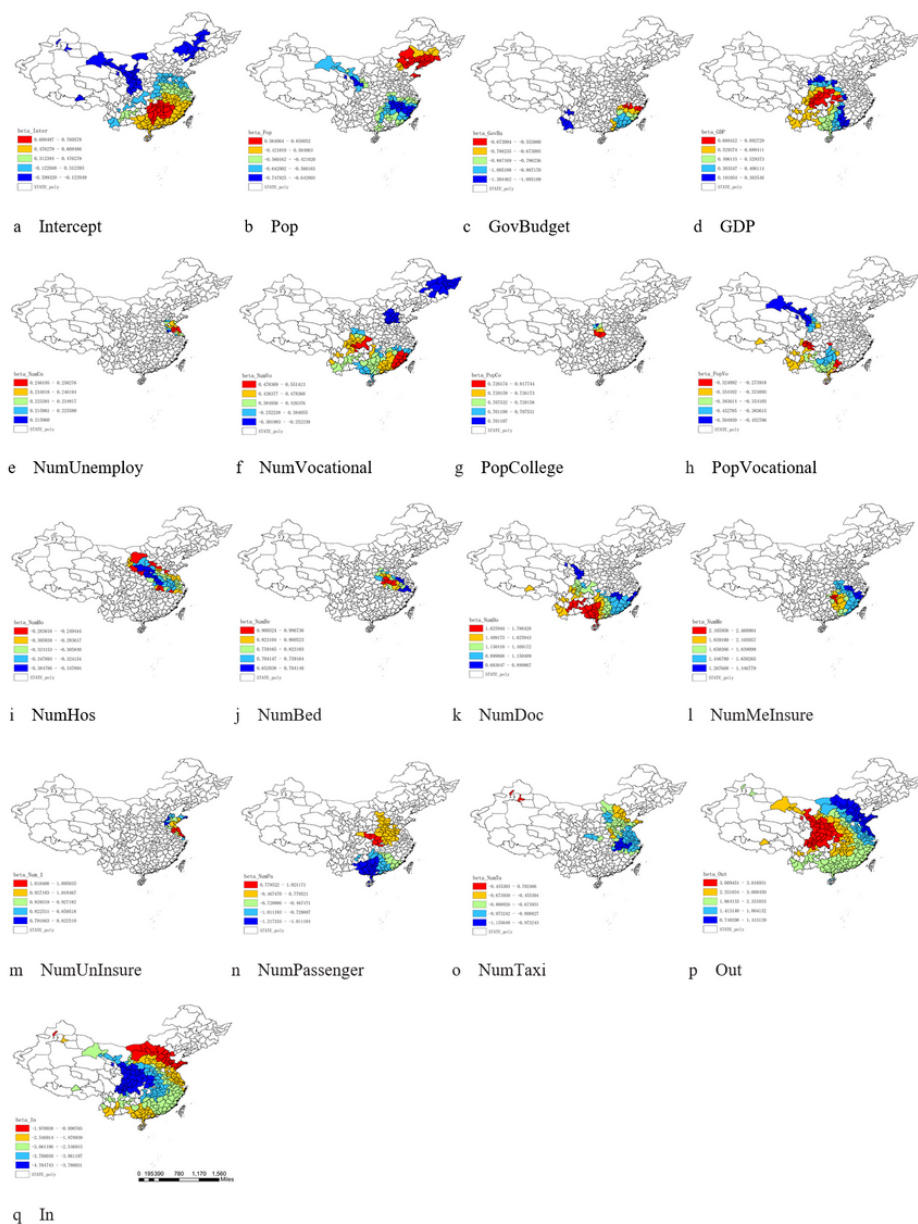
**Figure 1**

The cumulative number of confirmed cases in prefecture-level cities. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



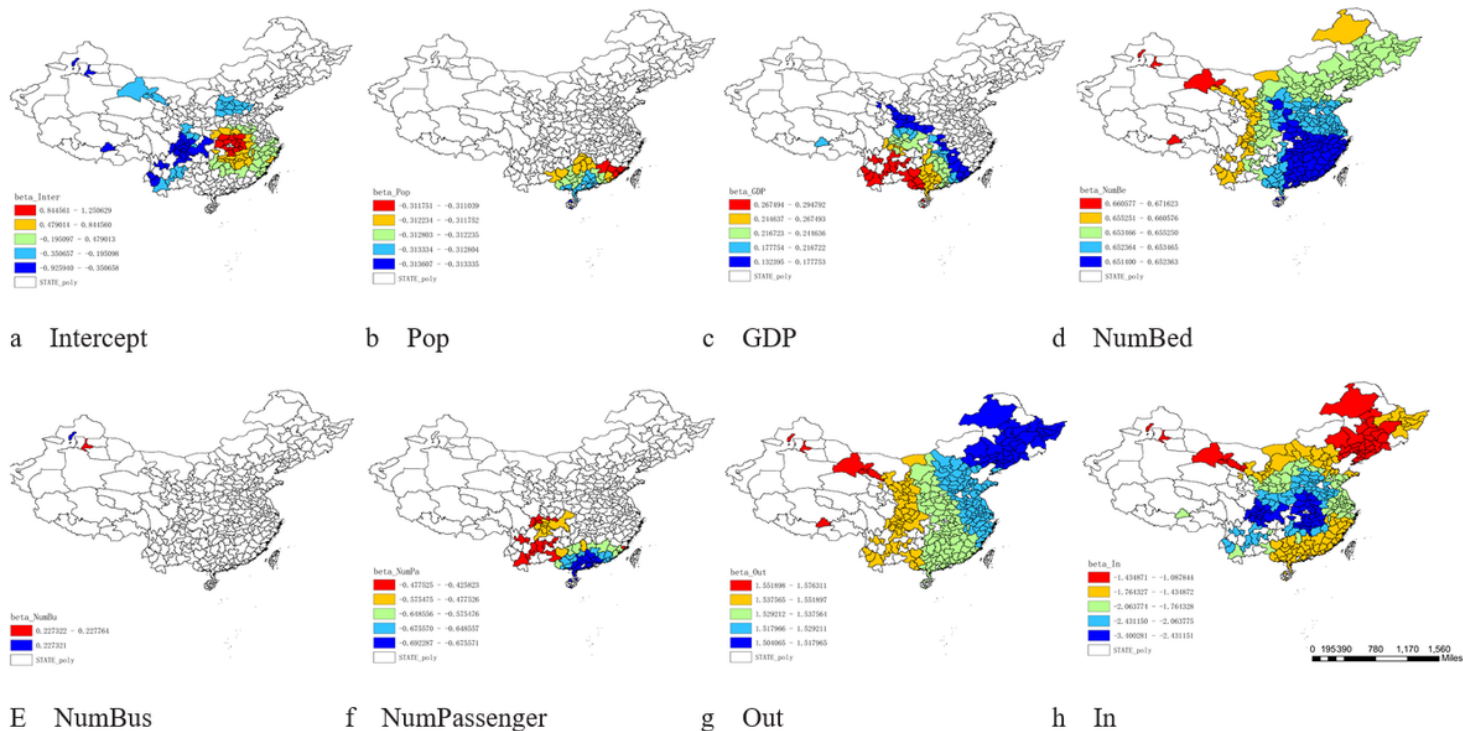
**Figure 2**

LISA map of the cumulative number of confirmed cases in prefecture-level cities. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 3**

The spatial pattern of GWR coefficients Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



**Figure 4**

The spatial pattern of MGWR coefficients Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.