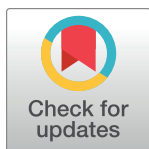


## RESEARCH ARTICLE

## Influential factors of tuberculosis in mainland China based on MGWR model

Zhipeng Ma , Hong Fan \*

State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

\* [hfan3@whu.edu.cn](mailto:hfan3@whu.edu.cn)

## OPEN ACCESS

**Citation:** Ma Z, Fan H (2023) Influential factors of tuberculosis in mainland China based on MGWR model. PLoS ONE 18(8): e0290978. <https://doi.org/10.1371/journal.pone.0290978>

**Editor:** Junyuan Yang, Shanxi University, CHINA

**Received:** April 1, 2023

**Accepted:** August 20, 2023

**Published:** August 31, 2023

**Copyright:** © 2023 Ma, Fan. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting Information](#) files.

**Funding:** Funder Name: Ministry of Science and Technology of the People's Republic of China Grant Number: 2019YFB1405600 Grant Recipient: Mr. Zhipeng Ma and Prof. Hong Fan Specific grant numbers: National Key R&D Program of China (2019YFB1405600) Initials of author who received each award: JW Full names of commercial companies that funded the study or authors: Ministry of Science and Technology of the People's Republic of China Initials of authors who received salary or other funding from commercial

## Abstract

Tuberculosis (TB), as a respiratory infectious disease, has damaged public health globally for decades, and mainland China has always been an area with high incidence of TB. Since the outbreak of COVID-19, it has seriously occupied medical resources and affected medical treatment of TB patients. Therefore, the authenticity and reliability of TB data during this period have also been questioned by many researchers. In response to this situation, this paper excludes the data from 2019 to the present, and collects the data of TB incidence in mainland China and the data of 11 influencing factors from 2014 to 2018. Using spatial autocorrelation methods and multiscale geographically weighted regression (MGWR) model to study the temporal and spatial distribution of TB incidence in mainland China and the influence of selected influencing factors on TB incidence. The experimental results show that the distribution of TB patients in mainland China shows spatial aggregation and spatial heterogeneity during this period. And the  $R^2$  and the adjusted  $R^2$  of MGWR model are 0.932 and 0.910, which are significantly better than OLS model (0.466, 0.429) and GWR model (0.836, 0.797). The fitting accuracy indicators MAE, MSE and MAPE of MGWR model reached 5.802075, 110.865107 and 0.088215 respectively, which also show that the overall fitting effect is significantly better than OLS model (19.987574, 869.181549, 0.314281) and GWR model (10.508819, 267.176741, 0.169292). Therefore, this model is based on real and reliable TB data, which provides decision-making references for the prevention and control of TB in mainland China and other countries.

## 1. Introduction

According to a report by the World Health Organization (WHO, 2012), one-third of the world's population, which is around two billion people infected with TB. As one of the top ten causes of death in the world, TB is a chronic infectious disease caused by *Mycobacterium tuberculosis* infection, which spreads mainly through respiratory transmission, and can be caused by recently infected patients or latently infected patients [1]. More than 90% of TB cases and deaths come from developing countries. In 2021, the estimated incidence of TB in China ranked third after India and Indonesia, so TB was listed as a Class-B respiratory infectious disease in China. In recent years, due to Chinese government's increased prevention and



companies: HF URL of sponsor's websites: <https://www.most.gov.cn/index.html> The funders had no role in study design, data collection and analysis, publication decisions, or manuscript preparation. And there was no additional external funding received for this study.

**Competing interests:** The authors have declared that no competing interests exist.

control efforts, the number of patients in China has been decreasing year by year. In addition, due to the large population base and the huge latent infection population in China, the number of patients in some areas has rebounded and increased in recent years. Therefore, TB is still one of the main infectious diseases that endangers the health of Chinese residents. Since the widespread spread of COVID-19, as COVID-19 has formed severe occupation of medical resources and disrupted the health care system [2], TB statistics have dropped significantly. Although some researchers believe that the temporary immunosuppressive effects and the corticosteroids used to treat COVID-19 have played a direct role in immunosuppression of TB [3], the authenticity and reliability of TB data are still questioned by many researchers.

In recent years, with the development of artificial intelligence data analysis and the improvement of data integrity, more and more researchers around the world have conducted varying degrees of research on the incidence distribution and pathogenic factors of TB. With the continuous development of the research level, the discussion of factors affecting TB incidence is no longer limited to the source of infection, the route of transmission, and the susceptible population. There is evidence to prove that meteorological factors and exposure to air pollutants have a certain impact on TB [4]. Due to transmission through the respiratory tract, TB has obvious seasonal peaks, with a larger peak in late spring (April) and a smaller peak in early autumn [5], and some studies have confirmed that long-term exposure to SO<sub>2</sub> [6] or occupational inhalation of silica dust [7] will increase the risk of men suffering from TB. Meanwhile, meteorological factors such as temperature and relative humidity may influence the risk of TB by altering the temporal and spatial distribution of air pollutants [8], resulting in spatial heterogeneity in the distribution of TB incidence in various regions of the world. TB is known as a disease reflecting socio-economic and environmental conditions [9], and regional differences in socio-economic are also one of the reasons for its spatial heterogeneity [10]. Different researchers use different models to analyze local TB incidence and socio-economic factors, and all conclude that there is a significant curvilinear relationship with socio-economic status [10, 11], which further verify the importance of socio-economic on the spread of TB. But the incidence of TB is affected by many factors, for example, TB incidence and corresponding mortality rates in S. Korea are unusual and unique compared to other economically developed countries [11]. At the same time, some researchers conduct a multilevel analysis of self-reported TB on the representative sample of South Africa, and find that TB is associated with cigarette smoking, alcohol consumption, low body mass index, lower level of personal education, lower household wealth and unemployment [12]. And other researchers conduct a retrospective analysis on routinely collected TB data in Kenya, and conclude that TB is not only related to age and gender, but also differences in nationality and job can lead to changes in the growth of TB [13].

