GeoHealth

RESEARCH ARTICLE
10.1029/2021GH000402

**Special Section:**
The COVID-19 pandemic: linking health, society and environment

**Key Points:**

- To quantitatively evaluate how the effect of population flows on COVID-19 spread varies over time and space in Hubei Province, China
- The association showed three stages of temporal variation characteristics and strong geographical disparities across Hubei Province
- The implementation of control measures by restricting population movements can play a key role in mitigating the spread of this epidemic

**Correspondence to:**
C. Wu,
chaowu@njupt.edu.cn

**Citation:**
Chen, Y., Chen, M., Huang, B., Wu, C., & Shi, W. (2021). Modeling the spatiotemporal association between COVID-19 transmission and population mobility using geographically and temporally weighted regression. *GeoHealth*, *5*, e2021GH000402. https://doi.org/10.1029/2021GH000402

Received 7 FEB 2021
Accepted 2 APR 2021



# Modeling the Spatiotemporal Association Between COVID-19 Transmission and Population Mobility Using Geographically and Temporally Weighted Regression

**Yixiang Chen[1,2], Min Chen[1], Bo Huang[3], Chao Wu[1,2], and Wenjia Shi[1]**

[1]School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing, China, [2]Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province, Nanjing, China, [3]Department of Geography and Resource Management, The Chinese University of Hong Kong, HongKong, China

**Abstract** The ongoing Coronavirus Disease 2019 (COVID-19) has posed a serious threat to human public health and global economy. Population mobility is an important factor that drives the spread of COVID-19. This study aimed to quantitatively evaluate the impact of population flow on the spread of COVID-19 from a spatiotemporal perspective. To this end, a case study was carried out in Hubei Province, which was once the most affected area of COVID-19 outbreak in Mainland China. The geographically and temporally weighted regression (GTWR) model was applied to model the spatiotemporal association between COVID-19 epidemic and population mobility. Two patterns of population flows, including the population inflow from Wuhan and intra-city population movement, were considered to construct explanatory variables. Results indicate that the GTWR model can reveal the spatial–temporal-varying relationships between COVID-19 and population mobility. Moreover, the association between COVID-19 case counts and population movements presented three stages of temporal variation characteristics due to the virus incubation period and implementation of strict lockdown measures. In the spatial dimension, evident geographical disparities were observed across Hubei Province. These findings can provide policymakers useful knowledge about the impact of population movement on the spatio-temporal transmission of COVID-19. Thus, targeted interventions, if necessary in certain time periods, can be implemented to restrict population flow in cities with high transmission risk.

## 1. Introduction

The ongoing Coronavirus Disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has posed a serious threat to human public health and global economy (Yang, Sha, et al., 2020). Given the unavailability of specific drugs for the treatment of this disease, timely public health interventions (e.g., travel restrictions, social distancing and wearing of facial masks) are one of the most effective ways to prevent and control the epidemic. On January 23, 2020, Wuhan, the capital city of Hubei Province, China, implemented an unprecedented lockdown measure to prohibit people and motor vehicles from entering and leaving the city. Then, this policy was extended to the entire province of Hubei and other cities in China. Studies indicate that these control measures have greatly reduced the spread of COVID-19 in China. The Wuhan shutdown resulted in the delayed arrival of COVID-19 in other cities by 2.91 days (Tian et al., 2020). If the implementation of public health interventions was delayed for 5 days, the extent of the epidemic would have tripled in Mainland China (Yang, Zeng, et al., 2020).

Population mobility is one of the major drivers of epidemic transmission, accelerating the dispersion of virus in space (Kraemer et al., 2019). Through the migration of population across regions, the virus is exported from one city to another. The association between population mobility and epidemic dynamics has attracted wide attention of researchers. Different data sources that record people's mobile behaviors, such as mobile phone, social media and public transport, can be used for this study (Yang, Sha, et al., 2020). Kraemer et al. (2020) quantitatively evaluated the impact of population movement on COVID-19 epidemic. Through linear regression modeling, they observed that the population movement data can well explain the distribution of COVID-19 cases in China. Jia et al. (2020) used mobile-phone data to track population flow from Wuhan and then correlated them with COVID-19 cases of 296 cities in Mainland China by using multiplicative exponential models. The results show that population flow data can be used not only to predict

the distribution of confirmed cases but to also identify areas with high-risk transmission. Liu et al. (2020) divided population mobility into different categories, including intra-city and inter-city patterns. Furthermore, they employed linear regression models to model the relationship between different mobility patterns and the spread of COVID-19 and evaluated the impact of travel restrictions on epidemic control.

However, the basic assumption of the above-mentioned studies is that the impact of population mobility on virus transmission is stationary, that is, the association between COVID-19 and population flows is homogeneous across the entire study area and time period. In fact, the spread of virus in space is a geographical phenomenon, which is inevitably constrained by geographical laws (Sun et al., 2020). The association between infectious diseases and their influence factors possibly varies among different spatial units and/or time frames (Ge et al., 2016; Song et al., 2019). Population mobility is usually limited by spatial distance, city level, traffic and other factors (Zhang et al., 2020). The neighboring cities in space usually experience more frequent population flow than distant ones, indicating the spatial dependence of population flow (Tao & Thill, 2020); on the other hand, differences in population flow exist among various cities, for example, a large city usually has greater population migration than a small one (Pan & Lai, 2019). These spatial characteristics of population mobility indicate the local spatial dependence and global spatial heterogeneity of the spread of infectious diseases (Merler & Ajelli, 2012; Mollalo et al., 2020). In addition to these spatial effects, temporal effects often occur in time dimension. The number of cases in a region is a time series and may show correlation and difference over time (Ge et al., 2016). As a result, traditional statistical models (e.g., linear regression) may be less effective in modeling the association between infectious diseases and their influence factors, because heterogeneous spatial and/or temporal effects may violate the basic assumption of statistical independence of observations (Huang et al., 2010).

To better address this problem, researchers have developed various localized modeling techniques to capture nonstationary associations (either spatially, temporally or both) between dependent and explanatory variables (Song et al., 2019). Geographically weighted regression (GWR) (Brunsdon et al., 1996; Fotheringham et al., 2002) is one of the most representative models. Different from ordinary least squares (OLS) regression, GWR is a spatially varying coefficient model that was originally designed to model spatial variation in housing markets. In the field of epidemiology, GWR has been widely applied to model the spatial heterogeneity of epidemics and identify their influencing factors (Hong et al., 2018; Ren et al., 2017). One limitation of this model is that it cannot deal with temporal variation well because of the neglect of time information. As an extension of GWR to overcome its limitation, geographically and temporally weighted regression (GTWR) was proposed to deal with spatial and temporal variations simultaneously (Huang et al., 2010). Many studies have proven the superiority of GTWR for spatio-temporal heterogeneity modeling in the fields of housing price prediction (Wu et al., 2019), PM2.5 estimation (He & Huang, 2018) and transit ridership modeling (Ma et al., 2018). Nevertheless, limited research have used these localized modeling techniques to model the impact of population mobility on infectious diseases from a spatiotemporal perspective.

