

Article

# How Urban Factors Affect the Spatiotemporal Distribution of Infectious Diseases in Addition to Intercity Population Movement in China

Xinyi Niu <sup>1,2</sup>, Yufeng Yue <sup>1</sup>, Xingang Zhou <sup>1,2,3,\*</sup> and Xiaohu Zhang <sup>4</sup>

<sup>1</sup> College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China; niuxinyi@tongji.edu.cn (X.N.); yufeng\_yue@tongji.edu.cn (Y.Y.)

<sup>2</sup> Key Laboratory of Ecology and Energy-Saving Study of Dense Habitat, Ministry of Education, Tongji University, Shanghai 200092, China

<sup>3</sup> Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China;

<sup>4</sup> Massachusetts Institute of Technology, 02142, USA; zhangxh@mit.edu (X.Z.)

\* Correspondence: zxcg@tongji.edu.cn

Received: 11 September 2020; Accepted: 19 October 2020; Published: 22 October 2020

**Abstract:** The outbreak of the 2019 novel coronavirus (COVID-19) has attracted global attention. During the Chinese New Year holiday, population outflow from Wuhan induced the spread of the epidemic to other cities in China. This study analyzed massive intercity movement data from Baidu and epidemic data to study how intercity population outflows affected the spatiotemporal spread of the epidemic. This study further investigated how urban factors influenced the spatiotemporal spread of COVID-19. The analysis indicates that intercity movement was an important factor in the spread of the epidemic in China, and the impact of intercity movement on the spread was heterogeneous across different classes of cities. The spread of the epidemic also varied among cities and was affected by urban factors including the total population, population density, and gross domestic product (GDP). The findings have implications for public health management. Mega-cities should consider tougher measures to contain the spread of the epidemic compared with other cities. It is of great significance for policymakers in any nation to assess the potential risk of epidemics and make cautious plans ahead of time.

**Keywords:** population movement; spatiotemporal distribution; infectious diseases

---

## 1. Introduction

Estimation of the spread and evolution of infectious disease is essential for public health during a pandemic [1]. Human travel is a major factor in the spread of epidemics from within one city to others. The incidence of infectious diseases such as influenza varies among cities and is affected by urban factors including population size, socioeconomic conditions, urban structure, and connectivity with other cities [2–4]. However, there are few studies on how intercity movements affect the spatiotemporal spread of infectious diseases together with urban factors.

The advent of the 2019 novel coronavirus (COVID-19) coincided with the Lunar New Year holiday, during which significant human migration occurred as people returned to their hometowns [5]. The population outflow spatially expanded the epidemic to other cities in China. When the travel restrictions in Wuhan came into effect on January 23, 2020, many Chinese cities had already accepted a number of infected travelers [6]. Around 5 million people traveled from Wuhan to other cities, and around one-third of them went to other cities outside of Hubei province within two weeks before the lockdown. Li and Pei et al. [7] estimated that 86% of all infections were undocumented before the travel restrictions were imposed in Wuhan. Human movement significantly affected the distribution and scale of the epidemics in China and can be used to predict the future spread of epidemics [8,9].

It is hypothesized that the population movement from Wuhan to other cities within the two weeks before the lockdown may have had a significant effect on the potential COVID-19 outbreaks in other Chinese cities. Therefore, it is necessary to understand how human movement affects the spatiotemporal distribution of COVID-19 in China.

Some studies have analyzed the correlation between intercity movements and infectious diseases [10], while few studies have examined how urban factors affect the spatiotemporal spread of infectious diseases. The spatial spread of epidemics is affected not only by the intercity movement of infected people, but also by the urban factors of different cities. This study further examines how urban factors affected the spatiotemporal distribution of COVID-19 in addition to intercity population movement in Chinese cities.

## 2. Literature Review

### 2.1. Human Mobility and Infectious Diseases

Many cities are likely to face more serious infectious diseases than ever before due to the increasing levels of human movement as globalization continues. Huge intercity human movement accelerates the spread of infectious diseases, and previous studies analyzing infectious diseases have varied from the global to intra-urban level in scale [11]. For example, Wells and Sah [12] examined the influence of cross-national travel and border-control policies on the spread of COVID-19 at a global level. Zhou and Xu [13] used mobile phone sightings data to examine the potential effects of intra-city mobility restrictions on the spread of COVID-19 in the city of Shenzhen. Massaro and Kondor [14] studied the interaction between human mobility and outbreaks of infectious diseases and found that intra-city human mobility is a major factor in the spread of infectious diseases in Singapore. Some studies have analyzed the spatiotemporal pattern of COVID-19 at the national or district levels [1,15,16]. However, few studies have examined the relationship between human movement and the spatiotemporal distribution of infectious diseases at the national scale because it is difficult to obtain detailed intercity travel records [17]. For example, Jia and Lu [18] investigated how population movement drove the spatial distribution of COVID-19 in Chinese cities using mobile phone location data at the national level.

### 2.2. Urban Factors in the Spread of Infectious Diseases

Human movement is an important factor leading to the spread of infectious diseases. However, studies examining how urban factors affect the spatiotemporal distribution of infectious diseases are rare. City size may affect the spatial spread of infectious diseases. Commuting distances increase as residents become farther away from city centers, so average commuting distances become longer as the size of a city becomes larger without considering spatial structure [19]. Residents may have more interactions during commuting and might be more prone to infection with longer commuting distances. Total population and built-up area are common indicators of city size. Dalziel and Pourbohloul [17] built an individual-level model of infectious disease transmission based on mobility data, confirming that the differences in population size and mobility patterns among cities are sufficient to affect the risk and severity of epidemics in different cities. Larger cities with larger populations and more social interactions are more vulnerable to infectious diseases [20].

High population density increases people's daily social interaction and close contact, which makes them vulnerable to infectious diseases and accelerates the spread of an epidemic [21]. China's urbanization has progressed rapidly in recent decades, characterized by high density. It is necessary to examine how density affects the spread of an epidemic, especially in Chinese cities.

In cities with higher GDP, people tend to have more face-to-face communications and are thus more vulnerable to infectious diseases. Garske T. [22] suggested that infectious diseases might spread faster in more developed areas than less developed areas in China. Economic activities such as wholesaling, retailing, and services usually require frequent and close interactions between people.

### 2.3. Examining Intercity Travel with Spatiotemporal Positioning Big Data

The scale of China's urbanization has been unprecedented in terms of the magnitude and rate of change, and the mass migration from rural to urban areas and from small to large cities [23, 24]. It is necessary to examine how the population movements affect the urban structure and population distribution in Chinese cities, especially in epidemics. Some studies use census data to examine intercity travel. For example, Mu and Yeh [25] used demographic survey data to divide China into new commuting areas based on the functional network of commuting flows. However, these kinds of census data are labor-intensive to collect and limited by the sampling size.

With the development and widespread applications of big spatiotemporal data [26–28], mobility data have been used to trace spatiotemporal intercity movements and disease transmission for their capability to track the travel behavior of inhabitants in fine granularity [29, 30]. Millions of location-based service records can not only map the spatiotemporal distribution of people and intercity travel [31, 32] but also help to elucidate the importation and transmission of diseases together with disease distribution data [33]. With the popularity of different sensors, it is possible to examine how intercity population movement affects the spatiotemporal distribution of infectious diseases [34].

