Natural and human environment interactively drive spread pattern of COVID-19: A city-level modeling study in China

Xiaoxu Wu, Jie Yin, Chenlu Li, Hongxu Xiang, Meng Lv, Zhiyi Guo

PII: S0048-9697(20)36874-1
DOI: https://doi.org/10.1016/j.scitotenv.2020.143343
Reference: STOTEN 143343

To appear in: Science of the Total Environment

Received date: 26 August 2020
Revised date: 20 October 2020
Accepted date: 24 October 2020

Please cite this article as: X. Wu, J. Yin, C. Li, et al., Natural and human environment interactively drive spread pattern of COVID-19: A city-level modeling study in China, Science of the Total Environment (2020), https://doi.org/10.1016/j.scitotenv.2020.143343

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.
Natural and human environment interactively drive spread pattern of COVID-19: a city-level modeling study in China

Xiaoxu Wu, Jie Yin, Chenlu Li, Hongxu Xiang, Meng Lv, Zhiyi Guo

a State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China

† Co-first authors

* Correspondence author
Abstract

A novel Coronavirus COVID-19 has caused high morbidity and mortality in China and worldwide. A few studies have explored the impact of climate change or human activity on the disease incidence in China or a city. The integrated study concerning environment impact on the emerging disease is rarely reported. Therefore, based on the two-stage modeling study, we investigate the effect of both natural and human environment on COVID-19 incidence at a city level. Besides, the interactive effect of different factors on COVID-19 incidence is analyzed using Geodetector; the impact of effective factors and interaction terms on COVID-19 is simulated with Geographically Weighted Regression (GWR) models. The results find that mean temperature (MeanT), Destination proportion in population flow from Wuhan (WH), Migration scale (MS), and WH*MeanT are generally promoting for COVID-19 incidence before Wuhan’s shutdown (T1); the WH and MeanT play a determinant role in the disease spread in T1. The effect of environment on COVID-19 incidence after Wuhan’s shutdown (T2) includes more factors (including relative humidity, precipitation, travel intensity within a city (TC), and their interactive terms) than T1, and their effect shows distinct spatial heterogeneity. Interestingly, the dividing line of positive-negative effect of MeanT and Pre on COVID-19 incidence is 8.5°C and 1mm, respectively. In T2, WH has weak impact, but the MS has the strongest effect. The COVID-19 incidence in T2 without quarantine is also modeled using the developed GWR model, and the modeled incidence shows an obvious increase for 75.6% cities compared with reported incidence in T2 especially for some mega cities. This evidences national quarantine and traffic control take determinant role in controlling the disease spread. The study indicates that both natural environment and human factors integratedly affect the spread pattern of COVID-19 in China.
Key words: COVID-19, two-stage, environment impact, interactive effect, GWR model, city-level

Introduction

A novel Coronavirus was identified in December 2019 in Wuhan city of Hubei Province, China. Thus the disease was officially named by the World Health Organization (WHO) as COVID-19 and later its virus was identified to be a new coronavirus SARS-CoV-2 (Zhu et al., 2020). Till now, no medical cure is available for this disease, with absence of both preventive vaccine and specific drug (Liu et al., 2020; Shi et al., 2020). Since mid January 2020, it has rapidly spread throughout China and 835 cases were reported on 23 January 2020 (China, 2020). Thus a shutdown of Wuhan city was started on that day and a travel ban from Hubei Province was enforced on the following day. Meantime, China also declared it as level 1 national emergency response, defined as an “extremely serious incident” (Chinadaily, 2020), thus a national quarantine was since promoted. Since January 2020, it was also spread worldwide and declared as an international public health emergency by the WHO on January 30 2020 (Lai et al., 2020). Till 19 October 2020, nearly 40.13 million cases have been reported globally, with the newly confirmed case of 389683 on the day (WHO, 2020).

People have struggled to control the transmission of COVID-19 worldwide, thus global researchers have conducted many studies in order to explain its spatial-temporal distribution pattern. Firstly, some studies have investigated the impact of climate factors on global (Bannister-Tyrrell et al., 2020; Ficetola and Rubolini, 2020; Sobral et al., 2020), national (Bariotakis et al., 2020; Prata et al., 2020; Qi et al., 2020; Shahzad et al., 2020; Shi et al., 2020; Xu et al., 2020) and municipal
COVID-19 pandemic, respectively. An early global study based on cases from the first imported case until 29 February 2020 showed preliminary evidence that higher temperature was strongly associated with lower COVID-19 incidence for temperature of 1°C and higher (Bannister-Tyrrell et al., 2020). A recent country-level study incorporated more bioclimatic factors and determined two major predictive variables, including minimum temperature of the coldest month (27.4%) and mean temperature of the wettest quarter (20.9%), thereupon potential distribution of global SARS-CoV-2 infection was predicted for April 2020 (Bariotakis et al., 2020). The impact of climate factors on COVID-19 incidence in different countries is respectively determined as temperature and absolute humidity in China (Shi et al., 2020), the interactive effect between daily temperature and relative humidity in mainland China (Qi et al., 2020), air quality in 33 locations of China (Xu et al., 2020), temperature in ten provinces in China (Shahzad et al., 2020), and temperature and humidity in the US (Gupta et al., 2020). The association between daily mortality of COVID-19 and meteorological factors and air pollutant index in Wuhan city indicated that diurnal temperature range \(r=0.44\) was positively and relative humidity \(r=-0.32\) was negatively associated with COVID-19 mortality (Ma et al., 2020). The correlation between temperature and COVID-19 incidence in New York and Jakarta, Indonesia was also be explored (Bashir et al., 2020; Tosepu et al., 2020). However, the effect of temperature on COVID-19 transmissions is debating while predicting the spread of the disease in certain warm countries (Shakil et al., 2020).