This study aimed to quantitatively evaluate how the effect of population flows on COVID-19 spread varies over time and space. Understanding the spatiotemporal heterogeneity of the association between the number of confirmed cases and population flows can provide useful knowledge for targeted interventions for the prevention and control of COVID-19 epidemic. To this end, a case study was carried out in Hubei Province, China. The GTWR model was applied to model the spatiotemporal association between the COVID-19 case counts and population migration indices. Based on the estimated regression parameters, temporal and spatial variation patterns were captured and visualized, and possible reasons were explained and discussed.

## 2. Materials and Methods

### 2.1. Study Area and Data

Hubei Province is located in the central part of China (108° 21′ 42″–116° 07′ 50″ E and 29° 01′ 53″–33° 06′ 47″ N), with a total area of 185,900 $km^2$. By the end of 2019, its permanent resident population had reached 59.27 million. Hubei has 12 prefecture-level cities, one autonomous prefecture (Enshi Tujia and Miao Autonomous Prefecture), three county-level cities (Xiantao City, Qianjiang City, Tianmen City) and one forest area (Shennongjia forest area).

Hubei is the most affected area of COVID-19 outbreak in China. COVID-19 disease was first discovered in Wuhan, the capital of Hubei Province. Then, with the massive population movement around Spring Festival, the epidemic spread rapidly to other cities in and outside Hubei Province. Wuhan is an important railway and aviation hub and the largest city in Central China, with a permanent population of 11.212 million. As of January 23, at least five million people had left Wuhan, two-thirds of whom had gone to other cities in Hubei Province (Chen et al., 2020; Liu et al., 2020). Therefore, Hubei Province was selected as a case to examine the spatiotemporal association between COVID-19 outbreak and population mobility for microcosm studies of COVID-19 across the country.

In this study, the data sources included the data on COVID-19 epidemic and population mobility. The numbers of daily cumulative cases of each administrative unit (excluding Wuhan) from 24 January 2020 to 5 March 2020 were obtained from the website of China Data Lab (https://projects.iq.harvard.edu/chinadata-lab/introduction-cdl), which integrates and releases daily official epidemic data from health commissions at all levels in China. The data of population mobility were derived from Baidu migration index. "Baidu migration" is a big data visualisation project launched by Baidu Inc., and it uses Baidu map to dynamically visualize population migration during Spring Festival in China. Baidu migration provides two types of indices, namely, inter-city and intra-city mobility indices, which are used to measure population flow among and within cities, respectively. For each city within Hubei (excluding Wuhan), the index of daily population inflow from Wuhan and the index of daily intra-city mobility were collected for this study.

The new coronavirus has an incubation period of 1–14 days (Li et al., 2020), which results in a lag in the occurrence of COVID-19 disease. Thus, daily new cases can be associated with the population flow in the previous 14 days. Huanggang was the first city outside of Wuhan in Hubei Province to report COVID-19 cases. The report was released on 20 January 2020. Given the maximum incubation period, the population flow associated with cases on that day can be traced back to the previous 14 days (i.e., from January 6 to January 19). Thus, the population flow data in our study were collected from 6 January 2020 to 4 March 2020.

### 2.2. GTWR Model

As an extension of the GWR model, GTWR not only considers the spatial nonstationarity of geographic data but also incorporates temporal effects into the calculation of the model, thus improving its goodness-of-fit. Formally, GTWR can be expressed as follows:

$$Y_i = \beta_0\left(u_i, v_i, t_i\right) + \sum_{k=1}^{K} \beta_k\left(u_i, v_i, t_i\right) X_{ik} + \varepsilon_i \tag{1}$$

For each observation $i$ ($i = 1, 2, \cdots, n$), $Y_i$ is the dependent variable, whereas $X_{ik}$ is the $k$th explanatory variable. $(u_i, v_i, t_i)$ represents the space-time coordinates of observation $i$; $u_i$ and $v_i$ are the projected spatial coordinates, whereas $t_i$ is the projected temporal coordinate. $\beta_0\left(u_i, v_i, t_i\right)$ is the intercept value, and $\beta_k\left(u_i, v_i, t_i\right)$ denotes the regression coefficient, which is a parameter measuring the influence of explanatory variable $X_{ik}$ on dependent variable $Y_i$. Unlike the fixed coefficient global regression model (e.g., OLS), GTWR allows the parameter estimates to vary across space and time. Therefore, this method can capture the spatio-temporal variations simultaneously.

Based on local weighted least squares, the estimation of parameters $\beta_k\left(u_i, v_i, t_i\right)$ can be given as follows:

$$\widehat{\beta_k}\left(u_i, v_i, t_i\right) = \left[X^T W\left(u_i, v_i, t_i\right) X\right]^{-1} X^T W\left(u_i, v_i, t_i\right) Y \tag{2}$$

where $W(u_i, v_i, t_i) = \mathrm{diag}(w_{i1}, w_{i2}, \cdots, w_{in})$ is the space-time weight matrix, and its diagonal elements $w_{ij} (1 \le j \le n)$ is the weight given to observation point $j$ adjacent to observation point $i$. The determination of $w_{ij}$ is the key in weight matrix calculation. GTWR defines $w_{ij}$ as a spatio-temporal distance decay function, which results in the higher weight of the data points close to observation $i$ in the space–time coordinate system than those located farther from point $i$ (Huang et al., 2010). The most commonly used weighting functions are Gaussian kernel functions:

$$w_{ij} = \exp\left(-\left(d_{ij}^{ST}\right)^2 / h^2\right) \tag{3}$$

where $h$ is a parameter called bandwidth, which is used to control the radial influence range. In practice, the optimal value for bandwidth $h$ can be obtained by cross-validation. $d_{ij}^{ST}$ is the spatial-temporal distance measuring the closeness between observations $i$ and $j$. In the GTWR model, the spatial-temporal distance is defined as a linear combination of spatial distance and temporal distance:

$$d_{ij}^{ST} = \sqrt{\lambda\left[\left(u_i - u_j\right)^2 + \left(v_i - v_j\right)^2\right] + \mu\left(t_i - t_j\right)^2} \tag{4}$$

where $\lambda$ and $\mu$ are scale parameters to balance the spatial and temporal effects, respectively. In particular, if $\mu = 0$ (or $\lambda = 0$), the spatiotemporal distance will be degraded to the traditional GWR distance (or temporally weighted regression (TWR) distance). In other words, GWR and TWR are special cases of GTWR. In practice, the optimized scale parameters can also be obtained by cross-validation (Huang et al., 2010).