## 3. Study Area and Data Processing

### 3.1. Study Area

This study focuses on Chinese prefectures, initially including 360 administrative units. These prefectures are referred to as “cities” in this study (Figure 1). All cities in Hubei Province were excluded, since intra-city transmission had already dominated the local spread of infectious diseases during the study period. The community-level spread of COVID-19 occurred much earlier in Hubei Province than in other parts of China, as Wuhan city has a close daily connection with those cities in Hubei.

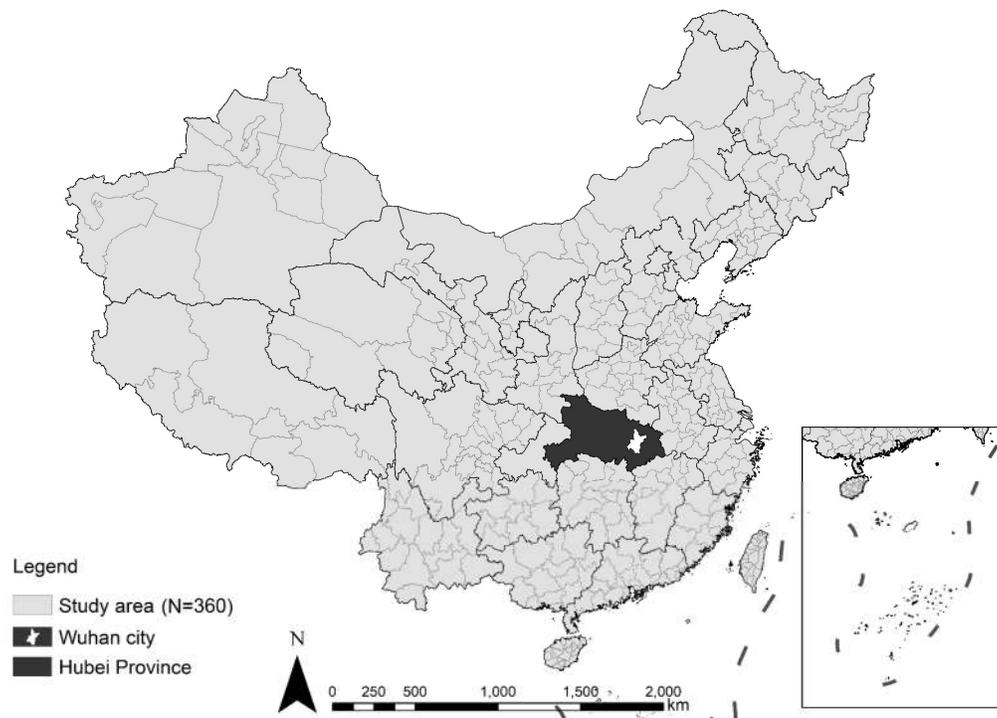


Figure 1. Study area.

### 3.2. COVID-19 Data

COVID-19 data were extracted from the epidemic report of the National Health Commission of China. The daily numbers of confirmed cases were summarized up to 6 February 2020 (China standard time), 14 days after the first day of the Wuhan lockdown (23 January 2020). The threshold of 14 days is the employed length of quarantine or active monitoring of people potentially exposed to the virus. The proposed length is supported by the findings of epidemiological studies [35].

### 3.3. Intercity Travel Data and Urban Factor Data

The intercity travel data adopted in this study were generally collected from the users of the mobile devices by location-based service operators. They are a type of mobile positioning data produced by mobile devices when their owners use software supported by the function of location-based services, such as the Internet, map services, booking services, express services, and entertainment services. Under the authorization of mobile device users, the operators of location-based services have access to visit the system interface of the mobile device to collect its GPS coordinate locations during daily use.

Intercity travel data reflect population movements across regions on the basis of positional shifts of personal mobile devices. The data come from the Baidu Qianxi data of Baidu Inc., the largest search engine operator in China (<http://qianxi.baidu.com>). It can be viewed as a counterpart of Google in the Western world. The Baidu Qianxi data are collected and verified by the Baidu Location-based Service Platform, which shows the dynamic regional population flow on a daily basis. Baidu states that over 120 billion location requests are supported by its location-based service platform daily. The data cover at least 70% of the Chinese population, and the sampling size is higher than that of national census data. An anonymization algorithm is used to ensure that all the travel data are collected without any sensitive information, for the consideration of confidentiality. Only spatial and temporal tags are recorded for spatial data analysis and data visualization. Wuhan entered lockdown on 23 January 2020, one day before Chinese New Year's Eve, when the spring-festival travel season had almost come to an end. We obtained the number of people flowing out of Wuhan to other cities within 14 days before the lockdown because the incubation period of the infection is thought to be about 14 days. The spring-festival travel season typically begins two weeks before Chinese New Year's Eve. Therefore, the travel data reflecting population outflow from Wuhan from 10 January to 23 January 2020 needed to be extracted. According to the Baidu data, there are at least 3.31 million passengers who left Wuhan for other domestic cities during the mentioned 14 days. We used de-identified and cumulative domestic population-movement data for the two weeks before and after the travel restrictions in and out of Wuhan derived from Baidu's location-based services, and the daily numbers of journeys from Wuhan to other Chinese cities were derived.

The resident population size, population density, and GDP of the cities were extracted from the 2019 City/Prefectures Statistical Yearbook of China [36].

### 3.4. Data Processing

The intercity travel data bar is composed of four columns, which represent the departure city, destination city, travel date, and flow population (Table 1). The departure city is typically identified as a place of residence or place of an overnight stay. The destination city was only identified as a meaningful destination if someone stayed in the city for at least 4 hours continuously; otherwise, it was ignored and assumed to be a place of temporary stay. The intercity flow from Wuhan to any other city within the two weeks before Spring Festival is described by the following equation:

$$F_{wi} = \sum_{d=1}^n F_{di} \quad (1)$$

where  $F_{wi}$  is the aggregate population outflow from Wuhan to city  $i$ , and  $F_{di}$  is the daily population outflow from Wuhan to city  $i$ .

**Table 1.** Raw dataset of mobile internet positioning data.

Departure city	Destination city	Travel date	Flow population (persons)
Wuhan	Beijing	10 January 2020	3350
Wuhan	Shanghai	10 January 2020	2432
Wuhan	...	...	...
Wuhan	Beijing	11 January 2020	...
Wuhan	Shanghai	11 January 2020	...
Wuhan	...	...	...
Wuhan	...	23 January 2020	...

Note: If there was no one that travelled from Wuhan to a city within the research period, the record value is “0” in the column of flow population.

#### 4. Research Methodology

##### 4.1. Ratio of the Risk of Imported COVID-19

The ratio of the risk of imported COVID-19 is introduced to evaluate the effect of intercity population movements on infectious disease transmission across cities at the beginning of the epidemic. The quotient of the cumulative confirmed cases over the population outflow from Wuhan is defined as the ratio of the risk of imported COVID-19 spread for each city using the following equation:

$$R_i = \frac{\sum_{d=1}^n C_{di}}{F_{wi}}, \quad (2)$$

where  $R_i$  is the ratio of the risk of imported COVID-19 in city  $i$ , and  $C_{di}$  is the daily confirmed COVID-19 cases in city  $i$  since the epidemic outbreak. The cumulative number of confirmed cases is summarized up to 6 February 2020. The ratio of the risk is described as the confirmed cases per thousand population as a result of outflow from Wuhan.

