



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

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



# Spatial disparities of self-reported COVID-19 cases and influencing factors in Wuhan, China

Gang Xu<sup>a</sup>, Yuhan Jiang<sup>b</sup>, Shuai Wang<sup>c</sup>, Kun Qin<sup>a,\*</sup>, Jingchen Ding<sup>a</sup>, Yang Liu<sup>b</sup>, Binbin Lu<sup>a,\*</sup>

<sup>a</sup> School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

<sup>b</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

<sup>c</sup> Wuhan Geomatics Institute, Wansongyuan Road, Wuhan 430022, China

## ARTICLE INFO

### Keywords:

Covid-19  
Self-reported cases  
Geographically weighted regression (GWR)  
Urban planning  
Urban sustainability

## ABSTRACT

The lack of detailed COVID-19 cases at a fine spatial resolution restricts the investigation of spatial disparities of its attack rate. Here, we collected nearly one thousand self-reported cases from a social media platform during the early stage of COVID-19 epidemic in Wuhan, China. We used kernel density estimation (KDE) to explore spatial disparities of epidemic intensity and adopted geographically weighted regression (GWR) model to quantify influences of population dynamics, transportation, and social interactions on COVID-19 epidemic. Results show that self-reported COVID-19 cases concentrated in commercial centers and populous residential areas. Blocks with higher population density, higher aging rate, more metro stations, more main roads, and more commercial point-of-interests (POIs) have a higher density of COVID-19 cases. These five explanatory variables explain 76% variance of self-reported cases using an OLS model. Commercial POIs have the strongest influence, which increase COVID-19 cases by 28% with one standard deviation increase. The GWR model performs better than OLS model with the adjusted  $R^2$  of 0.96. Spatial heterogeneities of coefficients in the GWR model show that influencing factors play different roles in diverse communities. We further discussed potential implications for the healthy city and urban planning for the sustainable development of cities.

## 1. Introduction

The COVID-19 epidemic has spread worldwide. As of September 30, 2021, there were 230 million COVID-19 cases with 4.7 million deaths across the world (Dong et al., 2020). Urban residents are the primary victims of this epidemic as more than 90% COVID-19 cases lived in cities (United Nations, 2020; Xu et al., 2021b). Therefore, it is of great significance to investigate spatial disparities and determinants of the COVID-19 epidemic through an urban lens (Acuto et al., 2020). The intra-urban analysis of the COVID-19 epidemic can not only reveal the spread of infectious diseases within cities but also support the building of a healthy city in the post-pandemic era (Frumkin, 2021; Megahed and Ghoneim, 2020).

Attack rates of the COVID-19 epidemic present apparent spatial heterogeneities (Das et al., 2021). Infectious diseases spread through human-to-human contact and diffuse with mobility of urban residents. Thus, spatial heterogeneities of infectious diseases are strongly correlated with urban population dynamics, including spatial aggregation of

infected individuals, individual dispersal characteristics, social interactions and contact patterns (Real and Biek, 2007; Sun et al., 2020). In addition, the demographic structure strongly influences the COVID-19 epidemic in space, because elderly people are more vulnerable to the virus (Team, 2020). Socioeconomic status, such as income and education, are also correlated with COVID-19 outcomes (Chang et al., 2021; Drefahl et al., 2020). Low-income people are less likely to follow social distancing and other prevention policy, which in turn leads to higher morbidity and mortality (Mena et al., 2021). Built environment and layout of urban form are associated with COVID-19 incidence rates (Kashem et al., 2021). Physical environment, such as air quality and meteorological factors, also influence the spread of COVID-19 epidemic (Zhang et al., 2020).

However, most of these studies are on the country or provincial level, while limited studies investigated the spatio-temporal variations of the COVID-19 epidemic at the intra-urban scale (Cordes and Castro, 2020; Maroko et al., 2020). Intra-urban studies of the COVID-19 epidemic require locations of COVID-19 cases at a fine scale, while official

\* Corresponding author.

E-mail addresses: [xugang@whu.edu.cn](mailto:xugang@whu.edu.cn) (G. Xu), [jiangyuhan415@whu.edu.cn](mailto:jiangyuhan415@whu.edu.cn) (Y. Jiang), [2010282050194@whu.edu.cn](mailto:2010282050194@whu.edu.cn) (S. Wang), [qink@whu.edu.cn](mailto:qink@whu.edu.cn) (K. Qin), [2017302590157@whu.edu.cn](mailto:2017302590157@whu.edu.cn) (J. Ding), [yang\\_liu@whu.edu.cn](mailto:yang_liu@whu.edu.cn) (Y. Liu), [binbinlu@whu.edu.cn](mailto:binbinlu@whu.edu.cn) (B. Lu).

<https://doi.org/10.1016/j.scs.2021.103485>

Received 3 February 2021; Received in revised form 30 September 2021; Accepted 21 October 2021

Available online 25 October 2021

2210-6707/© 2021 Elsevier Ltd. All rights reserved.

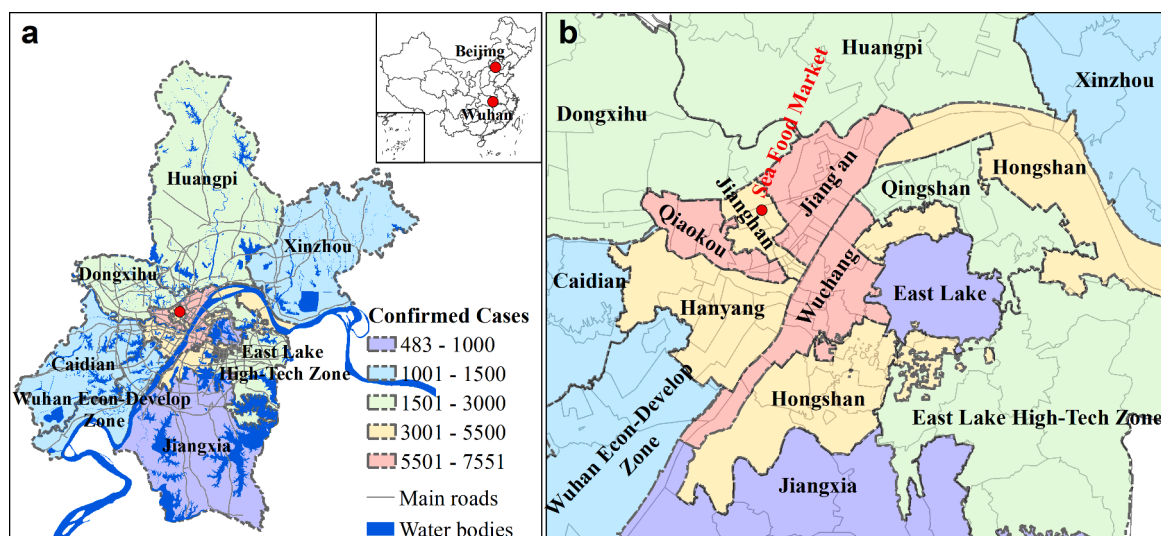


Fig. 1. Spatial distributions of accumulative confirmed COVID-19 cases at the district level in Wuhan, China (a) and in the main urban areas (b).

statistics of the epidemic are basically released in accordance with administrative units (Yang et al., 2021). It is difficult to obtain spatial distributions of individual cases. Social media data is an alternative data source in intra-urban scale modeling of the COVID-19 epidemic, which has been used in COVID-19 related studies (Li et al., 2020; Peng et al., 2020).