Secondly, the effects of human and control measures on COVID-19 spread have also been investigated. A study showed that the international travel and border control during the early stages of the epidemic could reduce the rate of case exportations, but
it could not fully arrest the global expansion of COVID-19 (Wells et al., 2020). Another study conceived that social lockdown was the only preventive measure against COVID-19 (Paital, 2020).

According to review on the published literatures, the impact of natural environment on COVID-19 incidence is mainly studied from aspect of climate factors. Presently, an integrated assessment of both natural and human environments on the disease transmission in China is lacking. Especially, the interactive effect between natural and human environmental factors on the COVID-19 incidence is never concerned. Therefore, this study will conduct a city-level study in China to investigate the correlation between both natural and human environment and COVID-19 spread, then model the disease transmission in two stages, and finally to evaluate the role of environment in the disease spread. Besides, modeling and prediction of disease incidence with absence of quarantine is conducted to quantitatively reflect the effectiveness of such measures.

**Methods**

**Study data**

In the study, city-level data were collected among 341 cities of 31 provinces in the mainland China (Figure 1), between Jan 20 and Feb 29, 2020. The collected data mainly include reported COVID-19 case, natural environment (climate factors and terrain indicator) and human environment factors. Daily data on the number of new confirmed COVID-19 cases was obtained from the DX Doctor (http://ncov.dxy.cn), which is a collection of data reported by 32 provincial Health Commissions in China. The accumulated confirmed COVID-19 cases at a city level were shown in Figure 1. The daily climate data, including minimum temperature (MinT), maximum
temperature (MaxT), average temperature (MeanT), precipitation (Pre), relative humidity (Rh), wind velocity (Wind) and air pressure (AirP), was downloaded from World Weather (https://en.tutiempo.net/). The terrain data, including minimum DEM (Min DEM), maximum DEM (Max DEM), and mean DEM (Mean DEM) was downloaded from Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences. (http://www.gscloud.cn). Among the human factors, GDP per capita (GDP), and population density (PD) was downloaded from Wind Economic Database (https://www.wind.com.cn); destination proportion in population flow from Wuhan (WH), migration scale (MS), and travel intensity within a city (TC) were downloaded from Baidu (http://qianxi.baidu.com/). The data type, variables and data source are listed in Table 1.

The lifetime of the COVID-19 in a city in China is characterized by two-stage process, uncontrolled infection in early times and decaying stage at later times once quarantines are being performed (Gross et al., 2020). Therefore, according to the date when the Wuhan was shutdown, we divided the study period into two stages: T1 (Jan 20 and Jan 24, 2020) and T2 (Jan 25 and Feb 29, 2020). Based on the reported cases during T1 and T2, 115 and 29 cities were included in the analysis, respectively.

**Statistical analysis**

Our study is performed following the order of factor determining, modeling and prediction (Figure 2). Firstly, after a correlation analysis and multicollinearity test, correlation factors were selected; then the Ordinary Least Squares (OLS) and Geodetector were used to finally determine the effective factors and interaction terms. Secondly, a Geographically Weighted Regression (GWR) model was developed and validated to quantify the effect of environment on COVID-19 incidence. Thirdly, migration scale data in 2019 was input the model to predict the disease incidence
without control. All variables were first standardized and then put into the model.

**Geodetector analysis**

A Geodetector is a method to detect spatial stratified heterogeneity and determine the factors that drive it (Wang et al., 2010). The theoretical basis of the method is that if the spatial variability of an independent variable Y is caused by a specific factor X, the similarity should exist between spatial distributions of X and Y. The Geodetector is applied to calculate the contribution of each factor to COVID-19 incidence and to detect synergies between factors with respect to COVID-19 incidence. The Geodetector determines the spatial correlation of factor X and Y based on the q-statistic calculated with the following equation:

\[
q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N \sigma^2}
\]

(1)

where N is the number of samples in the study area, L is the number of categories of factor X, \( \sigma^2 \) is the total variance of Y in the study area, \( \sigma_h^2 \) is the variance of Y within category h of factor X. The larger the q value is, the stronger the factor X explains Y.

The Geodetector includes four main detectors, and we mainly consider the interaction detectors (Yin et al., 2019) to determine the interaction terms of factors on the COVID-19 incidence. In this study, the notation * refers to the product of two variables, which is the interaction term. For example, \( Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \times X_2 \) means that the effect of \( X_1 \) on Y depends on the value of \( X_2 \). If an interaction exists in the data, interaction terms can provide a better description of the relationship between the independent and dependent variables, and offer a more accurate estimation of the relationship and explain more of the variation in the dependent variable.
Spatial temporal analysis