##### 4.2. Bivariate Aspect of Local Spatial Autocorrelation

This study mainly analyzed how population movements from Wuhan into other cities outside Hubei contributed to the spread of the COVID-19 epidemic. Local spatial autocorrelation was first analyzed to explore the spatial distribution characteristics of confirmed cases and detect their spatial correlation with population movement. Correlation and regression analyses were carried out to estimate the associations among the population outflow from Wuhan, gross domestic product (GDP), resident population size, population density, and cumulative number of confirmed COVID-19 cases.

The spatial autocorrelation analysis consisted of global spatial autocorrelation and local spatial autocorrelation. The former was used to reveal the integral characteristics of the spatial distribution and its significance in the study area, while the latter was used to explore the potential local distribution characteristics of spatial elements, such as aggregated or random distributions. Moran’s  $I$  and local Moran’s  $I$  were used to indicate the spatial relationship of global spatial autocorrelation and local spatial autocorrelation, respectively. Moran’s  $I$  is calculated as follows:

$$I = \frac{\sum_{i=1}^n \sum_{j \neq i}^n W_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j \neq i}^n W_{ij}}. \quad (3)$$

Local Moran’s  $I$  is calculated as follows:

$$I_i = \frac{(x_i - \bar{x}) \sum_{i=1, j \neq i}^n W_{ij} (x_j - \bar{x})}{S^2}. \quad (4)$$

Here,  $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ ,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $n$  is the number of spatial units for analysis,  $x_i$  and  $x_j$  represent the measured values of the spatial units  $i$  and  $j$ ,  $(x_i - \bar{x})$  represents the deviation between

the measured value of unit  $i$  and the mean value, and  $W_{ij}$  is the spatial weight matrix according to the contiguity relation.

The bivariate spatial autocorrelation was further evolved from spatial autocorrelation to explore the spatial correlation of multiple variables [37]. Bivariate local Moran's  $I$  indicator was employed to identify the spatial clusters and outliers of intercity flows with confirmed COVID-19 cases. The bivariate local spatial autocorrelation model can be described as follows:

$$I_{B,i} = X_i \sum_{j=1}^n W_{ij} Y_j, \quad (5)$$

where  $X_i = \frac{x_i - \bar{x}}{\sigma_x}$ ,  $Y_j = \frac{y_j - \bar{y}}{\sigma_y}$ ,  $x_i$  is the measured value of variable  $x$  of the spatial unit  $i$ ,  $y_j$  is the measured value of the variable  $y$  of the spatial unit  $j$ ,  $\bar{x}$  and  $\bar{y}$  represent the mean values of the variables  $x$  and  $y$ , and  $\sigma_x$  and  $\sigma_y$  represent the variance of the variables  $x$  and  $y$ . The spatial association between the values of variable  $x$  at location  $i$  ( $i = 1, 2, \dots, n$ ) and the average of the neighbor values for variable  $y$  at location  $j$  ( $j = 1, 2, \dots, n$ ) can be determined. The statistic  $I_{B,i}$  is the product of  $X_i$  with a spatial lag of  $Y_j$  [38,39]. Variable  $x_i$  in this study is the cumulative population outflow from Wuhan to city  $i$  in the two weeks before the Wuhan lockdown, while variable  $y_j$  is the cumulative number of COVID-19 cases confirmed in city  $j$  by 6 February 2020.

#### 4.3. Multivariate Regression

Regression analysis was performed to examine the effect of independent variables, i.e., the population outflow from Wuhan and other potential factors, on a dependent variable, i.e., the cumulative number of confirmed COVID-19 cases. Through testing the correlation between independent and dependent variables, a regression model can be established to describe the magnitude of the effects of the predictor variables on response variables. In this research, the effect was exactly the ratio of the risk of imported confirmed COVID-19 cases in China at the prefecture level, which can be regarded as the epidemic risk interpretation model. It was formulated with multivariate linear regression as follows:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon, \quad (6)$$

where  $Y_i$  is the cumulative number of confirmed cases in city  $i$  ( $i = 1, 2, \dots, n$ ),  $X_{i1}$  is the population outflow from Wuhan,  $X_{i2}$  is the GDP,  $X_{i3}$  is the resident population size,  $X_{i4}$  is the population density,  $\beta_0$  is a constant, and  $\beta_1, \dots, \beta_4$  represent the regression coefficients for each independent variable. All coefficients were obtained using the least-squares method and the constant  $\beta_0$ .  $\varepsilon$  is the error term.

In this study, the spatiotemporal data were called by PostgreSQL. The data visualization was realized by ArcGIS 10.7. The spatial autocorrelation analysis was processed using Geoda 1.14, while correlation and regression analyses were processed using SPSS 19.0.5.

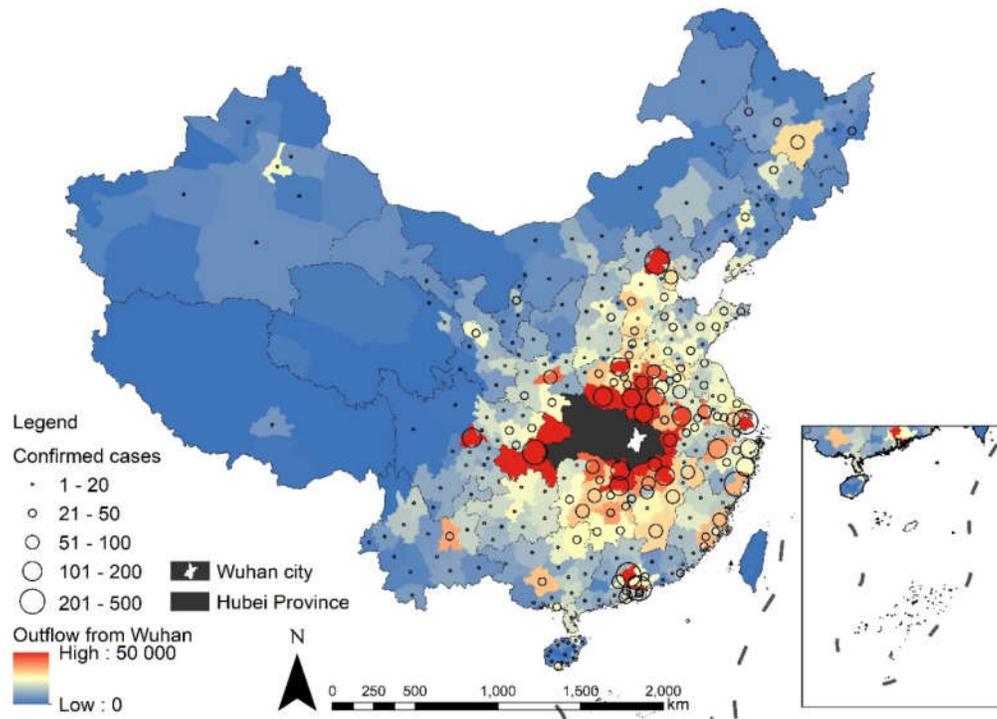
## 5. Spatial Cluster and Regression Analyses.