Wuhan was the first city in China to report the COVID-19 epidemic, and also the most affected city in China. The local government of Wuhan published confirmed COVID-19 cases in 13 administrative districts in the city. However, the district-level data is too coarse to reveal spatial disparities of infection intensity within a city. In this study, we collected nearly one thousand self-reported COVID-19 cases in Wuhan through a Chinese social media platform (*Weibo*, like *Twitter*), which was used to represent spatial disparities of attack rates of the COVID-19 epidemic within the city. We further chose five explanatory variables, namely, population density, aging rate, metro station, main roads, and commercial point-of-interests (POIs), to quantify their influences on spatial heterogeneities of COVID-19 epidemic in Wuhan. Considering the spatial nonstationary of self-reported COVID-19 cases and influencing factors, we adopted the geographically weighted regression (GWR) model and compared results of the GWR model with the global ordinary least squares (OLS) model (Maiti et al., 2021; Mollalo et al., 2020; Xu et al., 2021a).

## 2. Study area and data

### 2.1. Study area

Wuhan is the capital city of Hubei Province, located in the eastern part of the Jiangnan Plain and the middle reaches of the Yangtze River (Fig. 1). It is the intersection of the Yangtze River and its largest tributary, the Han River, and thus forms a pattern of three parts (Hankou, Wuchang, and Hanyang) (Fig. 1). In 2019, more than ten million people lived in Wuhan, with the GDP over 1.6 trillion RMB. Wuhan is the largest inland water-land-air transportation hub in Central China, and it is the only city in Central China with direct flights to five continents around the world.

Most initial COVID-19 cases reported in Wuhan were exposed to one seafood market in Jiangnan District, Wuhan (Fig. 1). As of May 18, 2020 that was the date of the last COVID-19 case in Wuhan, there were a total of 50,340 confirmed COVID-19 cases in Wuhan, with 3,869 deaths ([http://www.wuhan.gov.cn/zwjk/tzgg/202005/t20200521\\_1325022.shtml](http://www.wuhan.gov.cn/zwjk/tzgg/202005/t20200521_1325022.shtml)). The overall mortality rate was 7.7% of the whole city. The most

severely affected areas were Wuchang, Jiang' an, Jiangnan and Qiaokou districts on both sides of the Yangtze River. Wuchang District had the largest number of confirmed cases, reaching 7551 cases, and Jiangnan District had the highest morbidity that was 1.23%.

### 2.2. Self-reported COVID-19 cases

In the early stage of the COVID-19 epidemic in Wuhan, medical resources and other aspects of responses were not prepared in time. As a result, some infected or suspected cases could not be admitted to hospitals in time. One of the China's largest social media platform, *Weibo* (like *Twitter*), opened up a channel for help, allowing patients who were self-reported as infected cases to post their symptoms and onset time on the platform (Li et al., 2020; Peng et al., 2020). The information was collected, sorted and sent to the local government to better help patients.

We collected the information related to the self-reported COVID-19 cases on *Weibo* using the *Python* crawler technology. There were 910 self-reported cases in total, including the age, time of onset, where they live, date of post, and report text. Among them, there were a total of 693 cases who reported their specific time of illness. We completed part of the incomplete information based on the report text. We acquired the latitude and longitude of the location according to the text address for spatial analysis and modeling. Detailed information (age, location, post time, and confirmed time) about 910 self-reported COVID-19 cases was shared on the GitHub after removing personal information ([https://github.com/Inn905/COVID19\\_Self-reported\\_Data\\_Weibo](https://github.com/Inn905/COVID19_Self-reported_Data_Weibo)).

### 2.3. Explanatory variables

We choose the following five explanatory variables to explain spatial disparities of intensity of self-reported COVID-19 cases, namely, population density, aging rate (over 60 years old), metro stations, main roads, and commercial point-of-interests (POIs). The population density and aging rate are at the street block level, which were from the annual population survey in 2014. Transportation is an essential medium for disease transmission. We use metro stations and road net to represent public and private transportation, respectively. Metro stations and main road net were from the *Gaode Map* (<https://www.amap.com>) in 2019. The metro system counts for more than 50% inter-city commuters in Wuhan. Commercial activities correspond to the frequent contact and interaction between people within a city, and there is a high probability of indirect contact with strangers. We collect more than 120,000 commercial POIs from the *Gaode Map*, which cover shopping malls, hotels,