### 2.3. Modeling Space-Time Variation in COVID-19 Epidemic Using GTWR

In this study, GTWR was employed to model the space-time association between COVID-19 epidemic and population mobility. The dependent variable $Y_i$ is the number of daily cumulative confirmed cases for each city of Hubei (excluding Wuhan). Two Baidu mobility indices, that is, population inflow index and intra-city mobility index, were selected for the construction of explanatory variables. Given that COVID-19 has an incubation period of 1–14 days, the lag of space-time association between COVID-19 and population flow should be considered in the modeling process. The daily cumulative cases are the sum of all newly confirmed cases in the past days. Given the time lag, the population movement associated with the daily cumulative cases can be traced back to 14 days prior the day when COVID-19 cases were first reported. Then, for observation $i$, we constructed the following explanatory variables.

$$Inflow\_Wh_i\left(u_i, v_i, t_i\right) = \sum_{t=s-14}^{t_i-1} I_{i1}\left(u_i, v_i, t\right), \tag{5}$$

$$Intra\_Mov_i\left(u_i, v_i, t_i\right) = \sum_{t=s-14}^{t_i-1} I_{i2}\left(u_i, v_i, t\right), \tag{6}$$

where $I_{i1}$ is the index of population inflow from Wuhan, $I_{i2}$ is the index of intra-city mobility, and $t$ denotes the observation time, which ranges from the starting time $s$ (i.e., 20 January 2020 when COVID-19 cases were first reported outside of Wuhan in Hubei Province) to $t_i$−1 (i.e., the day before the current observation time $t_i$). Variables $Inflow\_Wh_i$ and $Intra\_Mov_i$ are cumulative population mobility indices. For each city, $Inflow\_Wh_i$ can be used to characterize the impact of population inflow from Wuhan on imported cases, whereas $Intra\_Mov_i$ represents the impact of intra-city population movements on local infection. In addition, the max-min normalization operator was used to normalize all variable values to achieve the same scale and unit for the dependent and explanatory variables. Then, the GTWR model was implemented using GTWR-Addins, a software package that can be installed and run on ArcGIS. This procedure can automatically complete the selection of optimal parameters through cross-validation.

For our dataset, it covered a long time series in temporal dimension. Due to the incubation period (maximum 14 days) of virus and the intervention of external control measures, the spread of COVID-19 over time may show significant changes at different stages. This assumption will be supported by the following experimental results (Section 3.2). More specifically, the time span (from January 24 to March 5) of the study can be divided into three consecutive stages: Stages 1 (from January 24 to February 6), 2 (from February 7 to February 20) and 3 (from February 21 to March 5). In other words, each stage comprised 2 weeks. Stage 1 corresponds to the first two weeks after Wuhan lockdown. In this stage, the implementation of relevant control measures prevented the inflow of population from external cities, and most of the imported latent cases would be found and confirmed. In the second stage, the population flow previously imported from

**Table 1**
*Parameter Estimate Summaries of OLS on the Entire Data Set*

| Variable | Coefficient | *t*-statistic | *t*-probability | VIF |
|---|---|---|---|---|
| Intercept | 0.010 | 3.038 | 0.002* | --- |
| *Inflow_Wh* | 0.117 | 30.162 | 0.000* | 1.007 |
| *Intra_Mov* | −0.010 | −1.837 | 0.067 | 1.007 |
| $R^2 = 0.576$, $R^2_{adj} = 0.575$, RSS = 0.477, AICc = −2958.69 | | | | |

*$p < 0.05$.

Wuhan would play a minimal role in the growth of the number of cases, whereas the confirmed cases might continue to increase due to the local infection. The third stage accounted for the last two weeks. By then, the spread of the epidemic would have been basically controlled. In our work, the GTWR model was used not only in the entire dataset to reveal the spatiotemporal variation trend of COVID-19 epidemic in the entire time period (from January 24 to March 5) but also in the subdatasets of different stages to capture more detailed information.

## 3. Results and Analysis

### 3.1. Model Results and Comparison

Before the implementation of GTWR, we first used global OLS regression to model the relationship between the COVID-19 epidemic and population movement. This model was used as a baseline to evaluate and compare the performance of local modeling techniques. Table 1 shows the estimated results of OLS. The adjusted $R^2$ was 0.575, which indicates that OLS can explain 57.5% of the total variation in the cumulative case counts. The selected influential factors were statistically significant, that is, population mobility is an important factor associated with the spread of COVID-19 in Hubei Province. Given that the variance inflation factor (VIF) values were smaller than 3, the choice of explanatory variables can avoid the problem of collinearity.

Based on the coefficient and t-probability of the variable, the explanatory variable *Inflow_Wh* was positively correlated with COVID-19 cases ($p$-value $< 0.01$). The coefficient value of variable *Intra_Mov* was −0.010 ($p$-value $< 0.1$), indicating its weak negative correlation with COVID-19 infection. The negative correlation may be caused by the intra-city movement of the people who provide public services to fight the epidemic.

Using the same dataset, the GTWR model was used to model the association between the cumulative number of cases and population flows. Table 2 presents the results. Compared with OLS, the fitting results of GTWR have been significantly improved in terms of the adjusted $R^2$, residual standard deviation (RSS) and corrected Akaike information criterion (AICc). Unlike the global OLS model with fixed coefficients, the coefficients of GTWR vary over space and time. Table 2 presents five statistics, including minimum (Min), maximum (Max), lower quartile (LQ), upper quartile (UQ) and mean, which were used to summarize the distribution of each estimated parameter. For variable *Inflow_Wh*, the Min and Max values of the coefficients were −0.072 and 0.338, respectively. These values indicate that the correlation between the number of cases and the population flow has great space-time variation. With LQ = 0.075 > 0, for most observation points, the coefficients were positive. In other words, for variable *Inflow_Wh*, although negative correlations existed for several spatial units or time frames, positive correlations were dominant. Similarly, the coefficients of variable *Intra_Mov* showed spatial-temporal variations ranging from −0.045 to 0.065. However, their absolute values were relatively smaller than those of *Inflow_Wh*, implying that the intra-city population movement has less influence on COVID-19 cases compared with population inflow from Wuhan. To evaluate whether the parameter estimates vary significantly across the study area, the Monte Carlo method was used to test the heterogeneity of parameter estimates, and the results show that the $p$-values of the two coefficients and intercept term are all 0.000(<0.01), which means they are statistically significant.

**Table 2**
*Parameter Estimate Summaries of GTWR on the Entire Data Set*

| Variable | Min | LQ | Mean | UQ | Max |
|---|---|---|---|---|---|
| Intercept | −0.027 | −0.004 | 0.006 | 0.018 | 0.049 |
| *Inflow_Wh* | −0.072 | 0.075 | 0.102 | 0.166 | 0.338 |
| *Intra_Mov* | −0.045 | −0.027 | −0.008 | 0.009 | 0.065 |
| $R^2 = 0.900$, $R^2_{adj} = 0.900$, RSS = 0.112, AICc = −3834.68 | | | | | |

For comparison, local modeling techniques GWR and TWR were also tested on the same dataset. Table 3 reports the evaluation results of these models. GTWR exhibited the best performance, including the highest $R^2$, lowest RSS and lowest AICc. Meanwhile, the TWR model performed better than the GWR model. A possible reason is that the data set covered a small geographical area but involved a relatively long period. Thus, temporal nonstationarity is more significant than spatial nonstationarity.