### 5.1. Population Outflow from Wuhan to Non-Hubei Cities and the Cumulative Confirmed Cases

China has seen rapid urbanization in the past 40 years. Massive migration from rural to urban areas has shaped the flow characteristics of Chinese intercity population movements. The recent construction of a high-speed railway system has further increased the scale and frequency of intercity flows, especially during the Spring Festival period. According to Baidu's daily migration data, 3.31 million mobile-device users traveled from Wuhan from 10 January to 23 January 2020, of which 1.02 million traveled from Wuhan to non-Hubei cities within 14 days. According to the daily epidemic reports from the China National Health Commission, the cumulative number of confirmed COVID-19 cases was 19,557 (only excluding Wuhan city) as of 6 February 2020, and there were 9063 COVID-19 cases confirmed with Wuhan and other cities in Hubei province excluded.

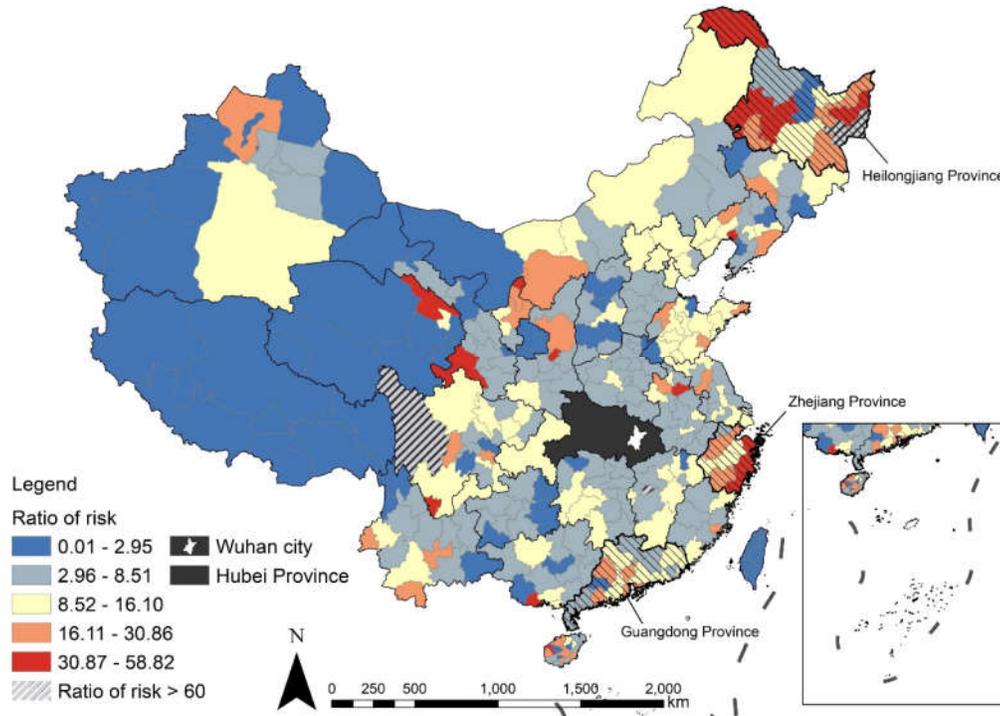
Most of the population outflow from Wuhan before the lockdown was to east parts of China. COVID-19 quickly spread to almost all cities with the large-scale population movement (Figure 2).

Additionally, the spatial distribution of the population outflows and confirmed cases in different cities did not strictly follow the rule of geographical proximity with Wuhan. The large-scale population outflows were not only distributed across the cities surrounding Hubei province but also over distant metropolitan regions such as Beijing, Shanghai, Guangzhou, and Shenzhen. The spatial distribution of cumulative confirmed cases was relatively concentrated in these regions, where the numbers of confirmed cases generally ranged from 201 to 500. By the comparison of the spatial distribution of the outflow from Wuhan and confirmed cases, more specific spatial characteristics of the spread of the epidemic can be explored.



**Figure 2.** Spatial distribution of population outflow from Wuhan and cumulative confirmed COVID-19 cases.

The ratio of the risk of imported COVID-19 transmission was used to describe the relationship between population flow and confirmed cases at the beginning of the epidemic (Figure 3). In all 360 cities, the median value was 8.2 confirmed cases per thousand population as a result of outflow from Wuhan. There were four cities in which the ratio of risk was beyond 60.0 cases per thousand population as a result of outflow, with values of 74.0, 96.8, 116.7, and 354.2. These four cities were located in marginal areas in terms of geography and population movement. The mean value of population outflow from Wuhan in these four cities was 374.0, while the mean value for all 360 cities was 2817.0. It was necessary to exclude these four marginal cities with minor outflow from Wuhan so that the classification outcome for the ratio of risk became more significant.



**Figure 3.** Ratio of the risk of imported 2019 novel coronavirus (COVID-19) transmission for each city (by natural Jenks).

According to the classification outcome for the ratio of risk using natural Jenks, the cities in which the ratio of risk was beyond 16.1 cases per thousand population as a result of outflow were not only distributed in the eastern and southeastern coastal parts of China (the most densely populated regions) but also in the northern parts. The ratios of the risk of imported COVID-19 in the regions around Wuhan were surprisingly low, even though Wuhan is recognized as the first epicenter of COVID-19. For example, the average ratio across China was 8.89 confirmed cases per thousand population as a result of outflow at the province level. Zhejiang, Heilongjiang, and Guangdong Province were ranked as the top three in terms of ratio of risk, with average values of 28.2, 24.16, and 16.1, respectively. However, the average was only 4.84 in Henan Province, which ranked first for population movement from Wuhan to a non-Hubei province. Compared with the scale of population outflow, the high ratios of risk in some parts of China such as Heilongjiang Province resulted from rather small-scale outflow from Wuhan. The population movement from Wuhan to Heilongjiang Province was only 9.93 thousand, while it was 35.71 thousand to Zhejiang Province and 63.11 thousand to Guangdong Province.