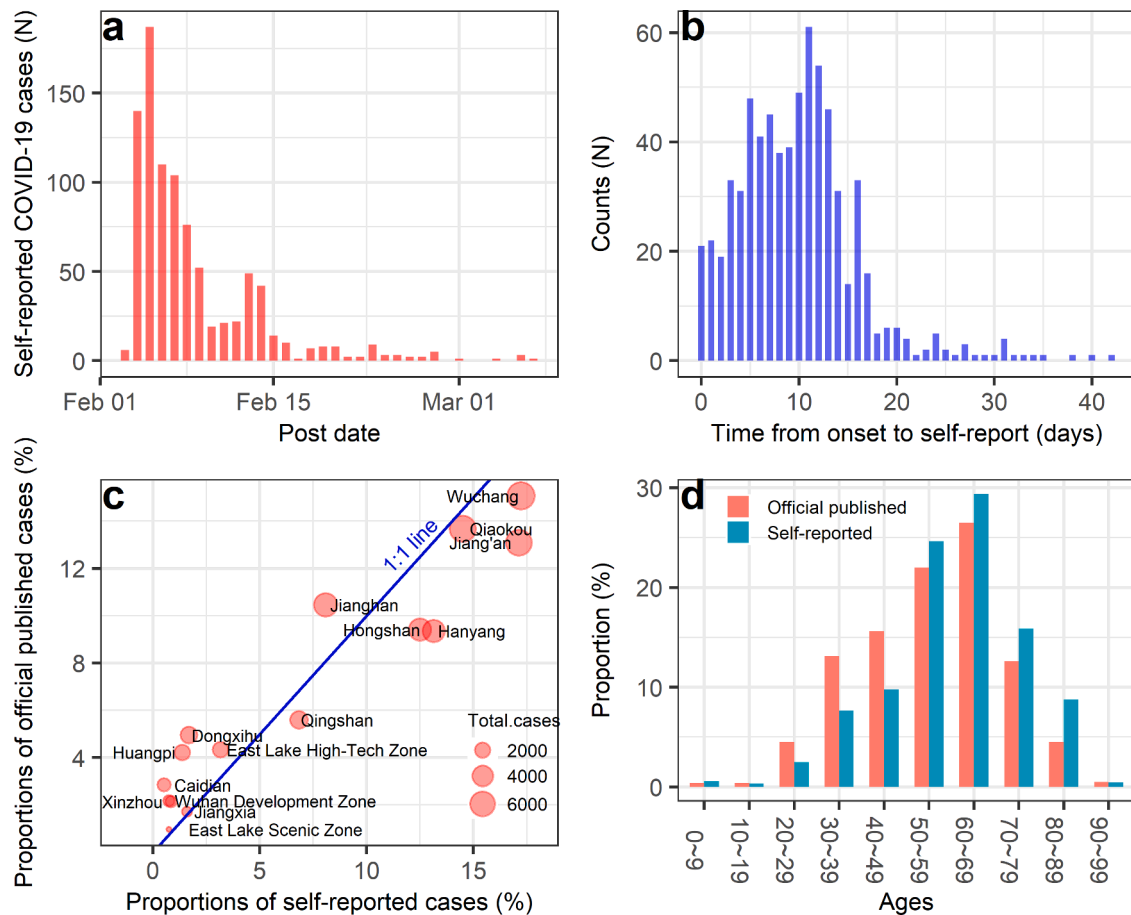


Fig. 2. Temporal variations of self-reported COVID-19 cases and comparisons with official data. (a) Numbers of self-reported COVID-19 cases on each day from February 3 to March 7, 2020; (b) Time delays between the time of onset and self-report on the social media (*Weibo* in China). (c) Proportions of self-reported COVID-19 cases and official published cases at the district level; (d) Frequency distributions of ages between self-reported COVID-19 cases and official published cases.

restaurants, leisure and entertainment sites, and life service sites.

The spatial extent is the main urban area of Wuhan with the block as the analysis unit. The average size of street blocks is 0.72 km<sup>2</sup>, with the smallest block of 0.012 km<sup>2</sup> and the largest block of 15.8 km<sup>2</sup>. There are 1107 blocks in total. A block is an area enclosed by main roads. In Chinese cities, this is a more refined spatial unit than district-level administrative divisions. By using blocks as the analysis unit, we can obtain more samples to reveal the spatial heterogeneity of the epidemic and the spatial differences of its influencing factors from a more refined perspective. The average of the kernel density estimation (KDE, see Methods for detailed information) of self-reported COVID-19 cases in each block is the dependent variable. Population density and aging rate (with their initial values) at the block level are two explanatory variables. In terms of transportation variables (metro stations and main roads) and commercial POIs, we also use the KDE of them in each block as the other three explanatory variables.

### 3. Methods

#### 3.1. Kernel density estimation

Kernel Density Estimation (KDE) takes the sample as the center (core) and calculates the density per unit area of the sample point within the search radius (bandwidth) to indicate the spatial distribution of densities of geographic elements. KDE has a wide range of applications in disease mapping. The mathematical expression of kernel density estimation is as Equation (1):

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \tag{1}$$

where  $K()$  is the kernel function;  $x-x_i$  represents the distance from the value point to the output grid,  $h$  is the bandwidth, that is, the radius of the circle, and  $n$  is the sample size.

This study uses the kernel density estimation function in ArcGIS 10.3 software to analyze the spatial intensity of self-reported COVID-19 cases and influencing factors at the block level in Wuhan. The adaptive bandwidth is used as the search radius.

#### 3.2. Global linear regression

We first build the ordinary least square (OLS) linear regression model (Equation (2)) to quantify influencing factors of COVID-19 cases.

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \mu \tag{2}$$

where  $y$  is the KDE of self-reported COVID-19 cases at blocks,  $x_i$  are five explanatory variables,  $k$  is the number of explanatory variables,  $\beta_0$  is a constant,  $\beta_i$  is the regression coefficient of  $x_i$ , and  $\mu$  is the error term.

We use percent changes (PC) of self-reported COVID-19 cases caused by the increase of one standard deviation in the explanatory variable to compare contributions from different explanatory variables, which is defined as Equation (3) (Xu et al., 2020):

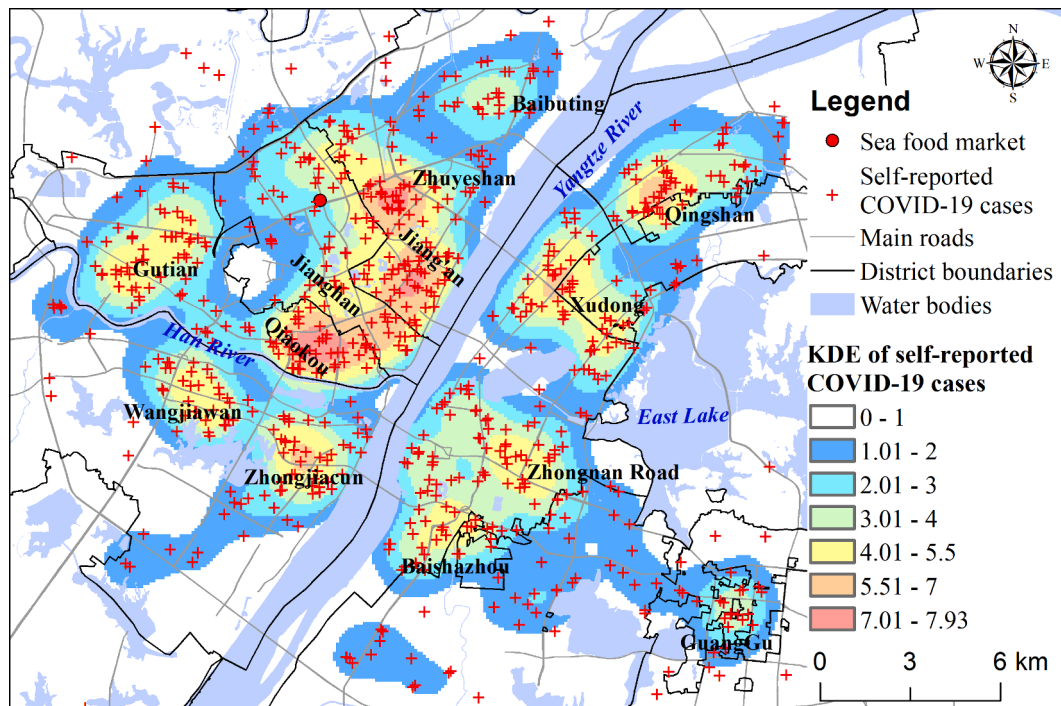


Fig. 3. Spatial distributions of self-reported COVID-19 cases and the result of kernel density estimation (KDE) in the main urban area of Wuhan, China.

$$PC_i = \frac{\beta_i s_{d_i}}{\bar{y}} \times 100\% \tag{3}$$

where  $PC_i$  is the percent change of  $y$  caused by the increase of one standard deviation of  $x_i$ ,  $\beta_i$  is the regression coefficient of  $x_i$ ,  $s_{d_i}$  is the standard deviation of  $x_i$ , and  $\bar{y}$  is the average of self-reported COVID-19 cases at all blocks.