To investigate the impact of the effective factors and interaction terms on the COVID-19 incidence, both OLS and GWR models were applied. As a global regression model, OLS captures the average strength and significance of the independent variables (Delmelle et al., 2016) and thus derives one-size-fits-all outcomes. In other words, it analyzes the impact of the effective factors on the COVID-19 incidence for the whole China. However, it fails to consider local variations of the effective factors for each city (Sumanasinghe et al., 2016). Global Moran’s I statistic for the COVID-19 incidence was conducted to measure spatial autocorrelation, in order to test whether the COVID-19 incidence is stationary across the China. By contrast, GWR is a local regression model, captures spatially relationships between the dependent variable and independent variables, which is different in varying locations (Zhao et al., 2020). Actually, for each location, GWR bases data from adjacent neighborhoods to conduct a local regression, thus estimating local regression coefficients for each of the predictor variables (Delmelle et al., 2016). Using COVID-19 incidence as the dependent variable, the GWR model can be expressed as follows:

\[ Y_i = \beta_0(i) + \sum_{k=1}^{N} \beta_k(i)X_k(i) + \epsilon_i \]  

(2)

where \( Y_i \) is the dependent variable, \( X_k(i) \) is the \( k \)-th independent variable \((k = 1...N)\), \( \beta_0(i) \) is the intercept at neighborhood \( i \), \( \beta_k(i) \) is the coefficient of \( X_k(i) \), \( \epsilon_i \) is the residual of the location \( i \). The adaptive bi-square kernel function and cross-validation were used to determine the ideal number of neighbors.

In our study, we used OLS forward stepwise to select the variables entering the regression model. Following this approach, variables are added sequentially until the coefficient of determination \((R^2)\) does not increase significantly anymore. Dominant
variables are used in the GWR model. We compare regression results to the ones obtained using an OLS approach.

Results

Factor identification and model development

Factor identification

Factor is determined based on correlation analysis, multicollinearity test, Geodetector analysis, and OLS analysis (Figure 2). The correlation between COVID-19 incidence and all the factors (Table 1) are listed in Table 2. For the T1 stage, most factors are correlated with COVID-19 incidence, except for Pre, Rh, Wind and Max DEM; and MaxT, MeanT, MinT, Min DEM, GDP, PD, WH, MS and TC are significantly correlated with COVID-19 incidence (p<0.01). While for the T2 stage, all the factors are significantly correlated with COVID-19 incidence, except Wind. After a multicollinearity analysis, some linearly correlated factors are excluded, such as AirP. Then for T1, two natural factors (MeanT, Mean DEM) and five human factors (GDP, PD, WH, MS, TC) are left as dominant factors; while for T2, more natural factors (MeanT, Mean DEM, Pre, Rh) and the same human factors (GDP, PD, WH, MS, TC) are dominant ones as the T1. All variance inflation factor values are ranged between 1.05-2.80 and 1.14-2.17 respectively for T1 and T2, indicating there is no multicollinearity.

The interactive effect of the above dominant factors on the COVID-19 incidence is analyzed using Geodetector (Figure 3). Clearly, the interactions of most pairs of natural factors and human factors were nonlinearly enhanced. The interaction terms with q>0.6 (in blue box in Figure 3) are then input OLS models to do final selection (Figure 2), such as WH*MeanT in T1 and MS*TC in T2. After the OLS
analysis, the effective factors and interaction terms are finally determined (Table 3). For T1, the effective factors include MeanT, WH, and MS, and the effective interaction term is WH*MeanT. While for T2, the effective factors are MeanT, Pre, Rh, Mean DEM, WH, MS, and TC, and the effective interaction terms include WH*Pre and MS*TC. All the determined factors and interaction terms are used to develop the model.

**Model development and validation**

For choosing proper model, the spatial autocorrelation analysis is conducted for the COVID-19 incidence itself. The calculated global Moran’s index for ln (case) in the study period is 0.45, indicating high spatial aggregation. This implies that GWR model is proper since GWR can capture spatially non-stationary relationships between the dependent variable and predictor variables by incorporating geographical information (Zhao et al., 2020). Then, using ln (case) as the dependent variable and the determined effective factors and effective interaction terms (Table 3) respectively as the independent variables, we developed both the GWR models and OLS models. To validate the accuracy of the developed models, we calculated the R², the adjusted R², and the Akaike Information Criterion (AICc) (Li et al., 2019; Zhao et al., 2020) for the models. First, the developed GWR models are compared to the OLS regression models (Table 4). Obviously, for both T1 and T2, the GWR models have higher values of R², adjusted R² and lower values of AICc than those of the OLS models, regardless of the model developed with the effective factors or the effective terms (Table 4). This comparison fully indicates the GWR model, with higher accuracy, is significantly more appropriate than the OLS model. Second, for both OLS and GWR models, the developed models with both the effective factors and interaction terms are much better than those meantime only with effective factors; the former models have
the higher $R^2$ and the adjusted $R^2$ and smaller values of AICc (Table 4). This implied that the interaction of different factors could greatly promote COVID-19 incidence in comparison with single factor, whenever and wherever.