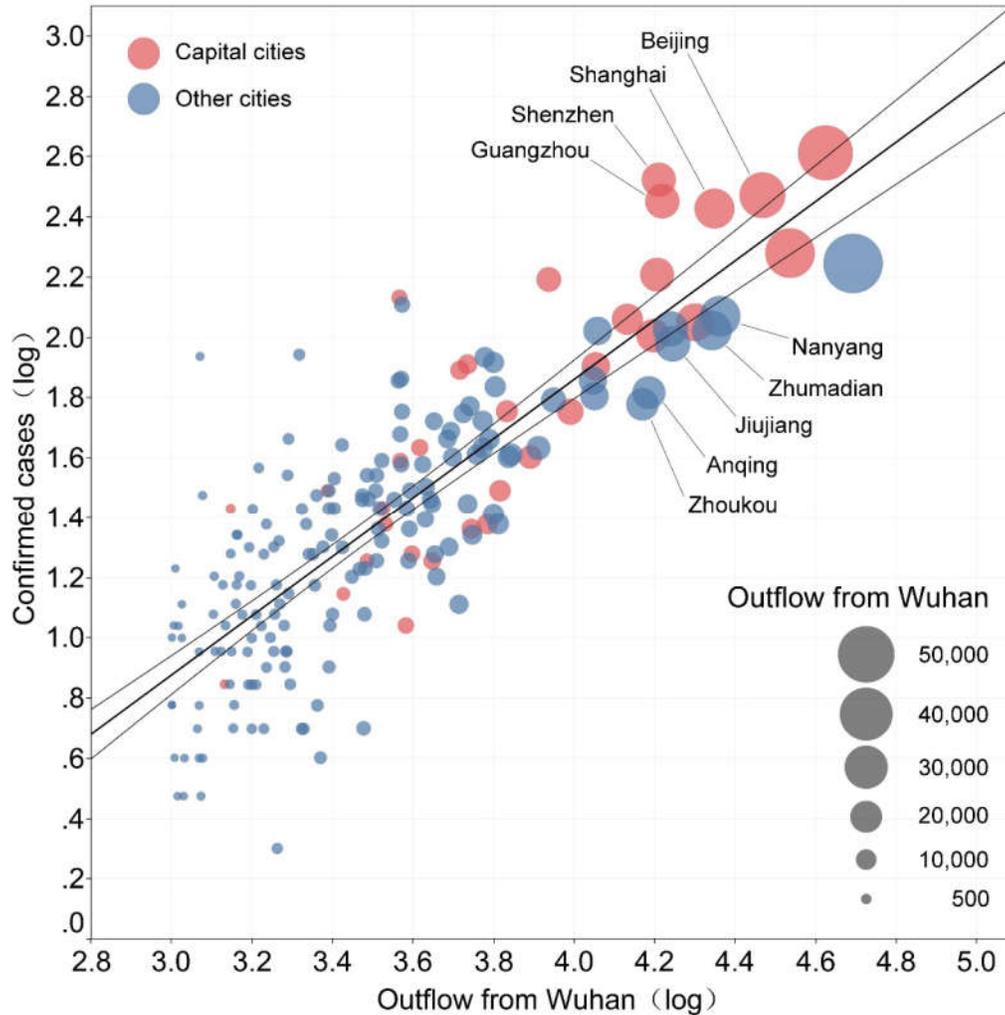
Other local factors, such as the travel behavior of the elderly or social events, may complicate COVID-19 transmission, even when there is a limited number of imported cases. However, this work mainly focused on the intercity population flow, as well as the spatial factors and socioeconomic factors of cities. It is, therefore, necessary to set a lower limit of population outflow to eliminate the effects of various local factors on the spread of COVID-19 in cities.

### 5.2. Correlation between Intercity Flow and COVID-19 Epidemic

It was more reasonable to focus on the cities with larger-scale population outflow from Wuhan to avoid the effect of accidental uncontrollable factors on the spread of COVID-19. For the regression analysis, a value of one thousand for population outflow within two weeks before Chinese New Year's Eve was set as the lower threshold for each city. Cities below this threshold were regarded as marginal areas of intercity flow. Eventually, 191 cities were selected as the study units for further

regression analysis, covering more than half of the domestic cities and 1.06 billion people, according to the China Statistical Yearbook 2019 [36].

A linear regression model was established to demonstrate the magnitude of the effect of the population outflow from Wuhan, as the single variable, on confirmed cases (Figure 4). For the 191 selected cities, the population outflows from Wuhan had significant impacts on the cumulative numbers of confirmed cases ( $R^2 = 0.650$ ,  $p < 0.001$ ). For each 10% increase in population outflow from Wuhan, the cumulative number of confirmed cases was predicted to increase by 8.06%.



**Figure 4.** The correlation between population outflow from Wuhan and cumulative confirmed COVID-19 cases.

A dummy variable was added to test the related hypothesis that differences in city administrative class had an effect of equivalent magnitude on the spread of COVID-19, as follows:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \lambda_1 Cap\_City_j + \varepsilon_{ij} \quad (7)$$

where  $Y_{ij}$  is the cumulative number of confirmed cases in city  $i$  of city class  $j$ .  $X_{ij}$  is the population outflow from Wuhan to city  $i$ .  $Cap\_City_j$  represents the city class difference ( $Cap\_City_j = 0, 1$ ), where  $Cap\_City_j = 1$  when the city is a capital city. Provincial-level capitals and sub-provincial cities are referred as “capital cities” in this study.  $\lambda_1$  represents the regression coefficient for the dummy variable.

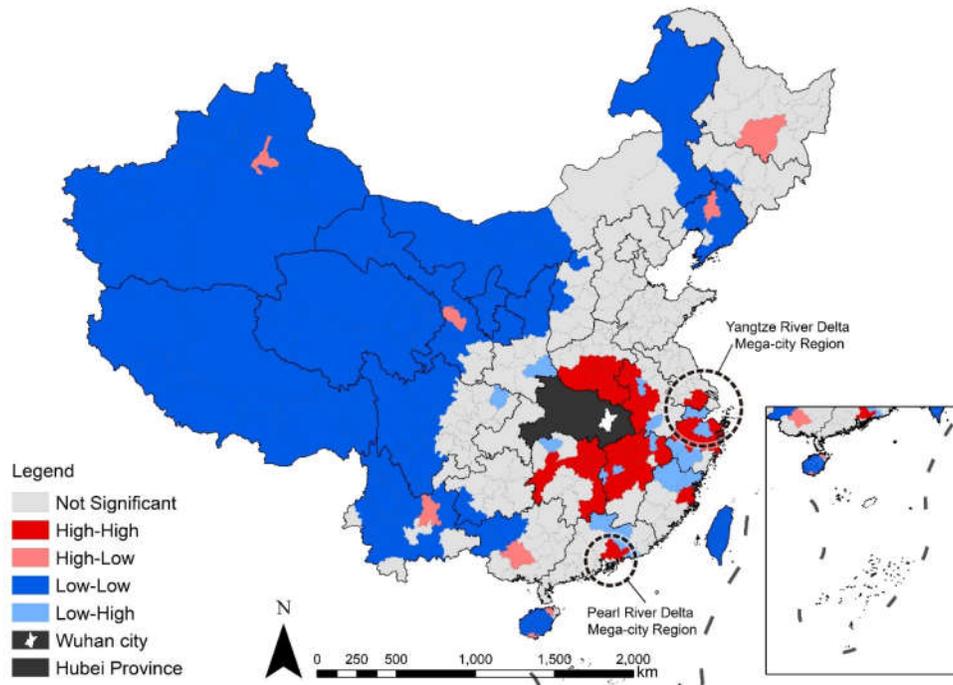
The  $R^2$  increased to 0.685 ( $p < 0.001$ ) when the dummy variable of city administrative class was added into the model, indicating a better goodness of fit. The standardized coefficient of city

administrative class is 0.202 ( $p < 0.001$ ), which means capital cities have a higher ratio of risk of COVID-19 transmission than other cities. For instance, the ratio of risk for mega-cities and provincial-level capitals, such as Beijing, Shanghai, Guangzhou, and Shenzhen, was 14.04. However, it was only 4.80 in the cities surrounding Wuhan city, such as Nanyang, Zhumadian, Jiujiang, Anqing, and Zhoukou, even though their population outflows all ranged from 15 to 30 thousand people within two weeks. It was estimated that the ratio of risk of imported disease spread in mega-cities and provincial capitals was nearly 2.93 times that in other cities, even with equivalent intercity population movements.