### 3.3. Geographically weighted regression

Spatial heterogeneity is a fundamental characteristic of geographical variables, leading to a spatial variance in their relationships (Fotheringham et al., 2015). The global ordinary least squares (OLS) regression model assumes the spatial stationarity in their relationships between explanatory variables and the dependent variable, failing in capturing the variance in space of relationships among geographical variables. A spatially varying coefficient modeling strategy is needed in geographical analysis (Murakami et al., 2019). The geographically weighted regression (GWR) model was proposed and widely used in many disciplines, whose generic formulation is shown by Equation (4) (Fotheringham et al., 2002):

$$y_i = \beta_0 + \sum_{j=1}^m \beta_{ij} X_{ij} + \varepsilon_i, \quad i = 1, 2, \dots, n \tag{4}$$

where at block  $i$ ,  $y_i$  is the averaged KDE of self-reported COVID-19 cases,  $\beta_{i0}$  is the intercept,  $\beta_{ij}$  is the  $j^{\text{th}}$  regression parameter,  $X_{ij}$  is the value of the  $j^{\text{th}}$  explanatory parameter, and  $\varepsilon_i$  is a random error term.

In the GWR model, closer observations have a higher influence in estimating the local set of coefficients than distant observations. Regression parameters in the GWR model at each block in matrix form are as Equation (5) (Fotheringham et al., 2002):

$$\hat{\beta}(i) = (X^T W(i) X)^{-1} X^T W(i) y \tag{5}$$

where  $\hat{\beta}(i)$  is the vector of parameter estimates for block  $i$ ,  $W(i)$  is the diagonal weights matrix specified for block  $i$ ,  $X$  is the matrix of the explanatory variable with a first column of  $1$ s for the intercept,  $y$  is the

vector of the dependent variable.

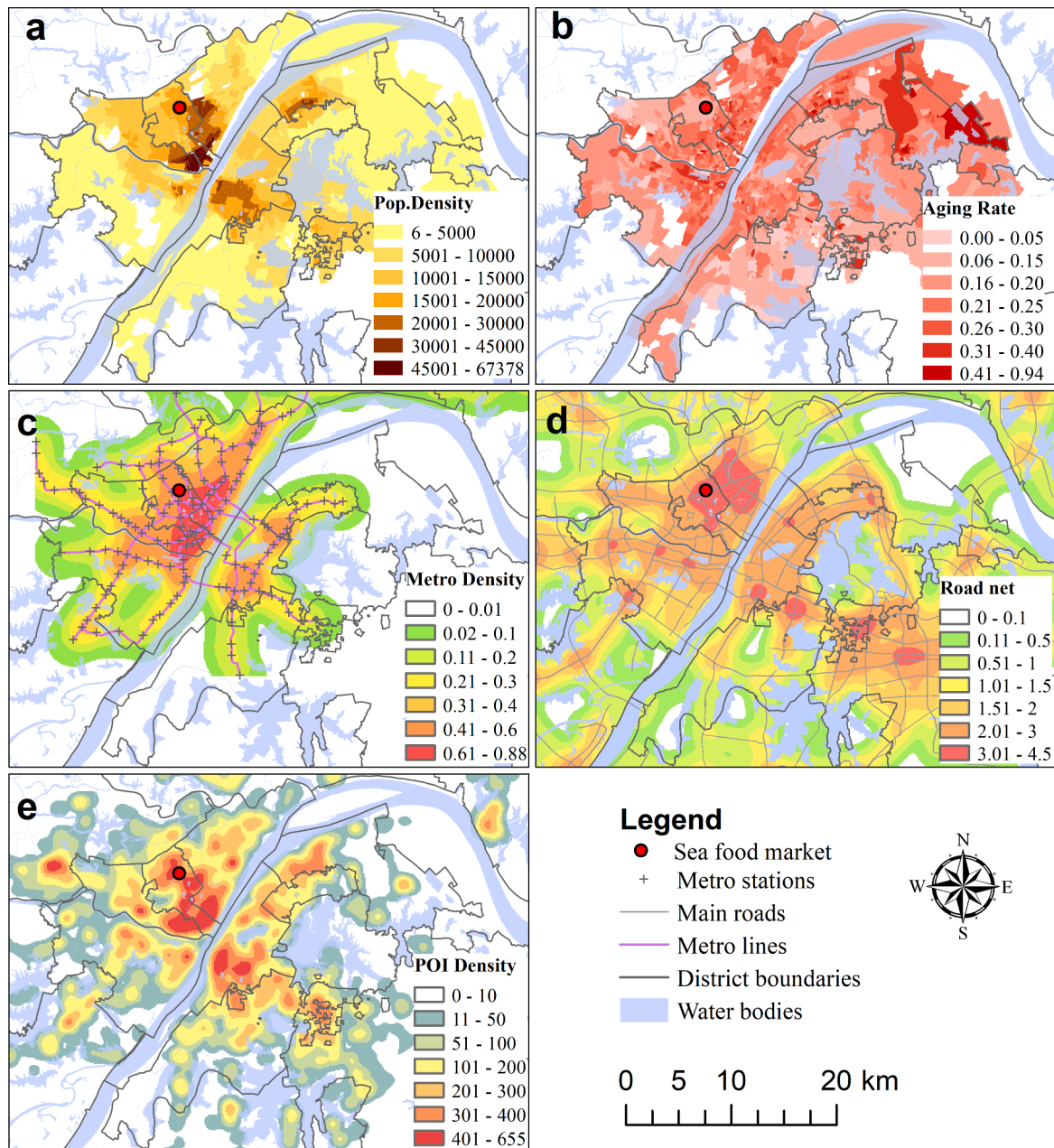
The weights matrix ( $W(i)$ ) is calculated with a specified kernel function and a bandwidth. The Gaussian function is a widely used kernel function. The bandwidth is specified either by a fixed distance or a fixed number of nearest neighbors, namely, a fixed bandwidth or an adaptive bandwidth, respectively (Lu B, Yang W, Ge Y, & Harris P, 2018). More detailed information on kernel function and bandwidth in GWR modeling can be found in references (Fotheringham and Oshan, 2016; Lu et al., 2017; Lu et al., 2014a; Wheeler and Tiefelsdorf, 2005). In this study, the GWR modeling was conducted using the package of “*GWmodel*” in R programming (Gollini I, Lu B, Charlton M, Brunson D, & Harris P, 2015; Lu, Harris, Charlton, & Brunson, 2014b).

## 4. Results

### 4.1. Spatio-temporal variations of self-reported COVID-19 cases

The temporal variations of self-reported COVID-19 cases are shown in Fig. 2. The self-reported COVID-19 cases started to post their symptoms for help from February 3, 2020, and numbers of self-reported cases quickly increased in the following days (Fig. 2a). On February 5, there were more than 180 self-reported cases who posted their symptoms on social media for help. Wuhan had urgently constructed two infectious hospitals and also opened up mobile cabin hospitals for the isolation and treatment of patients with mild symptoms. As the treatment capacity increased, self-reported cases were quickly admitted to hospitals. As a result, the number of self-reported COVID-19 cases also dropped significantly and there were fewer than 10 self-reported cases after February 15. We calculated the interval time from onset to self-report for each case and the histogram of the interval time is shown in Fig. 2b. The average interval time is 9.87 days and this time for most cases varies in 0-20 days. According to the interval time, it can be inferred that most self-reported COVID-19 cases were ill around January 25, 2020, which is close to the date (January 23, 2020) of the lockdown of Wuhan.