**GWR analysis of environment impact on COVID-19 incidence**

**GWR analysis of environment impact on COVID-19 incidence in T1**

The spatial distribution of local coefficient ($\beta_k$) for the four explanatory variables (Table 3) in the final GWR model for T1 (Table 4) is mapped in Figure 4. In general, the effect of the three factors on the COVID-19 incidence is similarly promoting in T1. Under the average destination proportion in population flow from Wuhan (WH), MeanT has the strongest positive correlation with the COVID-19 incidence, with the values of coefficient ranging 0.80~1.04. In other words, the hotter, the higher disease incidence occurs. While under average temperature level, WH is also positively correlated with the COVID-19 incidence, with the values of coefficient ranging 0.48~0.59. The migration scale (MS) is also positively correlated with the disease incidence, and its effect degree weakens gradually from north to the south. The interaction between WH and MeanT is the strongest, with the values of coefficient ranging 1.82~2.59. This implies that the linkage effect of WH and MeanT is very strong in this stage, and it is necessary to analyze the impact of two factors on the COVID-19 incidence at the same time. Under the same WH (MeanT) level, the higher COVID-19 incidence tends to occur in hotter (higher-WH) areas. The results indicate during early outbreak of the COVID-19, WH and MeanT are key determinants promoting disease transmission. Therefore, shutdown of Wuhan city is important and effective at the time.

**GWR analysis of environment impact on COVID-19 incidence in T2**

The spatial distribution of local coefficient for the final explanatory variables
(Table 3) used in the GWR model in T2 (Table 4) is mapped in Figure 5a. In comparison with T1, more (including Rh, Mean DEM, Pre, TC, and MS*TC) and different (WH*Pre) factors have significant effect on the COVID-19 incidence and their effect shows distinct spatial heterogeneity in T2; that is, the same factor has different impact on the disease in different areas. And the heterogeneity of the effect of natural factors is stronger than that of human factors. Even though with a few exception, the effect of temperature on COVID-19 incidence is spatially divided by the temperature about 8.5°C (Fig. 5b). Specifically, in areas with MeanT < 8.5°C, temperature is positively correlated with the COVID-19 incidence, and its effect degree strengthens gradually from west to the east; but in areas with MeanT > 8.5°C, this effect is negative and strengthens gradually from north to the south. Under the average WH level, there is obvious spatial heterogeneity in the impact of precipitation on the COVID-19 incidence, with a positive relationship in the west and a negative relationship in the east and northeast (Fig. 5c). Except for the northeast region, this positive-negative relationship is approximately divided by the precipitation equal to 1mm. Most areas with Pre < 1mm have a positive relationship, and its effect degree weakens gradually from west to the east. The positive relationship is strongest in Xinjiang province. Most areas with Pre > 1mm have a negative relationship, but this effect has no obvious spatial heterogeneity.

In T2, human factors generally have more significant effect on the COVID-19 incidence than natural factors. The MS has the strongest positive correlation with COVID-19 incidence; while its correlation with WH is not so strong. Because Wuhan has been shut down in this stage, and the MS has become a determinant promoting disease transmission. The effect of Mean DEM on the COVID-19 incidence is the weakest, and the relationship is mostly negative except in a few areas. For example, in
rugged areas within the “third topography ladder” of China, such as Fujian Province, the positive relationship is relatively strong.

The interactive effect of environmental factors on the COVID-19 incidence is far more complex than in T1. The interaction between WH and Pre is always negative (Fig. 5a), which means that one factor will reduce the effect of the other on the COVID-19 incidence. In other words, under the same WH (Pre) level, the less precipitation (WH), the higher disease incidence occurs. The interaction between MS and TC is generally positive (Fig. 5a), with some exception in areas such as Akesu and Yili in XinJiang Province. Overall, under the same MS (TC) level, the higher TC (MS), the higher disease incidence occurs.

The above two-stage modeling not only examines the different impact of environment on COVID-19 incidence in early and late period, but also quantitatively reflects the inhibition effect of shutdown of Wuhan city and following national quarantine.

**Modeling and risk analysis of COVID-19 in uncontrolled T2**

According to the above results, MS is the most important factor affecting COVID-19 spread for T2 (Figure 5). Therefore, a scenario is defined as “uncontrolled T2” to represent T2 with the meantime MS situation in 2019. Then MS data in 2019 is input as an alternative independent variable, and we model the COVID-19 incidence based on the developed model for T2 (Table 1). The spatial distribution of the reported incidence in T2 (Fig.6a) is compared with that of the modeled incidence under this scenario (Fig.6b). The difference between them (Fig.6c) indicates the changed MS alone could cause an increase in the COVID-19 incidence in 75.6% of cities nationwide. The modeled COVID-19 incidence has increased in almost all provincially capital cities (Fig.6b). Such increase is particularly prominent in some
mega cities, such as Beijing, Shanghai, Guangzhou and Chengdu, where booming outbreak of the COVID-19 is likely to occur as predicted (Fig.6c). This evidences the effectiveness of national quarantine and traffic control in inhibiting the disease spread.

In contrast, the predicted disease incidence is decreased in some areas, such as Wenzhou and cities around Wuhan. The population movement between Wenzhou and Wuhan is very frequent. According to the Wenzhou city government report on Jan 29, 2020, there are about 180,000 Wenzhou people doing business, working and studying in Wuhan. And a total of 433,000 people have been identified to return to Wenzhou from Wuhan and surrounding key areas (Yang et al., 2020). Similar phenomena exist in cities around Wuhan. These all imply the dominant factor for these cities is WH instead of MS, so the modeled incidence may not be a very accurate estimate.