### 5.3. Bivariate Local Spatial Association of Intercity Flows and Confirmed Cases

The spatial autocorrelation analysis in this research focused on the domestic prefecture level with queen contiguity-based spatial weights. The correlated results characterize the spatiotemporal distribution of the cumulative population outflow and the cumulative number of confirmed cases. The high-high and low-low clusters at the prefecture level represent the regions where the cumulative number of confirmed COVID-19 cases showed a significant positive correlation with the population outflow from Wuhan. The high-low and low-high clusters reflect the regions for which negative correlations were observable. Through the recognition of these high-high clusters and high-low clusters, the influence of population movement on the transmission of COVID-19 from a spatial perspective could be further explored.

The result showed that intercity flow had greatly contributed to the spread of COVID-19. The spatial distribution of outflows and confirmed cases was shown to have a positive spatial autocorrelation, with a Moran's  $I$  value of 0.161 ( $p < 0.05$ ). Two distinct spatial distributions of high-high clusters were identified (Figure 5). The high-high clusters were mainly concentrated around Wuhan city because of spatial proximity, while the other high-high clusters were concentrated in mega-city regions such as the Yangtze River Delta and Pearl River Delta mega-city regions, both of which are relatively far away and have strong population flows from Wuhan and other cities within the mega-city regions [40]. These high-high clusters represent much higher spatial correlation between population outflow and confirmed cases, which also indicates much higher COVID-19 infectious risks due to the intercity population movement. On the contrary, the spatial distributions of high-low clusters were in the provincial capitals. Compared with the low-low and low-high clusters, it is concluded that the cities located in mega-city regions and provincial capitals had much higher risks for the spread of COVID-19 at the beginning of epidemic transmission.



**Figure 5.** Bivariate local spatial autocorrelation with population outflow from Wuhan and cumulative confirmed COVID-19 cases.

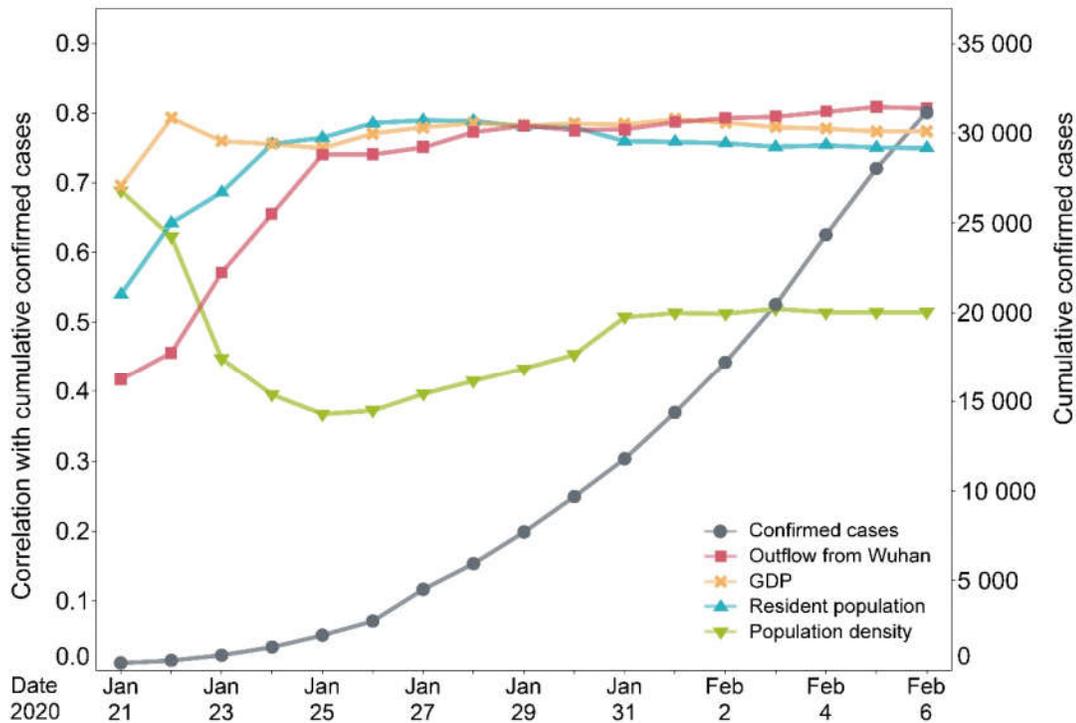
#### 5.4. Multivariate Regression Analyses at the Prefecture Level

The relationship over time between the number of confirmed cases and socioeconomic factors was evaluated with the Pearson's correlation coefficients. This descriptive statistic for the variables and its correlation with the confirmed cases up to 6 February 2020 is summarized (Table 2). The variation of the correlation coefficient was assessed from 21 January 2020, when the first COVID-19 case was detected outside Hubei province, to 6 February 6 2020, when Wuhan entered lockdown for 14 days (Figure 6). With the rapid increase in cumulative confirmed cases, the correlation between population outflow and confirmed cases increased from Pearson's  $r = 0.421$  to  $0.805$ . At the same time, the correlation between GDP and confirmed cases reached  $0.775$ , while that between resident population size and confirmed cases reached  $0.749$ . On the contrary, the correlation between population density and confirmed cases showed a distinct decline in the primary stage of the epidemic spread and then gradually increased to a stable level, with Pearson's  $r = 0.508$ , on 6 February 2020. The correlation coefficients for these socioeconomic factors all stabilized with slight fluctuations from January 31st. These four variables therefore showed strong correlations with the cumulative confirmed cases during the outbreak stage of COVID-19 intercity transmission.

**Table 2.** Descriptive statistics of population outflow and other socioeconomic variables and their relationships with cumulative confirmed COVID-19 cases.

Variables	Mean	St. Deviation	Coefficient
Accumulative confirmed cases	41.21	57.58	-
Population outflow from Wuhan (persons)	4952.49	6656.63	0.805***
Resident population (thousands)	5477.49	3711.82	0.749***
GDP (million USD)	55,211.39	68,678.67	0.775***
Population density (persons/km <sup>2</sup> )	393.26	613.88	0.508***

Note: The accumulative confirmed cases refer to the number of cumulative confirmed cases as of 6 February 2020. Population outflow from Wuhan refers to the intercity population movements within 14 days before Spring Festival. The exchange rate of the RMB against the USD is 6.75. \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , and \*  $p < 0.05$ .



**Figure 6.** The correlations between socioeconomic variables and cumulative confirmed cases according to Pearson coefficients.

Multivariate regression was preformed to establish the interpretation model for quantifying the effect of population intercity flows and other socioeconomic factors on the spread of COVID-19 in general conditions (Table 3). From the results of model, the regression passed the  $F$  test at the 0.01 significance level, and the Durbin–Watson index was 2.063, showing a good normal distribution of residuals. The determination coefficients  $R^2 = 0.603$  and adjusted  $R^2 = 0.594$  indicate good model fit, showing that the model was able to explain 59.4% of the variability in the cumulative confirmed cases.

**Table 3.** Association between cumulative confirmed COVID-19 cases and other socioeconomic variables according to multivariate regression model.