We assess the biases of the self-reported data by comparing it with official data. We calculate proportions of self-reported COVID-19 cases in each district in Wuhan, and compare proportions from the official



**Fig. 4.** Spatial distributions of explanatory variables and results of kernel density estimation (KDE). (a) Population density; (b) Aging rate; (c) KDE of metro stations; (d) KDE of main roads; (e) KDE of commercial POIs.

published data (Fig. 2c, 2d). Overall, all districts are scattered around the 1:1 line, showing a general consistence of proportions between self-reported COVID-19 cases and official data. The proportion of self-reported COVID-19 cases in highly infected districts (like Wuchang, Qiaokou, and Jiang'an districts) is higher than the official published proportion (Fig. 2c). By calculating the chi-square value between self-reported cases and official statistics (as of May 18, 2020), the result shows that the chi-square value is 122.55 ( $P < 0.001$ ), which shows that proportions among districts between the self-reported cases and official data are consistent. The average age of the self-reported COVID-19 cases is 59.8 years old. We also compare the age distributions of self-reported COVID-19 cases with official published data (Team, 2020) (Fig. 2d). The overall distribution trends of them are similar, with the peak group being 50-70 years old. However, the age distribution of self-reported cases is more biased towards the elderly.

The spatial distribution of more than nine hundred self-reported

COVID-19 cases is presented in Fig. 3. The result of kernel density estimation (KDE) shows the intensity of self-reported cases. Overall, most self-reported cases were concentrated in the main urban area of Hankou along the Yangtze River and the Han River. Hankou is the original birthplace and also the city center of Wuhan that has the highest population density and busiest commercial activities across the whole city. Apart from the concentrated and continuous cluster of self-reported COVID-19 cases in the city center, there are other three clusters of higher intensity of self-reported cases in Hankou, which are Zhuyeshan, Baibuting, and Gutian (Fig. 3). These three areas are residential areas with high population density, resulting in a higher attack rate of COVID-19 there. In the south bank of the Yangtze River (Wuchang), there are five clusters of higher intensity of self-reported cases, among which Zhongnan Road and Xudong are sub-centers and mixed commercial and residential areas with high population densities and large visitor flows. Another three clusters (Qingshan, Baishazhou, and Guanggu) are

**Table 1**  
Descriptive statistics of self-reported COVID-19 cases and explanatory variables.

Variables	Min.	Median	Mean	Max.
Self-reported cases	0	2.923	3.007	7.900
Pop. Density <sup>#</sup>	6	12,120	15,600	67,378
Aging rate	0	0.2091	0.2034	0.9424
Metro stations	0	0.3381	0.3440	0.8708
Main roads	0	2.318	2.202	4.456
Commercial POIs	0	245.0	248.0	647.3

<sup>#</sup> Person/km<sup>2</sup>.

residential areas with concentrated population there. In Hanyang region, the two clusters of high intensity of self-reported cases are Zhongjiacun and Wangjiawan, which are two local commercial and residential centers.

#### 4.2. Correlation analysis

The spatial distribution of explanatory variables and results of kernel density estimation (KDE) are shown in Fig. 4. Blocks with the highest population density are concentrated in the core area of Hankou (central business district). The core areas of Wuchang District and Qingshan District also experience higher population densities (Fig. 4a). The distribution of the aging rate is relatively discrete. The northeast part of the central urban area has the highest aging rate, which is the interface between urban and rural areas (Fig. 4b). There are eight metro lines in Wuhan (Fig. 4c) and areas with the highest density of metro stations are concentrated on the left bank of the Yangtze River, surrounding major business districts. The intersections of other metro stations are also hot spots. Two banks of the upper reaches of the Yangtze River are areas with the highest road density. High-density areas of commercial POIs are distributed along both banks of the Yangtze River, with business districts as the centers (Fig. 4e). There are many small clusters of the KDE of commercial POIs, suggesting that POIs reflect social activities in a more detailed resolution.

Descriptive statistics of self-reported COVID-19 cases and explanatory variables at blocks are shown in Table 1. Self-reported COVID-19

cases, metro stations, main roads, and POIs are aggregated from the KDE results of them. Correlation matrix between self-reported COVID-19 cases and explanatory variables is shown in Fig. 5. The Pearson's  $r$  between five explanatory variables and self-reported cases at the block level varies in 0.32-0.82, and they are all significantly and positively correlated ( $P < 0.001$ ). Among them, the Pearson's  $r$  between commercial POIs and self-reported cases is the strongest, reaching 0.82. The scatter plot between aging rate and self-reported cases is relatively dispersed, resulting in the lowest correlation.

#### 4.3. Regression modeling using OLS and GWR

We take the average KDE of self-reported COVID-19 cases in a block as the dependent variable and build the global ordinary least squares (OLS) regression model with five explanatory variables (Table 2). The adjusted  $R^2$  of the OLS model is 0.76, indicating that five explanatory variables can explain 76% variance of self-reported COVID-19 cases at the block level. In general, all explanatory variables are significantly ( $P < 0.001$ ) and positively correlated with the dependent variable.

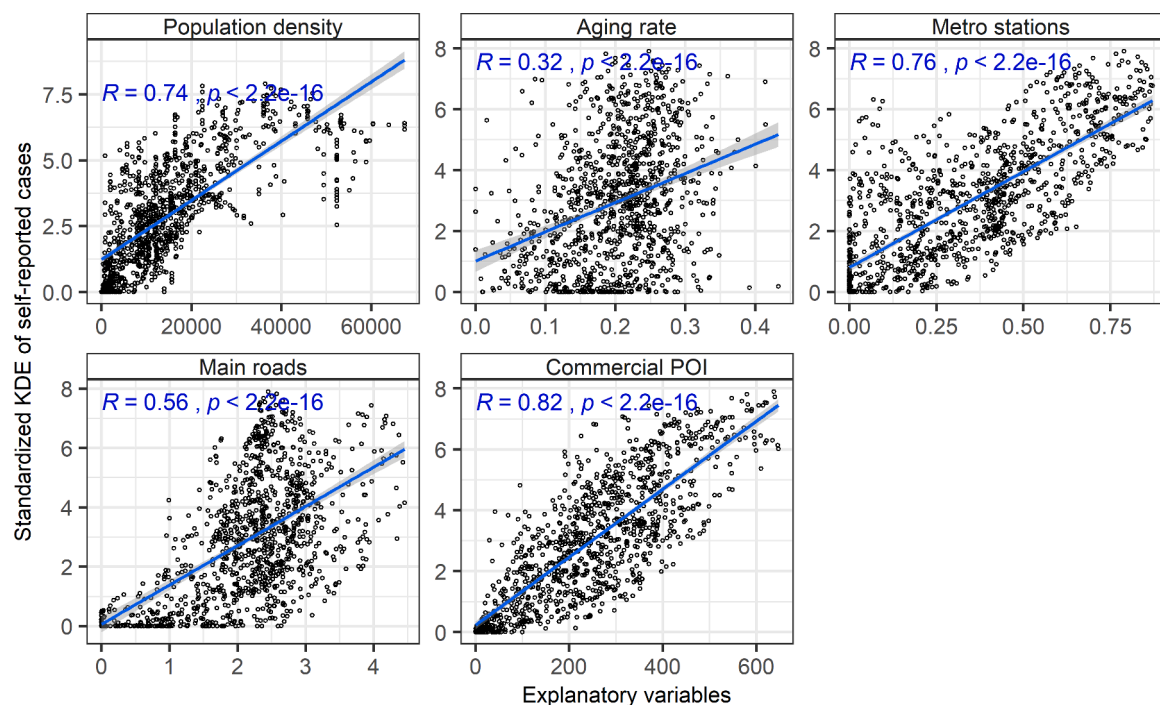
We calculate percent changes of self-reported COVID-19 cases when the explanatory variable increases by one standard deviation, holding all other explanatory variables constant at their arithmetic mean values