Given the stage characteristics of COVID-19 epidemic, the above-mentioned models (OLS, GWR, TWR and GTWR) were further used on three subdatasets of different stages. Table 4 presents the fitting performance of these models in the three stages. Compared with the results on the

**Table 3**
*Performance Comparison of Different Models on the Entire Data Set*

| | OLS | GWR | TWR | GTWR |
|---|---|---|---|---|
| $R^2$ | 0.576 | 0.655 | 0.837 | 0.900 |
| $R^2_{adj}$ | 0.575 | 0.654 | 0.836 | 0.900 |
| RSS | 0.477 | 0.389 | 0.184 | 0.112 |
| AICc | −2958.69 | −3060.44 | −3569.36 | −3834.68 |

entire dataset (Table 3), the fitting results of all models on the staged subdatasets significantly improved, and in each stage, the GTWR model performed the best in terms of $R^2$, adjusted $R^2$, RSS and AICc. Based on the adjusted $R^2$, the GTWRs of the three stages can explain 93% (Stage 1), 97.2% (Stage 2) and 99.1% (Stage 3) of the space-time variation in the daily cumulative case counts. From Stages 1 to 3, the fitting performance of the GTWR model improved continuously, implying that the final cumulative number of cases is almost completely determined by population flow data.

In addition, compared with the global OLS model, GWR and TWR showed different improvements in each stage. In the first stage, GWR and TWR exhibited close performances, indicating the balance between spatial and temporal nonstationarities in the case data. However, from Stage 1 to Stage 2, the adjusted $R^2$ for GWR decreased from 0.824 to 0.803, whereas for TWR, the adjusted $R^2$ increased from 0.826 to 0.875, suggesting that the spatial nonstationarity of the data decreased, whereas the temporal nonstationarity increased. That is, the temporal nonstationarity was dominant. In the third stage, for GWR, the adjusted $R^2$ increased to 0.991. However, compared with OLS, the performance of TWR showed almost no improvement, implying that the data were stationary in time, and only spatial heterogeneity existed. In this stage, the epidemic was under control in Hubei Province, and the number of cases in each city was stable and no longer exhibited changes over time. However, evident spatial differences in the number of cumulative cases were still observed among different cities. This finding may explain why the results of GTWR model were almost the same as those of the GWR model. In other words, the GTWR model has degenerated into GWR.

### 3.2. Temporal Variation of Estimated Coefficients

#### 3.2.1. Temporal Variation of *Inflow_Wh* Coefficients

The GTWR model can deal with temporal and spatial heterogeneity simultaneously in the COVID-19 epidemic data and capture more details related to location and time. Figure 1 illustrates the variation in *Inflow_Wh* coefficients over time in 16 cities in Hubei Province (excluding Wuhan). Except for Enshi, the *Inflow_Wh* coefficient values for other cities became all positive (greater than 0.05) over time, indicating that the cumulative number of cases in these cities was positively correlated with the population inflow from Wuhan. For most cities in Hubei Province, such as Xiangyang, Jingzhou and Suizhou, the *Inflow_Wh* coefficients showed similar temporal trends. Specifically, their coefficient values first increased in the first stage (from January 24 to February 6), decreased in the second stage (from February 7 to February 20) and finally stabilized in the third stage (from February 21 to March 5). This finding can be explained by the following: In the first stage (i.e., the first two weeks after Wuhan lockdown), most of the cases imported from Wuhan during the incubation period were identified. Thus, the cumulative confirmed cases increased significantly. In the second stage, given that most of the cases imported before Wuhan lockdown have occurred in the first stage, the impact of previous migration data in this stage rapidly decreased until it disappeared. Afterward, the epidemic has been controlled, and the *Inflow_Wh* coefficients no longer changed with time (Stage 3).

**Table 4**
*Performance Comparison of Models on Subdatasets of Three Different Stages*

| | Stage | OLS | GWR | TWR | GTWR |
|---|---|---|---|---|---|
| $R^2$ | 1 | 0.728 | 0.826 | 0.828 | 0.931 |
| | 2 | 0.706 | 0.805 | 0.876 | 0.973 |
| | 3 | 0.877 | 0.991 | 0.877 | 0.991 |
| $R^2_{adj}$ | 1 | 0.728 | 0.824 | 0.826 | 0.930 |
| | 2 | 0.706 | 0.803 | 0.875 | 0.972 |
| | 3 | 0.877 | 0.991 | 0.876 | 0.991 |
| RSS | 1 | 0.182 | 0.117 | 0.116 | 0.047 |
| | 2 | 0.067 | 0.044 | 0.028 | 0.006 |
| | 3 | 0.009 | 0.001 | 0.010 | 0.001 |
| AICc | 1 | −951.72 | −1016.01 | −1020.12 | −1149.22 |
| | 2 | −1177.51 | −1235.92 | −1347.67 | −1616.67 |
| | 3 | −1605.44 | −2172.15 | −1602.18 | −2144.13 |

However, the *Inflow_Wh* coefficient curves of certain cities, such as Shiyan, Shennongjia and Enshi, showed different patterns of variation. For Shiyan and Shennongjia, their *Inflow_Wh* coefficients were larger than those of other cities in the first stage, implying the stronger correlation between the population flow from Wuhan and COVID-19 cases in these cities. By contrast, the coefficient curve for Enshi notably differed from that of other cities. Particularly, from January 30 to February 15, the *Inflow_Wh* coefficients were all negative. This result was attributed to the status of Enshi as the only autonomous prefecture by ethnic minority in

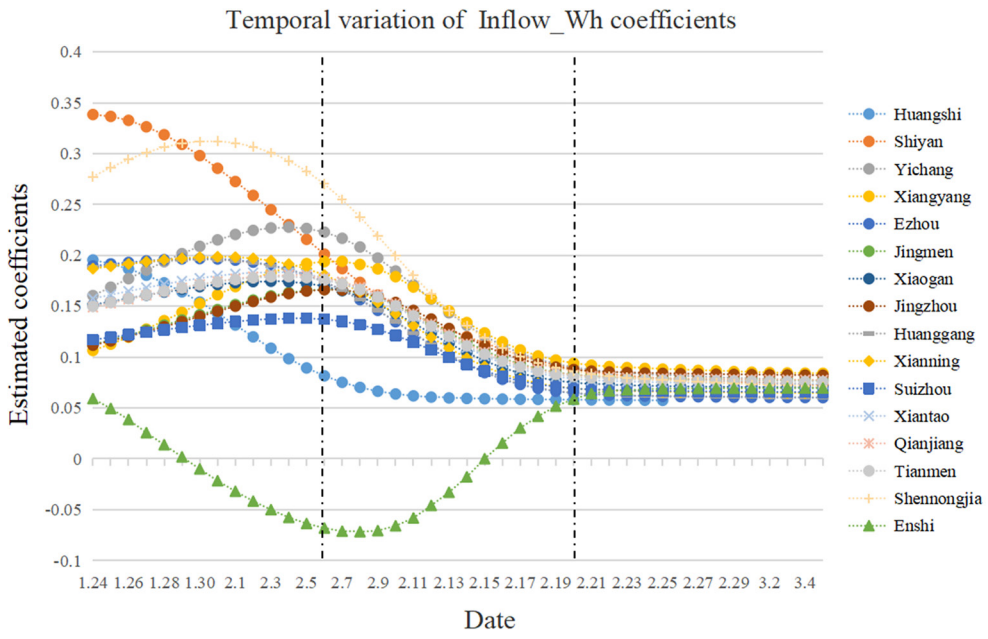


**Figure 1.** Temporal variation in the estimated coefficients for *Inflow_wh*.

Hubei Province. Most of the floating population (mainly rural migrant workers) in Enshi went to several cities outside Hubei rather than the provincial capital Wuhan. Relatively few people, including the infected ones, returned to Enshi from Wuhan during Spring Festival.