Independent variables	Standardized coefficients	95% CI
Population outflow from Wuhan (persons)	0.488***	(0.000, 0.000)
Resident population size (thousand people, logarithm)	0.244***	(0.241, 0.635)
GDP scale (million USD, logarithm)	0.216***	(0.203, 0.573)
Population density (persons/km <sup>2</sup> )	0.116*	(0.068, 1.631)
<b>Model summary</b>		
$R^2$	0.603	
adj_ $R^2$	0.594	
MSA	5.743	
MSE	0.081	
F (sig<0.01)	70.594	
Durbin–Watson	2.063	

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ .

According to the standardized coefficients, all the independent variables had significant impacts on the number of confirmed COVID-19 cases. Among them, the population outflow from Wuhan shows the most significant impact, followed by the population size, GDP, and population density. For every unit increase in population outflow from Wuhan, the cumulative number of confirmed cases was predicted to increase by 0.488. In addition, for every unit increase in population size, GDP, and population density, the cumulative number of confirmed cases was expected to increase by 0.244 (95% CI: 0.241–0.635), 0.216 (95% CI: 0.203–0.573), and 0.116 (95% CI: 0.068–1.631), respectively.

## 6. Findings and Discussion

The outbreak of COVID-19 resulted in a considerable crisis of epidemic spread due to the huge scale of intercity population flow during the peak travel season before Spring Festival, in consideration of the geographically central location of Wuhan city in China. Due to the intercity travel restriction policy implemented in Wuhan on 23 January 2020, the imported risk of COVID-19 from Wuhan was significantly controlled, which provided an ideal period for analyzing the relationships among imported confirmed cases, intercity population movement, and other potential urban factors.

On the basis of mobile positioning big data, the spatiotemporal population movement and the transmission characteristics of COVID-19 during the early outbreak stage were explored. We noted a value of 1.02 million for population outflow from Wuhan to non-Hubei cities, representing 30.82% of the total population outflow from Wuhan before the lockdown. Meanwhile, there were 9063 confirmed cases outside Hubei Province, representing 46.34% of the total number of cumulative confirmed cases in China on 6 February 2020. Intercity population movement played a vital role in the COVID-19 spatial spread in China.

First, the results reveal a significant positive correlation between population movement and the number of COVID-19 cases in different cities, and the correlation varied with time. The large-scale intercity population movement led to a higher transmission risk for cities in China.

Second, the impact of intercity movement on the spread of epidemics varies for different levels of cities, including capital and other cities, in China. Cities with a higher administrative class suffered from a higher risk of transmission when faced with the same level of population movement. Compared with noncapital cities, capital cities may need to take tougher quarantine measures to reduce the spread of epidemics because of their large social and economic scale.

Third, the spatial distribution characteristics of cities have an impact on the spread of infectious diseases. There was a positive spatial autocorrelation between population movement and the number of confirmed cases, with a Moran's  $I$  value of 0.161 ( $p < 0.05$ ). Large-scale and frequent intercity travel has a "positive" effect on the promotion of regional economic development, but it may also have "negative" effects. Examining intercity travel flows using big data from location-based services together with epidemic data may help to elucidate the "negative" effects during epidemics. Cities located in the Yangtze River Delta and Pearl River Delta mega-city regions faced higher transmission risk.

Fourth, the analysis further showed that the spread of epidemics varies among different cities because it is affected by urban factors including the total population, population density, and GDP, in addition to population movement. Cities with different socioeconomic attributes may assess the potential risk of epidemics ahead of time. In addition to population movement, resident population size had a great effect on the transmission risk, followed by the GDP and population density.

Lastly, the intercity population flows generated from mobile positioning big data made it possible to examine how urban factors affected the spatiotemporal distribution of the COVID-19 epidemic from an individual perspective at the national level, which is more efficient than labor-intensive personal surveys. The analytical framework provides a convenient way of preliminarily assessing the potential risk of epidemics in the early stage if mobile positioning big data are available for these cities. One limitation of this research is that the mode of intercity travel and the intercity travel time were not considered, as they were difficult to accurately ascertain from the data.

Geographical models such as gravity models or potential models supported with randomization will be used to forecast the spatial diffusion of the COVID-19 epidemic in future studies.

## 7. Conclusions

The emergence of large-scale, location-based-service data has provided an opportunity to explore intercity population movement from individuals' perspective rather than relying on census or travel survey data, which are limited by sampling size. It is possible to examine how intercity population movement and urban factors affect the spatiotemporal distribution of infectious diseases using intercity travel data. This study used massive intercity-movement data from Baidu's location-based service data to explore the impact of population outflow on the spread of an epidemic in China. Our data-analytical framework is generalizable to other datasets that capture intercity population flow.

The implications in China are applicable for other nations to some extent, and policymakers in any nation may introduce necessary control measures beforehand. Mega-city regions may face greater risks than other regions during epidemic events due to the large and frequent flows between the cities of mega-city regions. Such regions may manage population flows among cities using intercity-travel big data in a collaborative manner to contain infectious diseases [41]

**Author Contributions:** Conceptualization, Xinyi Niu and Xingang Zhou; formal analysis, Yufeng Yue; funding acquisition, Xingang Zhou; methodology, Yufeng Yue; resources, Xinyi Niu; writing—original draft, Yufeng Yue and Xingang Zhou; writing—review and editing, Xingang Zhou and Xiaohu Zhang. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Natural Science Foundation of China (41801147, 51878457) and Shanghai Pujiang Talent Program.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Ye, L.; Hu, L. Spatiotemporal distribution and trend of COVID-19 in the Yangtze River Delta region of the People's Republic of China. *Geospat. Health* **2020**, *15*, 1.
- Dalziel, B.D.; Kissler, S.; Gog, J.R.; Viboud, C.; Bjornstad, O.N.; Metcalf, C.; Grenfell, B.T. Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. *Science* **2018**, *362*, 75–79.
- Lai, S.; Ruktanonchai, N.W.; Zhou, L.; Prosper, O.; Tatem, A.J. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **2020**, 1–7. doi: 10.1038/s41586-020-2293-x.
- Tian, H.; Liu, Y.; Li, Y.; Wu, C.; Chen, B.; Kraemer, M.U.G.; Li, B.; Cai, J.; Xu, B.; Yang, Q.; et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **2020**, *368*, 638–642.
- Simiao, C.; Juntao, Y.; Weizhong, Y.; Chen, W.; Till, B. COVID-19 control in China during mass population movements at New Year. *Lancet Lond. Engl.* **2020**, *395*, 764–766.
- Chinazzi, M.; Davis, J.T.; Ajelli, M.; Gioannini, C.; Litvinova, M.; Merler, S.; Piontti, A.P.Y.; Mu, K.; Rossi, L.; Sun, K.; et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **2020**, *368*, 395–400.
- Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* **2020**, *368*, 489–493.
- Kraemer, M.U.G.; Yang, C.; Gutierrez, B.; Wu, C.; Klein, B.; Pigott, D.M.; Plessis, L.D.; Faria, N.R.; Li, R.; Hanage, W.P.; et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **2020**, *368*, 493–497.
- Fan, C.; Liu, L.; Guo, W.; Yang, A.; Ye, C.; Jilili, M.; Ren, M.; Xu, P.; Long, H.; Wang, Y. Prediction of Epidemic Spread of the 2019 Novel Coronavirus Driven by Spring Festival Transportation in China: A Population-Based Study. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1679.
- Chen, Z.; Zhang, Q.; Lu, Y.; Guo, Z.; Zhang, X. Distribution of the covid-19 epidemic and correlation with population emigration from wuhan, China. *Chin. Med. J.* **2020**, *133*. doi: 10.1097/CM9.0000000000000782
- Balcan, D.; Colizza, V.; Gonçalves, B.; Hu, H.; Ramasco, J.J.; Vespignani, A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 21484–21489.
- Wells, C.R.; Sah, P.; Moghadas, S.M.; Pandey, A.; Shoukat, A.; Wang, Y.; Wang, Z.; Meyers, L.A.; Singer, B.H.; Galvani, A.P. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7504–7509.