**Table 2**

The result of OLS regression with the KDE of self-reported COVID-19 cases as the dependent variable ( $N = 1071$ ).  $\beta$  is the coefficient of regression model. Standard error represents the standard deviation of the regression coefficient. T-value and p-value are the significance test results of coefficients.

Explanatory variables	$\beta$	Standard error	$t$	$p$	VIF
Intercept	-0.68	0.12	-5.71	< 0.001	
Population density <sup>#</sup>	0.031	0.004	8.78	< 0.001	2.44
Aging rate	2.28	0.41	5.57	< 0.001	1.05
Metro stations	2.02	0.20	9.98	< 0.001	2.73
Main roads	0.28	0.045	6.20	< 0.001	1.63
Commercial POIs	0.57	0.033	17.47	< 0.001	2.55

<sup>#</sup> Thousand people/km<sup>2</sup>.



**Fig. 5.** Correlation analysis between self-reported COVID-19 cases and explanatory variables. The blue straight line is the regression line with Pearson's  $R$  and  $p$  value in each plot.

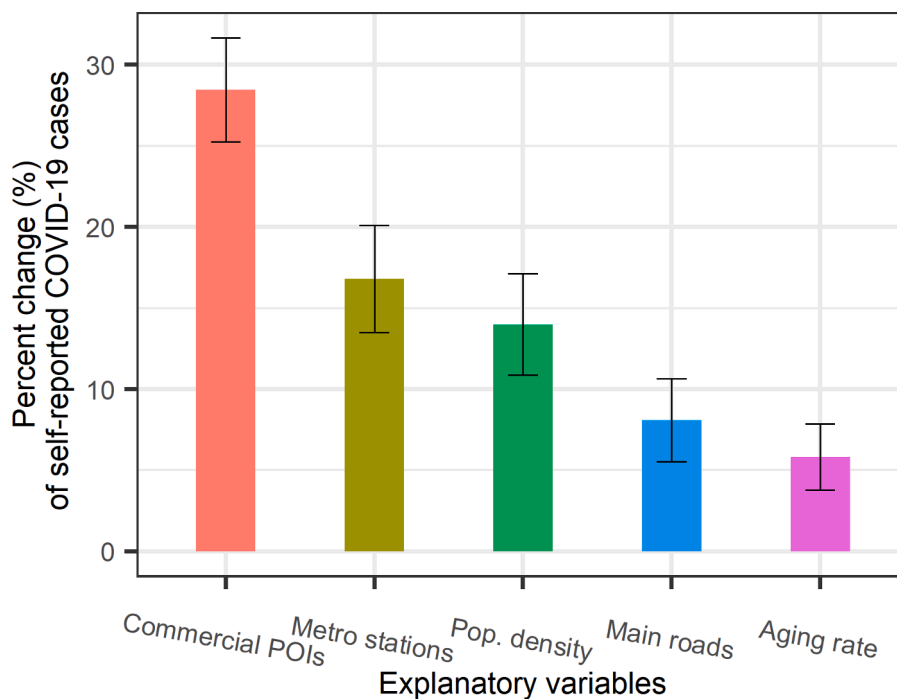


Fig. 6. Percent changes of self-reported COVID-19 cases associated with every one standard deviation increase from mean value of each explanatory variable. All other explanatory variables are held at their mean arithmetic values. Error bars are 95% confidence intervals.

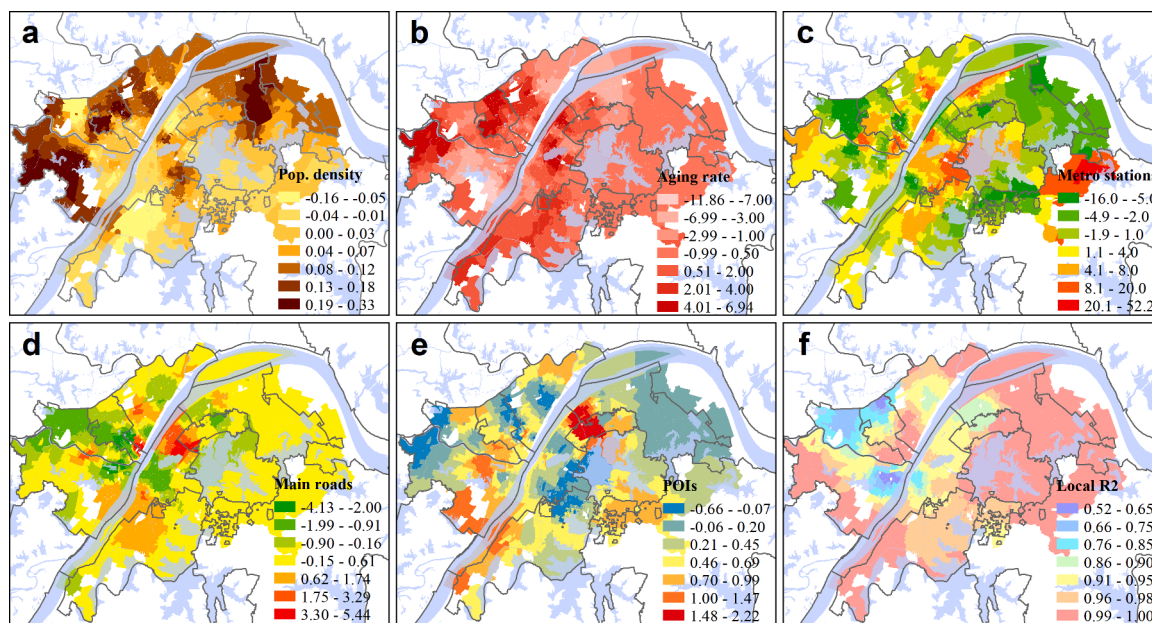


Fig. 7. Spatial distributions of estimated coefficients and local adjusted  $R^2$  in the GWR model. (a) population density; (b) aging rate; (c) Metro station; (d) Main roads; (e) Commercial POI; (f) Local adjusted  $R^2$ .