### 3.2.2. Temporal Variation of *Intra_Mov* Coefficients

The intra-city population mobility is another factor associated with the COVID-19 epidemic. Figure 2 presents the temporal variation of *Intra_Mov* coefficients for 16 cities in Hubei Province. Compared with the above *Inflow_Wh* coefficients, the coefficient estimates for *Intra_Mov* showed more diverse patterns of variation. They can be either positive or negative during the entire period, indicating that intra-city population mobility has either positive or negative impacts on COVID-19 spread over time.

Overall, the curves of *Intra_Mov* coefficients for most cities showed a similar trend of variation, that is, they decreased first and then gradually remained stable. This condition is due to the timely control and preventive measures of cities following the lockdown of Wuhan, such as travel restrictions, wearing of face masks and social distancing, which greatly reduced the impact of intra-city population flows on the epidemic. Among these cities, Jingzhou and Jingmen obtained relatively higher coefficient estimates, and the values were always positive for the former in the entire period and for the latter before February 23. As a result, continuous cases of local infection might have been caused by intra-city population movement in both cities.

In addition, the coefficient estimates of three cities, including Shiyan, Xiangyang and Enshi, showed another pattern of variation over time. That is, they first increased, then decreased and finally remained stable. Especially for the city of Enshi, the estimated coefficients were always positive before February 16, whereas for Shiyan and Xiangyang, from the second week to the fourth week, their coefficient estimates were also mostly positive. These results indicate that the intra-city population movement is an important factor driving the increase in the cumulative number of cases for these cities. In addition, although these cities have implemented closure measures after Wuhan lockdown, the estimated coefficients continually increased for a certain period, implying that before the closure of these cities, the intra-city population movement had caused the local spread of the virus, and those who were in incubation period were identified within the next two weeks. For example, the city of Xiangyang, one of the last closed cities in Hubei Province, was not

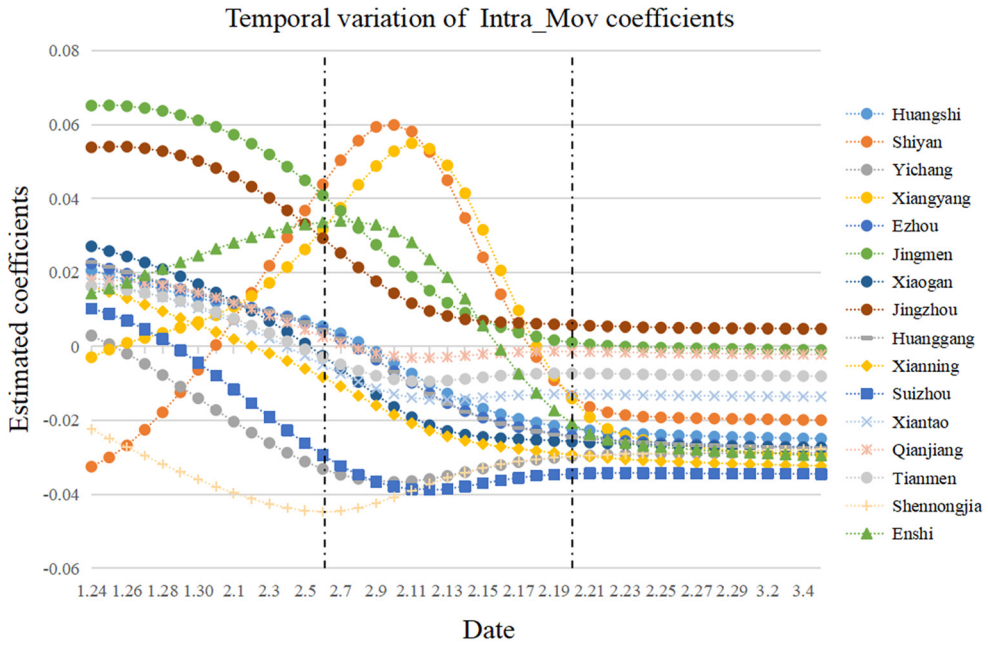


**Figure 2.** Temporal variation in the estimated coefficients for *Intra_Mov*.

officially closed until January 28. The delayed closure could have led to more local infections, which may explain the continuous increase in the estimated coefficients until February 11.

### 3.3. Spatial Variation of Estimated Coefficients

#### 3.3.1. Spatial Variation of *Inflow_Wh* Coefficients

The average values of *Inflow_Wh* coefficients (in the temporal dimension) were obtained to more intuitively understand the spatial variation in the estimated coefficients among different cities (Figure 3). The *Inflow_Wh* coefficients presented different patterns of spatial distribution in the three periods. Based on the natural breaks of average coefficients, the cities were divided into four groups (i.e., high-value, sub-high value, sub-low value and low-value). In the first stage, the *Inflow_Wh* coefficients ranged from −0.011 to 0.297, showing a large spatial variation. The high and sub-high values were mainly distributed in the west of Hubei Province (Shiyan, Shennongjia and Yichang) and the eastern areas around Wuhan (Huanggang, Ezhou, Huangshi, Xianning and Xiantao), implying that the population inflow from Wuhan had a great impact on these cities. In the second and third stages, although the *Inflow_Wh* coefficients for most cities had declined greatly, they still presented evident spatial differences and aggregations. The high and sub-high values were mainly concentrated in the central cities (e.g., Xiangyang and Jingmen) of Hubei Province, moving from the east and west of Hubei Province to the central region.

#### 3.3.2. Spatial Variation of *Intra_Mov* Coefficients

Figure 4 presents the spatial distribution of the average *Intra_Mov* coefficient estimates (in the temporal dimension) for the three stages. The values showed remarkable spatial variation in each stage. In the first stage, for most cities (such as those around Wuhan) in Hubei Province, the average of their estimated coefficients was positive, indicating that the growth in the number of cases in these cities had a positive correlation with local population flows. That is, the intra-city population movement accelerated the local spread of COVID-19. The high-value group of cities include Jingzhou and Jingmen, which are large cities in central and southern Hubei Province. With a negative coefficient estimate, the low-value group only included Shennongjia Forestry District. This city is located in the mountainous western region. The forest coverage rate is as high as 85%, and the population density is the lowest in Hubei Province. In the second stage, the high-value group included more cities (e.g., Jingzhou, Xiangyang and Shiyan.), which are mainly