Discussion

Research implication

This study has implications for the two aspects. On the one hand, climate change is well known to have significant effect on infectious diseases (Li et al., 2018; Wu et al., 2020; Wu et al., 2016). Our study indicates that temperature, precipitation and relative humidity are three important climate indicators for COVID-19 incidence. The three factors are also found effective on the other coronavirus transmitted diseases, including SARS-CoV (Tan et al., 2005) and MERS (Alghamdi et al., 2014). Furthermore, temperature is usually a powerful indicator for these diseases (Tan et al., 2005) and the impact of temperature is proven to vary within different temperature ranges. Therefore, even though the temperature and humidity have significant effect on COVID-19 in China (Qi et al., 2020), Brazil (Auler et al., 2020), and the US (Gupta et al., 2020), the effect of temperature on COVID-19 is different among provinces (Qi et al., 2020; Shahzad et al., 2020) and cities in China. The improvement
of our study lies in that it not only indicates that the effect of climate factors on COVID-19 has strong spatial heterogeneity, but also determines the dividing line of positive-negative effect of MeanT and Pre on COVID-19 incidence is 8.5°C and 1 mm, respectively. A study in mainland China finds that there is a potential interactive effect between daily temperature and relative humidity on COVID-19 incidence (Qi et al., 2020). Another study in 33 locations in China finds that the degree of the effect of air quality on COVID-19 incidence is related to climate factors, when the temperature of 10-20 °C or the relative humidity of 10-20%, the air quality has a stronger effect (Xu et al., 2020). The improvement of our study lies in that the former studies only consider the interaction between climate factors, while ours consider interactions between all factors and quantified these interactions. Furthermore, we find interaction between natural and human factors is stronger than the interaction between natural factors.

On the other hand, human activity has been reported as an important driver for the transmission of infectious diseases (Wu et al., 2014). Our study infers that destination proportion in population flow from Wuhan, travel intensity within a city, and migration scale were three contributing factors for the COVID-19. It is evidenced that the transportation network of highways and railways (Fang et al., 2009) and distribution of ring road (Wang et al., 2008; Wang et al., 2006) are important in spreading an SARS epidemic. This indicates traffic transportation is crucial to instant spread of emerging coronavirus diseases. Prior studies on COVID-19 find that lockdown of Wuhan has a great efficacy, the population outflow from Wuhan has a strong correlation with the confirmed cases in each county (Jia et al., 2020), and measures taken to close schools and workplaces are conducive to controlling the outbreak of COVID-19 in Wuhan (Prem et al., 2020). The selected five human factors
in this study not only include those corresponding to the indicators in previous studies, but also consider more aspects than the previous studies. For example, “GDP” in our study contains both “the number of doctors” (Qiu et al., 2020) and “the number of companies” (Briz-Redón and Serrano-Aroca, 2020); and our “MS” is highly correlated with “the number of travelers”. Besides, the human factors we selected are more comprehensive and representative. For example, WH, MS, and TC can reflect the human mobility in key disease outbreak cities, between different cities, and within cities, respectively. A great significance lies in that our study find that the interactive effect of two human factors is much more significant than the single factor. We believe that the interaction between migration scale and travel intensity within a city is typical in reflecting human mobility, and thus both is determinant in the disease transmission in T2. Therefore effective preventive control on human mobility should be promoted in other pandemic areas.

**Research contribution**

This study has advanced our understanding of the COVID-19 from the four aspects. Firstly, this is a systematic study on integrated assessment of environment on COVID-19 in China. Natural and human environment is geographically correlated, which is rarely detected. Compared with single-perspective environmental study, such a comprehensive evaluation could truly reflect the underlying environment driving of the newly emerging disease. This could also shed light on the impact of environment on other diseases. Secondly, the interactive effect of natural and human environmental factors has been innovatively addressed. This investigation contributes to explaining the spatial-temporal spread pattern of COVID-19 in China. Thirdly, this is a novel spatial modeling of COVID-19 incidence in China using GWR model. Environment and disease distribution are spatially correlated in essence and our model
quantitatively depicts this. Fourthly, we have conducted two-stage investigation on impact of environment on COVID-19 incidence at city level, considering the transmission and virulence of the SARS-CoV-2 virus varies in different conditions (Shi et al., 2020). The stronger interactive effect of natural and human environment in some zone has further proven and evidenced this.

**Conclusion**

In this study, we investigate the effect of both natural and human environment on COVID-19 pattern at a city level, based on the two-stage modeling using GWR models. The study finds four effective factors (MeanT, WH, MS) and interaction terms (WH*MeanT) are generally promoting for COVID-19 incidence before Wuhan’s shutdown (T1), and the WH and MeanT play a determinant role in the disease spread in T1. The effect of environment on COVID-19 incidence after Wuhan’s shutdown (T2) includes more factors than T1, and their effect shows distinct spatial heterogeneity. The impact is spatially divided by temperature about 8.5°C and precipitation of 1mm. In T2, human factors have more impact than the natural ones. WH has weak impact than in T1, but the MS has the strongest effect. The COVID-19 incidence in T2 without quarantine is also modeled using the developed GWR model, and the modeled incidence show an obvious increase compared with reported incidence in T2 especially for some mega cities. The results indicate that both natural environment and human factors integratedly affect the spread pattern of COVID-19 in China, and national quarantine and traffic control take determinant role in controlling the disease spread. Therefore, our future research will investigate the effect of different social and traffic control measures on controlling the COVID-19 spread, so as to provide a reference for other pandemic countries.
Acknowledgements

This work was supported by the National Key Research and Development Program of China (2016YFA0600104).
References

Alghamdi IG, Hussain II, Almalki SS, Alghamdi MS, Alghamdi MM, El-Sheemy MA. 2014. The pattern of Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive epidemiological analysis of data from the Saudi Ministry of Health. Int J. Gen. Med. 7, 417-423. https://doi.org/10.2147/IJGM.S67061.