13. Zhou, Y.; Xu, R.; Hu, D.; Yue, Y.; Li, Q.; Xia, J. Effects of human mobility restrictions on the spread of COVID-19 in Shenzhen, China: A modelling study using mobile phone data. *Lancet Digit. Health* **2020**, *2*, 417–424.
14. Massaro, E.; Kondor, D.; Ratti, C. Assessing the interplay between human mobility and mosquito borne diseases in urban environments. *Sci. Rep.* **2019**, *9*, 16911.
15. Mo, C.; Tan, D.; Mai, T.; Bei, C.; Qin, J.; Pang, W.; Zhang, Z. An analysis of spatiotemporal pattern for COVID-19 in China based on space-time cube. *J. Med Virol.* **2020**, *92*, 1587–1595.
16. Yang, W.; Deng, M.; Li, C.; Huang, J. Spatio-Temporal Patterns of the 2019-nCoV Epidemic at the County Level in Hubei Province, China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2563.
17. Dalziel, B.D.; Pourbohloul, B.; Ellner, S.P. Human mobility patterns predict divergent epidemic dynamics among cities. *Proc. Biol. Sci.* **2013**, *280*, 20130763.
18. Jia, J.S.; Lu, X.; Yuan, Y.; Xu, G.; Jia, J.; Christakis, N.A. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* **2020**, *582*, 389–394.
19. Engelfriet, L.; Koomen, E. The impact of urban form on commuting in large Chinese cities. *Transportation* **2018**, *45*, 1269–1295.
20. Hamidi, S.; Sabouri, S.; Ewing, R. Does Density Aggravate the COVID-19 Pandemic? *J. Am. Plan. Assoc.* **2020**, 1–15. doi: 10.1080/01944363.2020.1777891
21. Glaeser, E. Cities, Productivity, and Quality of Life. *Science* **2011**, *333*, 592–594.
22. Garske, T.; Yu, H.; Peng, Z.; Ye, M.; Zhou, H.; Cheng, X.; Wu, J.; Ferguson, N. Travel Patterns in China. *PLoS ONE* **2011**, *6*, e16364.
23. Gaughan, A.E.; Stevens, F.R.; Huang, Z.; Nieves, J.J.; Sorichetta, A.; Lai, S.; Ye, X.; Linard, C.; Hornby, G.M.; Hay, S.I.; et al. Spatiotemporal patterns of population in mainland China, 1990 to 2010. *Nature* **2016**, *3*, 160005.
24. Liu, T.; Qi, Y.; Cao, G.; Liu, H. Spatial patterns, driving forces, and urbanization effects of china's internal migration: County-level analysis based on the 2000 and 2010 censuses. *J. Geogr. Sci.* **2015**, *25*, 236–256.
25. Mu, X.; Yeh, A. Regional delineation of China based on commuting flows. *Environ. Plan. A* **2020**, *52*, 478–482.
26. Yang, C.; Clarke, K.; Shekhar, S.; Tao, C. Big Spatiotemporal Data Analytics: A research and innovation frontier. *Int. J. Geogr. Inf.* **2020**, *34*, 1075–1088.
27. Zhou, X.; Yeh, A.; Li, W.; Yue, Y. A commuting spectrum analysis of the jobs–housing balance and self-containment of employment with mobile phone location big data. *Environ. Plan. B Urban Anal. City Sci.* **2018**, *45*, 434–451.
28. Kwan, M.P.; Neutens, T. Space-Time Research in GIScience. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 851–854.
29. Menach, A.L.; Tatem, A.J.; Cohen, J.M.; Hay, S.I.; Randell, H.; Patil, A.P.; Smith, D.L. Travel risk, malaria importation and malaria transmission in Zanzibar. *Sci. Rep.* **2011**, *1*, 93.
30. Buckee, C.; Balsari, S.; Chan, J.; Crosas, M.; Dominici, F.; Gasser, U.; Grad, Y.; Grenfell, B.; Halloran, M.; Kraemer, M.; et al. Cumulative mobility data could help fight COVID-19. *Science* **2020**, *368*, eabb8021.
31. Liu, Y.; Sui, Z.; Kang, C.; Gao, Y. Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from Social Media Check-In Data. *PLoS ONE* **2014**, *9*, e86026.
32. Yang, X.; Fang, Z.; Xu, Y.; Shaw, S.; Zhao, Z.; Yin, L.; Zhang, T.; Lin, Y. Understanding Spatiotemporal Patterns of Human Convergence and Divergence Using Mobile Phone Location Data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 1.
33. Mu, X.; Yeh, A.; Zhang, X. The interplay of spatial spread of COVID-19 and human mobility in the urban system of China during the Chinese New Year. *Environ. Plan. B Urban Anal. City Sci.* **2020**. doi: 10.1177/2399808320954211.
34. Yang, C.; Sha, D.; Liu, Q.; Li, Y.; Lan, H.; Guan, W.; Hu, T.; Li, Z.; Zhang, Z.; Thompson, J.; et al. Taking the pulse of COVID-19: A spatiotemporal perspective. *Int. J. Digit. Earth* **2020**, doi:10.1080/17538947.2020.1809723.
35. Lauer, S.; Grantz, K.; Bi, Q.; Jones, F.; Zheng, Q.; Meredith, H.; Azman, A.; Reich, N.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine.* **2020**, *172*, 577–582.
36. China Statistical Yearbook. 2019. Available online: <http://www.stats.gov.cn/tjsj/ndsj/2019/indexeh.htm> (accessed on 5 July 2020).
37. Wartenberg, D. Multivariate spatial correlation: A method for exploratory geographical analysis. *Geogr. Anal.* **1985**, *17*, 263–283.

38. Anselin, L.; Syabri, I.; Smirnov, O. Visualizing Multivariate Spatial Correlation with Dynamically Linked Windows. In *New Tools for Spatial Data Analysis: Proceedings of the Specialist Meetin*; Anselin, L., Rey, S., Eds.; Center for Spatially Integrated Social Science (CSISS): Santa Barbara, CA, USA, 2002.
39. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115.
40. Yeh, A.; Chen, Z. From cities to super mega city regions in China in a new wave of urbanisation and economic transition: Issues and challenges. *Urban Stud.* **2019**, *57*, 636–654.
41. Bai, X.; Nagendra, H.; Shi, P.; Liu, H. Cities: Build networks and share plans to emerge stronger from COVID-19. *Nature* **2020**, *584*, 517–520.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).