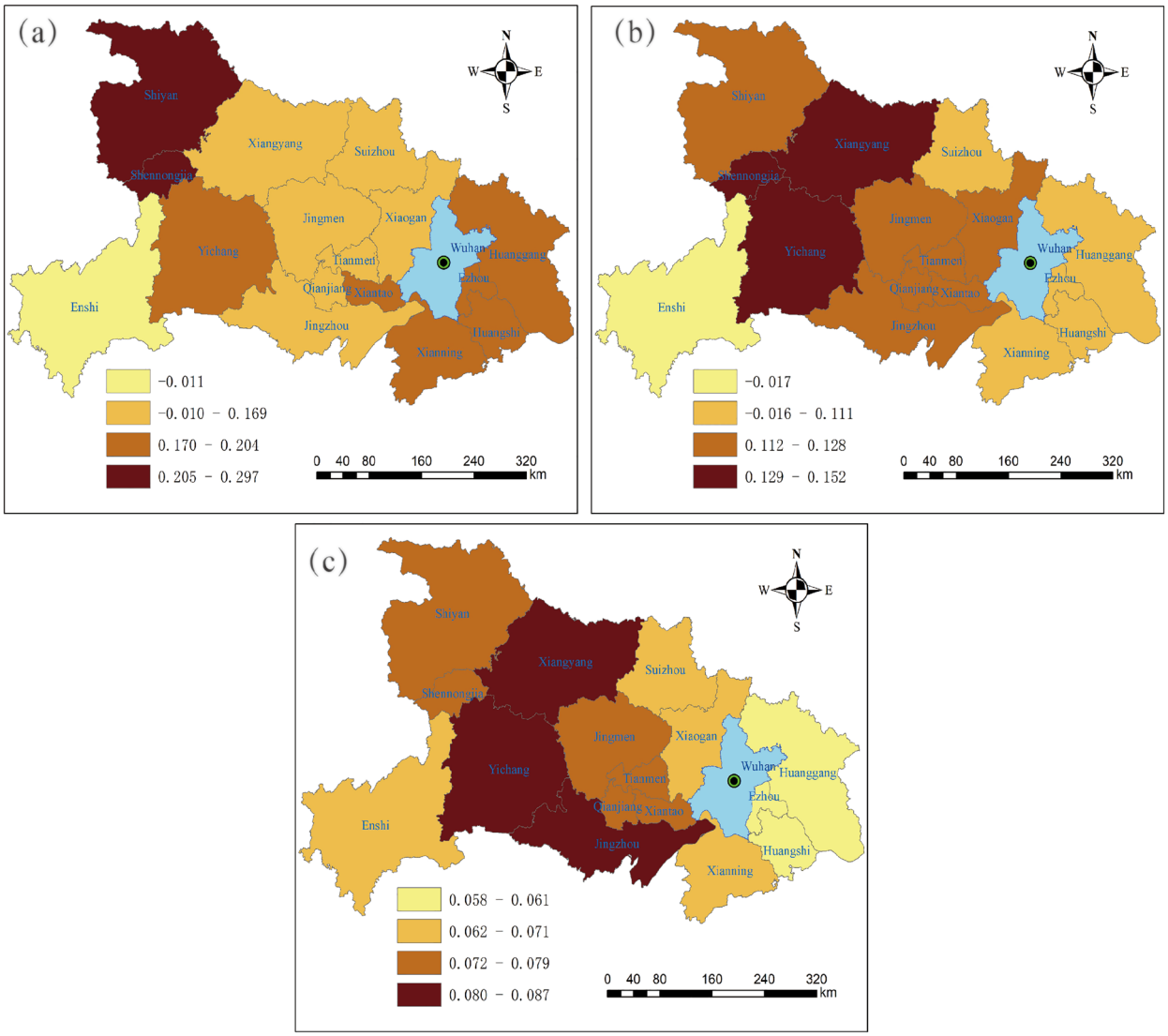


**Figure 3.** Spatial distribution of the average coefficients for *Inflow_Wh* in three different stages. (a) Stage 1; (b) Stage 2; (c) Stage 3.

concentrated in the central and western regions of Hubei Province. Particularly, a zonal distribution was observed along the "Jingzhou-Jingmen-Xiangyang-Shiyan". In addition, these cities are regional central cities of Hubei Province with relatively large population flows. However, in the third stage, three cities (i.e., Jingzhou, Jingmen and Qianjiang) belonged to the high-value group. Jingzhou and Jingmen are always in the high-value group, implying that the impact of intra-city population movements on the epidemic almost occurred in the entire period of our study.

In addition, in the second and third stages, except for the high-value group, the average coefficient estimates for the other group of cities were all negative, indicating the negative correlation between the COVID-19 epidemic and intra-city population flows. The reason may be that with the implementation of prevention and control measures, residents were strictly restricted to travel and required to wear face masks when going out, which prevented the further spread of the virus among people. In these stages, the floating population mainly included those who provide public services for the fight against the epidemic, such as doctors, police, drivers and volunteers. These population flows had played a key role in reducing the number of cases.