Auler AC, Cássaro FAM, da Silva VO, Pires LF. 2020. Evidence that high temperatures and intermediate relative humidity might favor the spread of COVID-19 in tropical climate: A case study for the most affected Brazilian cities. Sci. Total Environ. 729, 139090. https://doi.org/10.1016/j.scitotenv.2020.139090.

Bannister-Tyrrell M, Meyer A, Faverjon C, Cameron A. 2020. Preliminary evidence that higher temperatures are associated with lower incidence of COVID-19, for cases reported globally up to 29th February 2020. medRxiv, 2020.03.18.20036731. https://doi.org/10.1101/2020.03.18.20036731.

Bariotakis M, Sourvinos G, Castanas E, Pirintsos SA. 2020. Climatic influences on the worldwide spread of SARS-CoV-2. medRxiv, 2020.03.19.20038158. https://doi.org/10.1101/2020.03.19.20038158.

Bashir MF, Ma B, Bial, Komal B, Bashir MA, Tan D, et al., 2020. Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci. Total Environ. 728, 138835. https://doi.org/10.1016/j.scitotenv.2020.138835.

Briz-Redón Á, Serrano-Aroca Á. 2020. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. Sci. Total Environ. 728, 138811. https://doi.org/10.1016/j.scitotenv.2020.138811.

China NHCoPrsRo. 2020. Novel Coronavirus infected pneumonia situation 24 January.
Delmelle E, Hagenlocher M, Kienberger S, Casas I. 2016. A spatial model of socioeconomic and environmental determinants of dengue fever in Cali, Colombia. Acta Trop. 164, 169-176. https://doi.org/10.1016/j.actatropica.2016.08.028.

Fang L-Q, de Vlas SJ, Feng D, Liang S, Xu Y-F, Zhou J-P, et al., 2009. Geographical spread of SARS in mainland China. Trop. Med. Int. Health 14, 14-20. https://doi.org/10.1111/j.1365-3156.2008.02189.x.

Ficetola GF, Rubolini D. 2020. Climate affects global patterns of COVID-19 early outbreak dynamics. medRxiv, 2020.03.23.20040501. https://doi.org/10.1101/2020.03.23.20040501.

Gross B, Zheng Z, Liu S, Chen X, Sela A, Li J, et al., 2020. Spatio-temporal propagation of COVID-19 pandemics. medRxiv, 2020.03.23.20041517. https://doi.org/10.1101/2020.03.23.20041517.

Gupta S, Raghuwanshi GS, Chanda A. 2020. Effect of weather on COVID-19 spread in the US: A prediction model for India in 2020. Sci. Total Environ. 728, 138860. https://doi.org/10.1016/j.scitotenv.2020.138860.

Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. 2020. Population flow drives spatio-temporal distribution of COVID-19 in China. Nature 582, 389–394. https://doi.org/10.1038/s41586-020-2284-y.
Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int. J. Antimicrob. Agents. 55, 105924. https://doi.org/10.1016/j.ijantimicag.2020.105924.

Li C, Lu Y, Liu J, Wu X. 2018. Climate change and dengue fever transmission in China: Evidences and challenges. Sci. Total Environ. 622-623, 493-501. https://doi.org/10.1016/j.scitotenv.2017.11.326.

Li Y, Yang L, Shen H, Wu Z. 2019. Modeling intra-destination travel behavior of tourists through spatio-temporal analysis. J. Dest. Mark. Manage. 11, 260-269. https://doi.org/10.1016/j.jdmm.2018.05.002.

Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, et al., 2020. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. ACS Cent. Sci. 6, 315-331. https://doi.org/10.1021/acscentsci.0c00272.

Ma Y, Zhao Y, Liu J, He X, Wang B, Fu S, et al., 2020. Effects of temperature variation and humidity on the mortality of COVID-19 in Wuhan. medRxiv, 2020.03.15.20036426. https://doi.org/10.1101/2020.03.15.20036426.

Paital B. 2020. Nurture to nature via COVID-19, a self-regenerating environmental strategy of environment in global context. Sci. Total Environ. 729, 139088. https://doi.org/10.1016/j.scitotenv.2020.139088.

Prata DN, Rodrigues W, Bermejo PH. 2020. Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil. Sci. Total Environ. 729, 138862. https://doi.org/10.1016/j.scitotenv.2020.138862.

Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, et al., 2020. The effect of control strategies to reduce social mixing on outcomes of the
COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health 5, e261-e270. https://doi.org/10.1016/S2468-2667(20)30073-6.

Qi H, Xiao S, Shi R, Ward MP, Chen Y, Tu W, et al., 2020. COVID-19 transmission in Mainland China is associated with temperature and humidity: A time-series analysis. Sci. Total Environ. 728, 138778. https://doi.org/10.1016/j.scitotenv.2020.138778.

Qiu Y, Chen X, Shi W. 2020. Impacts of social and economic factors on the transmission of coronavirus disease (COVID-19) in China. medRxiv, 2020.03.13.20035238. https://doi.org/10.1101/2020.03.13.20035238.