(Equation (3)), shown in Fig. 6. Commercial POIs show the strongest influences on self-reported COVID-19 cases, which increase COVID-19 cases by 28% with the increase of one standard deviation. In terms of transportation, metro stations have stronger influences on COVID-19 cases than main roads. It is easy to understand that public transportation has a higher risk for infectious diseases. As for population, the increase of one standard deviation in population density and aging rate increase COVID-19 cases by 14% and 6%, respectively.

We further investigate influences of explanatory variables on self-reported COVID-19 cases using the geographically weighted regression

(GWR) method. The spatial distribution of coefficients and local  $R^2$  in the GWR model are presented in Fig. 7. Generally, the adjusted  $R^2$  of the GWR model is 0.96, which is significantly higher than that of OLS model (0.76). It's clear that the GWR model can explain the dependent variable more effectively, indicating that GWR has a high explanatory power and better fit ability in most blocks. In this model, the corrected Akaike information criterion (AICc) value of the GWR model (1305.6) is reduced by 57% compared with the global OLS model ( $AICc = 3052.5$ ). Obviously, the performance of the GWR model is significantly improved.

Coefficients of five explanatory variables vary across blocks, showing



obvious spatial heterogeneities of their influences on self-reported COVID-19 cases. For population density, blocks in the north of the central urban area of Wuhan have higher coefficients, showing stronger influences of population density on self-reported COVID-19 cases in this region (Fig. 7a). For aging rate, blocks with higher coefficients are located in Hankou and the south of the central urban area (Fig. 7b). For transportation factors, blocks with higher coefficients of metro stations are mainly located in the central urban areas (Fig. 7c), while blocks with higher coefficients of main roads are mainly located in faraway suburban areas (Fig. 7d). The highest coefficients of commercial POIs are located in the Qingshan District, showing a stronger influence of social interactions there (Fig. 7e). Most areas around the central urban area of Wuhan have higher local  $R^2$  in the GWR model (Fig. 7f). Two clusters of lower local  $R^2$  are located in the core urban area of Hanyang and northwest part of Hankou, which reveals that other influencing factors account for local intensity of self-reported COVID-19 cases there.

## 5. Discussion

Numerous studies investigating the COVID-19 epidemic are in a large scale like country, province, and state levels, while detailed exploration at smaller spatial scales is limited. This study contributes to reveal spatial disparities of COVID-19 cases and influencing factors at a fine spatial resolution in Wuhan, China. We collected more than nine hundred self-reported COVID-19 cases in Wuhan through a large Chinese social media platform (*Weibo*, like *Twitter*), compensating for the vacancy of detailed confirmed COVID-19 cases at the intra-urban scale. The proportions of self-reported cases at the district level are consistent with official published data, and they also share a very similar distribution of ages with official published data, suggesting the representativeness of self-reported cases to quantifying spatial heterogeneities of the COVID-19 epidemic in Wuhan. Such social media data and other big data have great potential applications in response to disasters and public emergencies (Zhou et al., 2020). Nevertheless, self-reported COVID-19 cases are not final cases after all, and there may be a bias in the spatial distribution of final cases.

Overall, there are obvious spatial clusters of self-reported COVID-19 cases, showing obvious spatial heterogeneities of the COVID-19 epidemic, which was demonstrated by previous related studies (Mena et al., 2021; Yang et al., 2021). Areas with higher morbidity rates are mainly concentrated in commercial centers and populous residential areas, where there are higher population densities and a higher frequency of social interactions. The OLS model shows that population dynamics, transportation, and social interaction account for 76% variance in self-reported COVID-19 cases. The GWR model has a better performance (adjusted  $R^2 = 0.96$ ) than the OLS model and reveals spatial disparities of influences of explanatory variables on self-reported COVID-19 cases.

The COVID-19 epidemic asked us to rethink the city, to reflect on urban planning, construction and governance: including city size, urban density, and community design (Batty, 2020; Moosa and Khatatbeh, 2021; Sharifi and Khavarian-Garmsir, 2020). The core of a city is the agglomeration of people, which is reflected in two aspects, size and density (Li et al., 2021). Size and density support the economic output and knowledge innovation of cities, but also provide hotbeds for the breeding and spread of infectious diseases (Bettencourt et al., 2007; Lei et al., 2021). It is especially necessary to incorporate healthy city in urban planning. Firstly, in high-density cities, the focus is on improving the accessibility of public facilities and services, and increasing urban green space and open space (Liu et al., 2021). Secondly, we should improve urban resilience and the ability of cities to deal with emergencies through the improvement of self-sufficiency in resource and reducing in ecological footprint (Yang et al., 2021). Finally, local realities should be taken into consideration in community design.

This study also has limitations. Many other factors may have influences on the spread of COVID-19 epidemic in cities, such as

socioeconomic status, occupation, urban structure, housing quality, etc. (Hu et al., 2021; Mansour et al., 2021; Megahed and Ghoneim, 2020). In addition, those demographical and urban characteristics have complex influences on the infectious diseases. For example, the enclosed community in China is usually criticized for isolation of human mobility and increasing in traffic congestion. However, it is the enclosed community that allows the stay-at-home order to be implemented well in Wuhan and other Chinese cities (Huang et al., 2021). From this point of view, it is very meaningful to analyze the COVID-19 infection rate in enclosed and open communities in Wuhan.

## 6. Conclusions

Wuhan was the initial epicenter of the COVID-19 epidemic. This study used self-reported COVID-19 cases in Wuhan from a Chinese social media platform (*Weibo*) to quantify spatial intensity of COVID-19 epidemic within the city. The self-reported cases share consistent properties with official published data at the macro level but with detailed locations. Self-reported cases were mainly concentrated in commercial centers and populous residential areas, verifying spatial clusters of COVID-19 epidemic there. Population dynamics, transportation, and social interactions strongly determine the spatial disparities of the COVID-19 epidemic at the block level. The GWR model characterizes local variations of influences of five explanatory variables on the COVID-19 epidemic and performs better than the OLS model. Our findings enlighten us to optimize urban design, transform urban infrastructure, and create a healthy city.

## Declaration of Competing Interest

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.

## Acknowledgements

This research was supported by the National Key Research and Development Program of China (2017YFB0503600), the National Natural Science Foundation of China (42101460, 42171448, 42071368) and the Postdoctoral Program of Hubei Province, China (2021-2022).