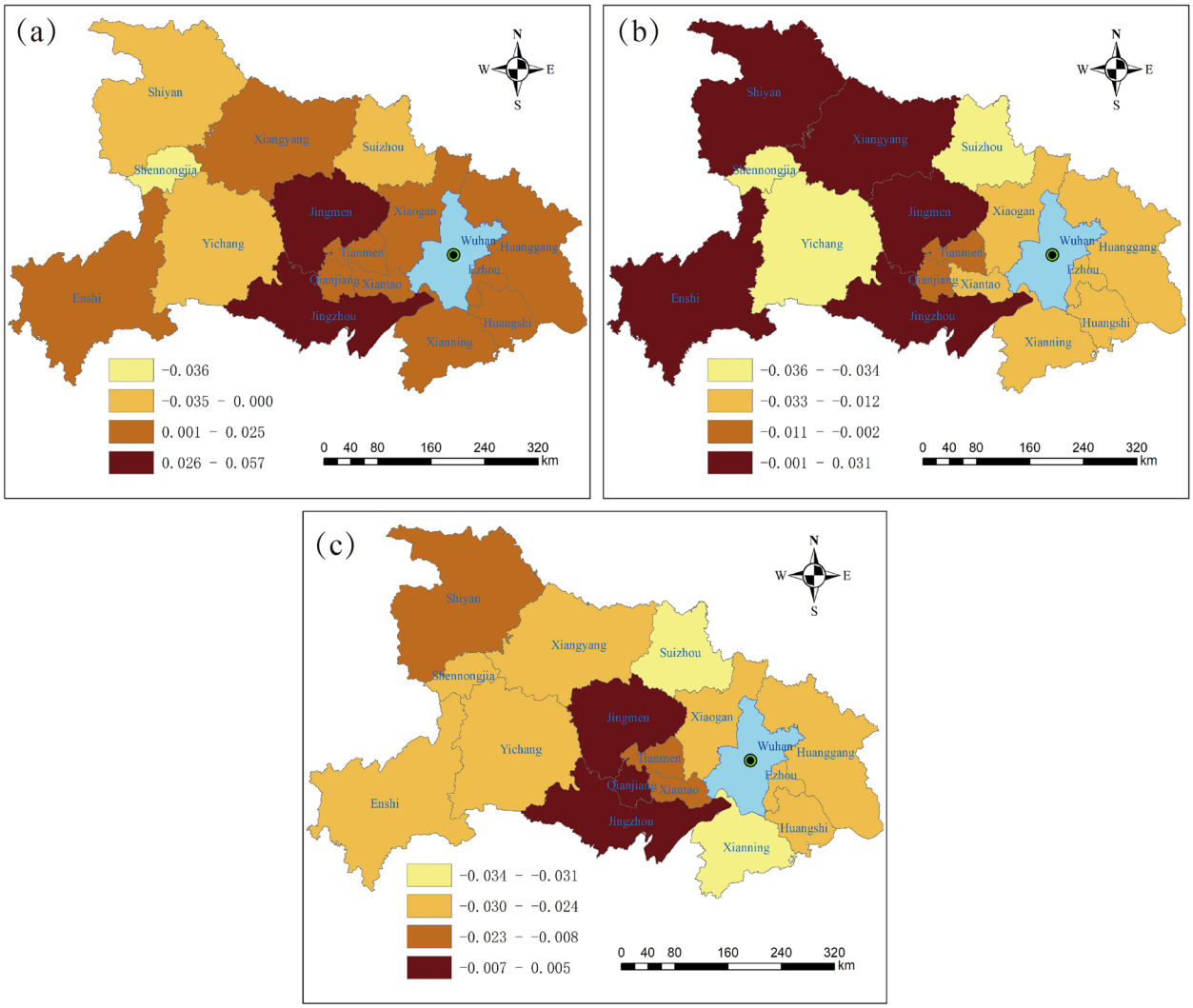


**Figure 4.** Spatial distribution of the average coefficients for *Intra_Mov* in three different stages. (a) Stage 1; (b) Stage 2; (c) Stage 3.

## 4. Discussion

This study investigated the spatiotemporal relationships between COVID-19 epidemic and population mobility. Considering Hubei Province in China as a case study, we used the GTWR model to model the relationship between daily cumulative case counts and population flows. Compared with the classical regression techniques (e.g., OLS, GWR, and TWR), the GTWR model showed the best performance based on the goodness of fit. Particularly, on the staged sub-datasets, the performance of GTWR was further improved. The goodness of fit increased from 93% in the first stage to 99.1% in the third stage, implying that the final cumulative number of cases was almost completely determined by population flow data. On the one hand, this result verified the effectiveness of GTWR in modeling the spatiotemporal heterogeneity of COVID-19 transmission. On the other hand, this result also supports previous findings indicating that the size of the epidemic at the early stage can be well predicted by the volume of population outflow from Wuhan (Kraemer et al., 2020), and the spreading pattern of COVID-19 gradually converged to the distribution of human movement from Wuhan to other cities in China (Jia et al., 2020).

Local human mobility for a city includes inter-city and intra-city patterns (Liu et al., 2020). For the inter-city pattern, this study only considered the population inflow from Wuhan, because at the early stage, the imported cases mainly came from Wuhan. For most of the cities in Hubei Province, the population flow from Wuhan had a positive impact on the number of COVID-19 cases, and this effect showed three stages

of spatiotemporal variation. From the first stage to the third stage, the effect increased first, then decreased and finally remained stable. The stage characteristics of this epidemic can be explained as follows: First, the virus has a maximum incubation period of 14 days (i.e., 2 weeks); second, the implementation of strict control measures (Wuhan lockdown followed by other cities) prevented the further occurrence of imported cases.

Meanwhile, intra-city population mobility showed either positive or negative impacts on the COVID-19 spread over time. For most cities in Hubei Province, this effect showed a time-varying trend of decreasing at the first two stages and then stabilizing at the last stage. Population movements within a city can result in local infection of this epidemic, thereby increasing the number of COVID-19 cases. On the other hand, population movements can also play a significant role in tackling the COVID-19 emergency by providing essential public health services and relief forces in relevant sectors (Sannigrahi et al., 2020).

From the perspective of space, the coefficients of variables *Inflow_Wh* and *Intra_Mov* presented notable spatial variations and clustering. In epidemiological research, except for person and time, location has been regarded as an important dimension of disease processes (Sun et al., 2020). Several studies have focused on the spatial characteristics and correlation factors of COVID-19 transmission (Andersen et al., 2020; Franch-Pardo et al., 2020; Sannigrahi et al., 2020). For COVID-19 disease, heterogeneous spatial distribution may be associated with a variety of factors, such as geography, socioeconomics, demographics, environment, topography and behavior (Mollalo et al., 2020). However, according to the model results of this study, population mobility is the most direct factor that resulted in the outbreak of epidemic in each city of Hubei Province. Various cities, due to the differences in geographical conditions, social economy and urban policies, exhibited varied population movements, which led to spatial disparities in COVID-19 cases. Particularly, under the strict lockdown policies of China, the variations in *Intra_Mov* coefficients can reflect the implementation effect of prevention and control measures in a city. For example, the delayed lockdown for the city of Xiangyang resulted in the continuously increasing impact of intra-city population movements on the local infection until February 11. Likewise, the abnormality of coefficient curves (always positive) for the cities of Jingzhou and Jingmen imply the continuous occurrence of local infection caused by intra-city population movement.

This study encountered several limitations. For example, the dataset used covered a relatively long period of time but a small spatial area, which might have led to inadequacy in spatial nonstationarity. In a larger spatial area (e.g., the whole area of China), spatial effects, including spatial dependence and spatial heterogeneity, are usually more significant. How population mobility affects the spread of COVID-19 deserves further study. When GTWR is used to model this relationship, it is necessary to design a more efficient spatiotemporal weight matrix to better quantify the complex spatiotemporal relationship caused by the increased spatial effects. In addition to population mobility, the temporal and spatial distributions of cumulative cases may be affected by other factors, such as geographic location, temperature, humidity, topography, city size, population density and control measures. These factors can be considered to construct explanatory variables to further improve the fitting performance of the GTWR model.

## 5. Conclusions

This study employed the GTWR model to quantitatively evaluate the spatiotemporal association between COVID-19 and population movements. A case study was carried out in Hubei Province, China. Results indicate that the GTWR model can explain over 90% of the variation and possible associated factors for COVID-19 cases in terms of the adjusted $R^2$. The association between COVID-19 case counts and population movements presented significant variations over time and space. In the time dimension, the association showed three stages of variation characteristics due to the virus incubation period and the implementation of strict lockdown measures to restrict resident travel. For most cities in Hubei Province, the population flow from Wuhan has a positive impact on the spread of COVID-19, and this impact shows a trend of increasing in the first stage, then decreasing in the second stage, and finally maintaining stable in the third stage. In contrast, the impact of intra-city population movement on the spread of the epidemic is more diverse, either positive or negative. For most cities, it keeps declining in the first two stages and then remains stable in the third stage. In space, strong geographical disparities were observed across Hubei Province. The

direct reason is that the population inflow from Wuhan and intra-city population movement varied among cities due to the differences in geographical location, population size and urban policy.