Shahzad F, Shahzad U, Fareed Z, Iqbal N, Hashmi S, Ahmad F. 2020. Asymmetric nexus between temperature and COVID-19 in the top ten affected provinces of China: A current application of quantile-on-quantile approach. Sci. Total Environ. 736, 139115. https://doi.org/10.1016/j.scitotenv.2020.139115.

Shi P, Dong Y, Yan H, Li X, Zhao C, Liu W, et al., 2020. The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak - evidence from China. medRxiv, 2020.03.22.20038919. https://doi.org/10.1101/2020.03.22.20038919.

Sobral MFF, Duarte GB, da Penha Sobral AIG, Marinho MLM, de Souza Melo A. 2020. Association between climate variables and global transmission of SARS-CoV-2. Sci. Total Environ. 729, 138997. https://doi.org/10.1016/j.scitotenv.2020.138997.

Sumanasinghe N, Mikler A, Tiwari C, Muthukudage J. 2016. Geo-statistical dengue risk model using GIS techniques to identify the risk prone areas by linking rainfall and population density factors in Sri Lanka. Ceylon Journal of Science 45, 39. https://doi.org/10.4038/cjs.v45i3.7399.
Tan J, Mu L, Huang J, Yu S, Chen B, Yin J. 2005. An initial investigation of the association between the SARS outbreak and weather: with the view of the environmental temperature and its variation. J. Epidemiology Community Health 59, 186. https://doi.org/10.1136/jech.2004.020180.

Tosepu R, Gunawan J, Effendy DS, Ahmad LOAI, Lestari H, Bahar H, et al., 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. Sci. Total Environ. 725, 138436. https://doi.org/10.1016/j.scitotenv.2020.138436.

Wang J-F, Christakos G, Han W-G, Meng B. 2008. Data-driven exploration of 'spatial pattern-time process-driving forces' associations of SARS epidemic in Beijing, China. J Public Health (Oxf) 30, 234-244. https://doi.org/10.1093/pubmed/fdn023.

Wang J, McMichael AJ, Meng B, Ecobichon NG, Han W, Glass K, et al., 2006. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. Bull World Health Organ. 84, 965-968. https://doi.org/10.2471/blt.06.030247.

Wang JF, Li XH, Christakos G, Liao YL, Zhang T, Gu X, et al., 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. Int. J. Geogr. Inf. Sci. 24, 107-127. https://doi.org/10.1080/13658810802443457.

Wells CR, Sah P, Moghadas SM, Pandey A, Shoukat A, Wang Y, et al., 2020. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. PNAS 117, 7504. https://doi.org/10.1073/pnas.2002616117.

WHO. 2020. WHO Coronavirus Disease (COVID-19) Dashboard.
https://covid19.who.int/ (Accessed 19 October 2020).

Wu X, Liu J, Li C, Yin J. 2020. Impact of climate change on dysentery: Scientific evidences, uncertainty, modeling and projections. Sci. Total Environ. 714, 136702. https://doi.org/10.1016/j.scitotenv.2020.136702.

Wu X, Lu Y, Zhou S, Chen L, Xu B. 2016. Impact of climate change on human infectious diseases: Empirical evidence and human adaptation. Environ. Int. 86, 14-23. https://doi.org/10.1016/j.envint.2015.09.007.

Wu X, Tian H, Zhou S, Chen L, Xu B. 2014. Impact of global change on transmission of human infectious diseases. Sci. China Earth Sci. 57, 189-203. https://doi.org/10.1007/s11430-013-4635-0.

Xu H, Yan C, Fu Q, Xiao K, Yu Y, Han D, et al., 2020. Possible environmental effects on the spread of COVID-19 in China. Sci. Total Environ. 731, 139211. https://doi.org/10.1016/j.scitotenv.2020.139211.

Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A, et al., 2020. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): A multi-center study in Wenzhou city, Zhejiang, China. J. Infection. 80, 388-393. https://doi.org/10.1016/j.jinf.2020.02.016.

Yin J, Wu X, Shen M, Zhang X, Zhu C, Xiang H, et al., 2019. Impact of urban greenspace spatial pattern on land surface temperature: a case study in Beijing metropolitan area, China. Landscape Ecology 34, 2949-2961. https://doi.org/10.1007/s10980-019-00932-6.

Zhao R, Zhan L, Mingxing Y, Yang L. 2020. A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5. Sustain. Cities Soc. 56, 102106.
Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727-733.

https://doi.org/10.1056/NEJMo2001017.
Fig. 1 Distribution of accumulated COVID-19 cases (till Feb 29, 2020) at city level in mainland China.

Fig. 2 Flowchart of data processing in the study.

Fig. 3 The interactive effect of dominant factors on COVID-19 incidence, where the blue box are the interaction terms with q>0.6.

Fig. 4 The spatial distribution of the local coefficient (β_k) for the four variables used in the GWR model for T1.

Fig. 5 The spatial distribution of the local coefficient (β_k) for the nine variables used in the GWR model for T2. (a) All the factors; (b) MeanT divided by the temperature about 8.5°C; (c) Pre divided by the precipitation equal to 1mm.