## References

- Acuto, M., Larcom, S., Keil, R., Ghojeh, M., Lindsay, T., Camponeschi, C., et al. (2020). Seeing COVID-19 through an urban lens. *Nature Sustainability*, 3, 977–978.
- Batty, M. (2020). The Coronavirus crisis: What will the post-pandemic city look like? *Environment and Planning B: Urban Analytics and City Science*, 47, 547–552.
- Bettencourt, L. M., Lobo, J., Helbing, D., Kuhnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104, 7301–7306.
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., et al. (2021). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, 589, 82–87.
- Cordes, J., & Castro, M. C. (2020). Spatial analysis of COVID-19 clusters and contextual factors in New York City. *Spatial and Spatio-temporal Epidemiology*, 34, Article 100355.
- Das, A., Ghosh, S., Das, K., Basu, T., Dutta, I., & Das, M. (2021). Living environment matters: Unravelling the spatial clustering of COVID-19 hotspots in Kolkata megacity, India. *Sustainable Cities and Society*, 65, Article 102577.
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20, 533–534.
- Drefahl, S., Wallace, M., Mussino, E., Aradhya, S., Kolk, M., Brandén, M., et al. (2020). A population-based cohort study of socio-demographic risk factors for COVID-19 deaths in Sweden. *Nature Communications*, 11, 1–7.
- Fotheringham, A. S., Brunson, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. John Wiley & Sons.
- Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and Temporal Weighted Regression (GTWR). *Geographical Analysis*, 47, 431–452.
- Fotheringham, A. S., & Oshan, T. M. (2016). Geographically weighted regression and multicollinearity: Dispelling the myth. *Journal of Geographical Systems*, 18, 303–329.
- Frumkin, H. (2021). COVID-19, the built environment, and health. *Environmental Health Perspectives*, 129, Article 075001.

- Gollini I, Lu B, Charlton M, Brunsdon C, & Harris P. (2015). GWmodel: an R Package for Exploring Spatial Heterogeneity using Geographically Weighted Models. *Journal of Statistical Software*, 63(17), 1–50. <https://doi.org/10.18637/jss.v063.i17>
- Hu, M., Roberts, J. D., Azevedo, G. P., & Milner, D. (2021). The role of built and social environmental factors in Covid-19 transmission: A look at America's capital city. *Sustainable Cities and Society*, 65, Article 102580.
- Huang, X., Yang, Q., & Yang, J. (2021). Importance of community containment measures in combating the COVID-19 epidemic: From the perspective of urban planning. *Geospatial Information Science*, 1–9.
- Kashem, S. B., Baker, D. M., González, S. R., & Lee, C. A. (2021). Exploring the nexus between social vulnerability, built environment, and the prevalence of COVID-19: A case study of Chicago. *Sustainable Cities and Society*, 75, Article 103261.
- Lei, W., Jiao, L., Xu, G., & Zhou, Z. (2021). *Urban scaling in rapidly urbanising china*. Urban studies, 00420980211017817.
- Li, S., Ma, S., & Zhang, J. (2021). Association of built environment attributes with the spread of COVID-19 at its initial stage in China. *Sustainable Cities and Society*, Article 102752.
- Li, X., Zhou, L., Jia, T., Peng, R., Fu, X., & Zou, Y. (2020). Associating COVID-19 Severity with Urban Factors: A Case Study of Wuhan. *International Journal of Environmental Research and Public Health*, 17, 6712.
- Liu, C., Liu, Z., & Guan, C. (2021). The impacts of the built environment on the incidence rate of COVID-19: A case study of King County, Washington. *Sustainable Cities and Society*, 74, Article 103144.
- Lu B, Yang W, Ge Y, & Harris P. (2018). Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. *Computers, Environment and Urban Systems*, 71, 41–57. <https://doi.org/10.1016/j.compenurbysys.2018.03.012>
- Lu, B., Brunsdon, C., Charlton, M., & Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, 31, 982–998.
- Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014a). Geographically weighted regression with a non-Euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*, 28, 660–681.
- Lu, B., Harris, P., Charlton, M., & Brunsdon, C. (2014b). The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, 17, 85–101.
- Maiti, A., Zhang, Q., Sannigrahi, S., Pramanik, S., Chakraborti, S., Cerda, A., et al. (2021). Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States. *Sustainable Cities and Society*, 68, Article 102784.
- Mansour, S., Al Kindi, A., Al-Said, A., Al-Said, A., & Atkinson, P. (2021). Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). *Sustainable Cities and Society*, 65, Article 102627.
- Maroko, A. R., Nash, D., & Pavilonis, B. T. (2020). COVID-19 and Inequity: A Comparative Spatial Analysis of New York City and Chicago Hot Spots. *Journal of Urban Health*, 97, 461–470.
- Megahed, N. A., & Ghoneim, E. M. (2020). Antivirus-built environment: Lessons learned from Covid-19 pandemic. *Sustainable Cities and Society*, 61.
- Mena, G. E., Martinez, P. P., Mahmud, A. S., Marquet, P. A., Buckee, C. O., & Santillana, M. (2021). Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile. *Science (New York, N.Y.)*, 372.
- 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, Article 138884.
- Moosa, I. A., & Khatatbeh, I. N. (2021). The density paradox: Are densely-populated regions more vulnerable to Covid-19? *The International Journal of Health Planning and Management*.
- Murakami, D., Lu, B. B., Harris, P., Brunsdon, C., Charlton, M., Nakaya, T., et al. (2019). The Importance of Scale in Spatially Varying Coefficient Modeling. *Annals of the American Association of Geographers*, 109, 50–70.
- Peng, Z., Wang, R., Liu, L., & Wu, H. (2020). Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data. *ISPRS International Journal of Geo-Information*, 9, 402.
- Real, L. A., & Biek, R. (2007). Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *Journal of the Royal Society Interface*, 4, 935–948.
- Sharifi, A., & Khavarian-Garmsir, A. R. (2020). The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Science of the Total Environment*, Article 142391.
- 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*.
- Team, T. N. C. P. E. R. E. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Weekly*, 2, 113–122.
- Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7, 161–187.
- Xu, G., Wang, W., Lu, D., Lu, B., Qin, K., & Jiao, L. (2021a). Geographically varying relationships between population flows from Wuhan and COVID-19 cases in Chinese cities. *Geo-spatial Information Science*, 1–11.
- Xu, G., Xiu, T., Li, X., Liang, X., & Jiao, L. (2021b). Lockdown Induced Night-Time Light Dynamics during the COVID-19 Epidemic in Global Megacities. *International Journal of Applied Earth Observation and Geoinformation*, 102, 1–10.
- Xu, G., Zhou, Z., Jiao, L., & Zhao, R. (2020). Compact Urban Form and Expansion Pattern Slow Down the Decline in Urban Densities: A Global Perspective. *Land Use Policy*, 94, Article 104563.
- Yang, Y., Li, Y., Kral, K., Hupert, N., & Dogan, T. (2021). Urban design attributes and resilience: COVID-19 evidence from New York City. *Buildings and Cities*, 2.
- Zhang, Z., Xue, T., & Jin, X. (2020). Effects of meteorological conditions and air pollution on COVID-19 transmission: Evidence from 219 Chinese cities. *Science of the Total Environment*, 741, Article 140244.
- Zhou, C., Su, F., Pei, T., Zhang, A., Du, Y., Luo, B., et al. (2020). COVID-19: Challenges to GIS with Big Data. *Geography and Sustainability*, 1, 77–87.
- United Nations. (2020). Policy Brief: COVID-19 in an Urban World.