Given the significant impact of population mobility on COVID-19 transmission, the implementation of control measures by restricting population movements can play a key role in mitigating the spread of this epidemic. Moreover, the timing and strictness of control measures may result in the variation in spatiotemporal distribution of COVID-19 cases. Thus, targeted interventions, if necessary in certain periods, can be considered to further restrict population movement in cities with high transmission risk.

## Conflict of Interest

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

## Data Availability Statement

Datasets for this research can be downloaded from the website of China Data Lab, Harvard University(https://projects.iq.harvard.edu/chinadatalab/introduction-cdl). More specifically, the COVID-19 daily case data are available online in https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/MR5IJN, while the Baidu migration index data are in https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FAEZIO.

**Acknowledgments**
This study was funded by the National Natural Science Foundation of China (No.41901326), the Natural Science Foundation of Jiangsu Province (No. BK20190742) and the Natural Science Research of Jiangsu Higher Education Institutions of China (No. 19KJB170009). This work was also funded by the Open Foundation of Smart Health Big Data Analysis and Location Services Engineering Lab of Jiangsu Province (No. SHEL221003).

## References

Andersen, L. M., Harden, S. R., Sugg, M. M., Runkle, J. D., & Lundquist, T. E. (2020). Analyzing the spatial determinants of local COVID-19 transmission in the United States. *Science of the Total Environment*, *754*, 142396. https://doi.org/10.1016/j.scitotenv.2020.142396

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, *28*, 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Chen, S., Yang, J., Yang, W., Wang, C., & Bärnighausen, T. (2020). COVID-19 control in China during mass population movements at New Year. *The Lancet*, *395*(10226), 764–766. https://doi.org/10.1016/s0140-6736(20)30421-9

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Chichester, UK: John Wiley & Sons.

Franch-Pardo, I., Napoletano, B. M., Rosete-Verges, F., & Billa, L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment*, *739*, 140033. https://doi.org/10.1016/j.scitotenv.2020.140033

Ge, L., Zhao, Y., Sheng, Z., Wang, N., Zhou, K., Mu, X., et al. (2016). Construction of a seasonal difference-geographically and temporally weighted regression (SD-GTWR) model and comparative analysis with gwr-based models for hemorrhagic fever with renal syndrome (HFRS) in Hubei Province (China). *IJERPH*, *13*, 1062. https://doi.org/10.3390/ijerph13111062

He, Q., & Huang, B. (2018). Satellite-based mapping of daily high-resolution ground PM2.5 in China via space-time regression modeling. *Remote Sensing of Environment*, *206*, 72–83. https://doi.org/10.1016/j.rse.2017.12.018

Hong, Z., Hao, H., Li, C., Du, W., Wei, L., & Wang, H. (2018). Exploration of potential risks of hand, foot, and mouth disease in inner Mongolia autonomous region, China using geographically weighted regression model. *Scientific Reports*, *8*, 17707. https://doi.org/10.1038/s41598-018-35721-9

Huang, B., Wu, B., & Barry, M. (2010). Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, *24*(3), 383–401. https://doi.org/10.1080/13658810802672469

Jia, J. S., Lu, X., Yuan, Y., Xu, G., Jia, J., & Christakis, N. A. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*, *582*, 389–394. https://doi.org/10.1038/s41586-020-2284-y

Kraemer, M. U. G., Golding, N., Bisanzio, D., Bhatt, S., Pigott, D. M., Ray, S. E., et al. (2019). Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Scientific Reports*, *9*, 5151. https://doi.org/10.1038/s41598-019-41192-3

Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*, *368*(6490), 493–497. https://doi.org/10.1126/science.abb4218

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, *382*(13), 1199–1207. https://doi.org/10.1056/nejmoa2001316

Liu, H., Bai, X., Shen, H., Pang, X., Liang, Z., & Liu, Y. (2020). Synchronized travel restrictions across cities can be effective in COVID-19 control. *medRxiv* https://doi.org/10.1101/2020.04.02.20050781

Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems*, *70*, 113–124. https://doi.org/10.1016/j.compenvurbsys.2018.03.001

Merler, S., & Ajelli, M. (2012). Human mobility and population heterogeneity in the spread of an epidemic. *Procedia Computer Science*, *1*, 2237–2244. https://doi.org/10.1016/j.procs.2010.04.250

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

Pan, J., & Lai, J. (2019). Spatial pattern of population mobility among cities in China: Case study of the National Day plus Mid-Autumn Festival based on Tencent migration data. *Cities*, *94*, 55–69. https://doi.org/10.1016/j.cities.2019.05.022

Ren, H., Zheng, L., Li, Q., Yuan, W., & Lu, L. (2017). Exploring determinants of spatial variations in the dengue fever epidemic using geographically weighted regression model: A case study in the joint Guangzhou-Foshan Area, China, 2014. *IJERPH*, *14*, 1518. https://doi.org/10.3390/ijerph14121518

Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., & Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*, *62*, 102418. https://doi.org/10.1016/j.scs.2020.102418

Song, C., Shi, X., Bo, Y., Wang, J., Wang, Y., & Huang, D. (2019). Exploring spatiotemporal nonstationary effects of climate factors on hand, foot, and mouth disease using Bayesian Spatiotemporally Varying Coefficients (STVC) model in Sichuan, China. *Science of the Total Environment*, *648*, 550–560. https://doi.org/10.1016/j.scitotenv.2018.08.114

Sun, F., Matthews, S. A., Yang, T.-C., & Hu, M.-H. (2020). A spatial analysis of the COVID-19 period prevalence in U.S. counties through June 28, 2020: where geography matters? *Annals of Epidemiology*. https://doi.org/10.1016/j.annepidem.2020.07.014

Tao, R., & Thill, J.-C. (2020). BiFlowLISA: Measuring spatial association for bivariate flow data. *Computers, Environment and Urban Systems*, *83*, 101519. https://doi.org/10.1016/j.compenvurbsys.2020.101519

Tian, H., Liu, Y., Li, Y., Wu, C.-H., Chen, B., Kraemer, M. U. G., et al. (2020). An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*, *368*(6491), 638–642. https://doi.org/10.1126/science.abb6105

Wu, C., Ren, F., Hu, W., & Du, Q. (2019). Multiscale geographically and temporally weighted regression: exploring the spatiotemporal determinants of housing prices. *International Journal of Geographical Information Science*, *33*(3), 489–511. https://doi.org/10.1080/13658816.2018.1545158

Yang, C., Sha, D., Liu, Q., Li, Y., Lan, H., Guan, W. W., et al. (2020). Taking the pulse of COVID-19: A spatiotemporal perspective. *International Journal of Digital Earth*, *13*(10), 1186–1211. https://doi.org/10.1080/17538947.2020.1809723

Yang, Z., Zeng, Z., Wang, K., Wong, S.-S., Liang, W., Zanin, M., et al. (2020). Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, *12*(3), 165–174. https://doi.org/10.21037/jtd.2020.02.64

Zhang, W., Chong, Z., Li, X., & Nie, G. (2020). Spatial patterns and determinant factors of population flow networks in China: Analysis on Tencent location big data. *Cities*, *99*, 102640. https://doi.org/10.1016/j.cities.2020.102640