Fig. 6 Risk analysis of COVID-19 incidence based on T2 situation: (a) Reported incidence in T2; (b) Modeled incidence using GWR model in T2 with 2019 migration scale; (c) The difference between the modeled and reported incidence.
### Tables

**Table 1** Data type, variables and data source

| Data type               | Variable                                      | Scale | Data source                             |
|-------------------------|-----------------------------------------------|-------|-----------------------------------------|
| Disease                 | COVID-19 data                                 |       |                                         |
|                         | Confirmed case                                | Daily | [http://ncov.dxy.cn](http://ncov.dxy.cn) |
| Natural environment     | Climate data                                  |       |                                         |
|                         | Minimum Temperature                           | Daily | [https://en.tutiempo.net/](https://en.tutiempo.net/) |
|                         | Maximum Temperature                           | Daily |                                         |
|                         | Mean Temperature                              | Daily |                                         |
|                         | Precipitation                                 | Daily |                                         |
|                         | Relative humidity                             | Daily |                                         |
|                         | Wind speed                                    | Daily |                                         |
|                         | Air pressure                                  | Daily |                                         |
| Terrain data            | Minimum DEM                                   | Yearly| [http://www.gscloud.cn](http://www.gscloud.cn) |
|                         | Maximum DEM                                   | Yearly|                                         |
|                         | Mean DEM                                      | Yearly|                                         |
| Human environment       | Economic data                                 |       |                                         |
|                         | GDP per capita                                | Yearly| [https://www.wind.com.cn](https://www.wind.com.cn) |
| Population data         | Population density                            | Yearly|                                         |
|                         | Destination proportion in population flow from Wuhan | Daily|                                         |
| Human mobility          | Travel intensity within a city                | Daily | [http://qianxi.baidu.com](http://qianxi.baidu.com) |
|                         | Migration scale                               | Daily |                                         |
Table 2 Pearson correlations between COVID-19 incidence and environmental factors

| Natural factor | T1         | T2         | Human factor | T1         | T2         |
|----------------|------------|------------|--------------|------------|------------|
| MinT           | 0.326**    | 0.298**    | GDP          | 0.338*     | 0.301**    |
| MaxT           | 0.282**    | 0.211**    | PD           | 0.518**    | 0.438**    |
| MeanT          | 0.313**    | 0.268**    | WH           | 0.413**    | 0.445**    |
| Pre            | 0.010      | 0.300*     | MS           | 0.574**    | 0.489**    |
| Rh             | 0.127      | 0.419**    | TC           | -0.371**   | -0.409**   |
| Wind           | -0.037     | 0.080      |              |            |            |
| AirP           | 0.215*     | 0.418**    |              |            |            |
| Min DEM        | -0.247**   | -0.396**   |              |            |            |
| Max DEM        | -0.073     | -0.337**   |              |            |            |
| Mean DEM       | -0.224*    | -0.414**   |              |            |            |

* Correlation is significant at the 0.05 level
** Correlation is significant at the 0.01 level

Table 3 Comparison between OLS and GWR model

| Stage | Variables                        | Model | AICc  | R²     | Adjusted R² |
|-------|----------------------------------|-------|-------|--------|-------------|
| T1    | Effective factors                | OLS   | 241.39| 0.56   | 0.55        |
|       |                                  | GWR   | 240.31| 0.58   | 0.56        |
|       | Effective factors and            | OLS   | 236.86| 0.59   | 0.57        |
|       | interaction terms                 | GWR   | 235.56| 0.61   | 0.58        |
| T2    | Effective factors                | OLS   | 557.84| 0.59   | 0.57        |
|       |                                  | GWR   | 470.58| 0.74   | 0.70        |
|       | Effective factors and            | OLS   | 517.32| 0.64   | 0.63        |
|       | interaction terms                 | GWR   | 440.23| 0.77   | 0.73        |
Credit Author Statement

Xiaoxu Wu conceived and designed the study. Jie Yin processed the data, and Xiaoxu Wu & Jie Yin analyzed the data. Chenlu Li, Hongxu Xiang, Meng Lv and Zhiyi Guo contributed to data collection. Xiaoxu Wu and Jie Yin wrote the paper.
Declaration of interests

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

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:
Highlights

- A city-level two-stage model is developed to quantify the environmental impact on COVID-19 in China.

- Destination proportion in population flow from Wuhan and mean temperature play a determinant role in the disease spread before Wuhan’s shutdown.

- After Wuhan’s shutdown (T2), human factors have more impact than the natural ones, and their effect shows distinct spatial heterogeneity.

- The dividing line of positive-negative effect of mean temperature and precipitation on COVID-19 incidence in T2 is 8.5°C and 1 mm, respectively.

- The modeled incidence shows an obvious increase compared with reported incidence in T2 especially for some mega cities.
Figure 1

Cumulative COVID-19 cases in cities before Feb 29, 2020

- Beijing
- Wuhan

Legend:
- No data
- 0
- 10
- 100
- 500
- 1000
- 2000
- 48,557

Scale:
- 0, 550, 1,100 Km
Figure 2

1. Ln(cases)
2. All factors
3. Correlation analysis
4. Multicollinearity test
5. Correlation factors
6. Geodetector
7. Interaction terms
8. Adjust
9. OLS analysis
10. Validate
11. Good
12. Effective factors and interaction terms
13. GWR models
14. 2019 MS data
15. Modelling
16. Uncontrolled T2 scenarios
Figure 3
Figure 5
Figure 6