The current incidence of TB in China also shows spatial heterogeneity [14], so systematic investigations of social and environmental factors influencing TB are necessary for the prevention and control of the disease [15]. In recent years, researchers have successively carried out research on areas with high incidence of TB in economically underdeveloped areas in western China. Some researchers applied Kulldorff's spatial-temporal scanning statistics to identify the temporal and spatial clusters of county-level TB prevalence in Yunnan, detected aggregated time interval and regions for TB at county-level of Yunnan Province, and found similarity prevalence patterns in the borders of China and the Great Mekong Subregion (GMS) region [16]. The spatial clustering of TB among students in Nanning, mainly located in the urban center and its surrounding areas, and the clustering gradually decreased from the urban center to the surrounding areas [17]. Meanwhile, some studies have shown that air quality and economic level have hysteresis with different lag time [15], and find that using "proportion of minorities" as a predictor may help to guide TB control programs and targeting interventions [18].

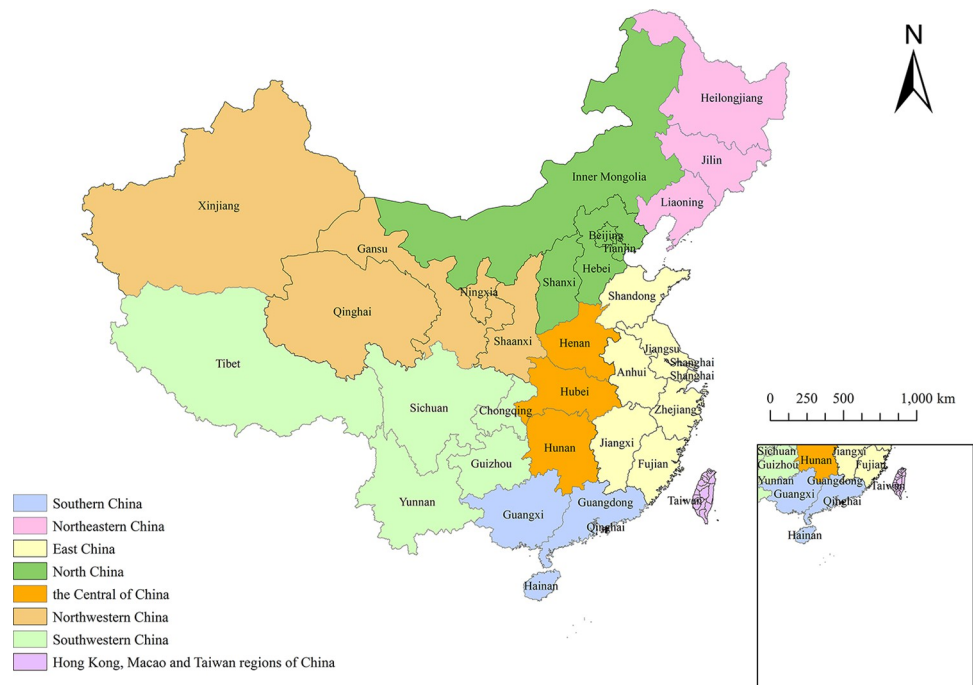


Since most infectious diseases have the characteristics of spatial aggregation and spatial heterogeneity, traditional regression models often have problems such as insufficient discussion of spatial distribution characteristics when dealing with such problems, resulting in unsatisfactory analysis results. At present, most research methods of TB are relatively single, mainly including Spatial Lag Model [1, 18], Spatial Error Model [18, 19], Multivariate Poisson Regression Model [20], etc. And the analysis of TB needs to consider spatial distribution characteristics [21, 22], so a single analysis model can't accurately analyze such problems. Therefore, this paper takes Moran's I as the analysis index to carry out the research on global and local spatial autocorrelation. On this basis, MGWR model, which is often applied to urban housing land prices [23, 24], economic development [25], medical and health field [26], but not often used in the exploration of disease influencing factors, is selected to analyze the influencing factors of TB and its influence degree. By analyzing the spatial distribution of TB and reasonably inferring the potential relationship between TB and other influencing factors, this paper provides decision-making references for the prevention and control of TB in the future.

## 2. Materials and methods

### 2.1 Research area

China, a country of East Asia with a land area of about 9.6 million square kilometers, has 34 provincial-level administrative regions, including 23 provinces, 5 autonomous regions, 4 municipalities and 2 special administrative regions. According to data released by National Bureau of Statistics of China on January 17, 2020, the total population of mainland China exceeded 1.4 billion, ranking first in the world. In order to formulate effective prevention and control measures to reduce the serious harm caused by TB, it is urgent to select real and reliable TB data and find out the factors that affect the spread of TB. The research area is 31 provinces in mainland China (municipalities and autonomous regions, excluding Hong Kong, Macao and Taiwan), as shown in Fig 1. The map comes from Natural Earth (<http://www.naturalearthdata.com/>).



**Fig 1. Locations of study areas in mainland China.**

<https://doi.org/10.1371/journal.pone.0290978.g001>

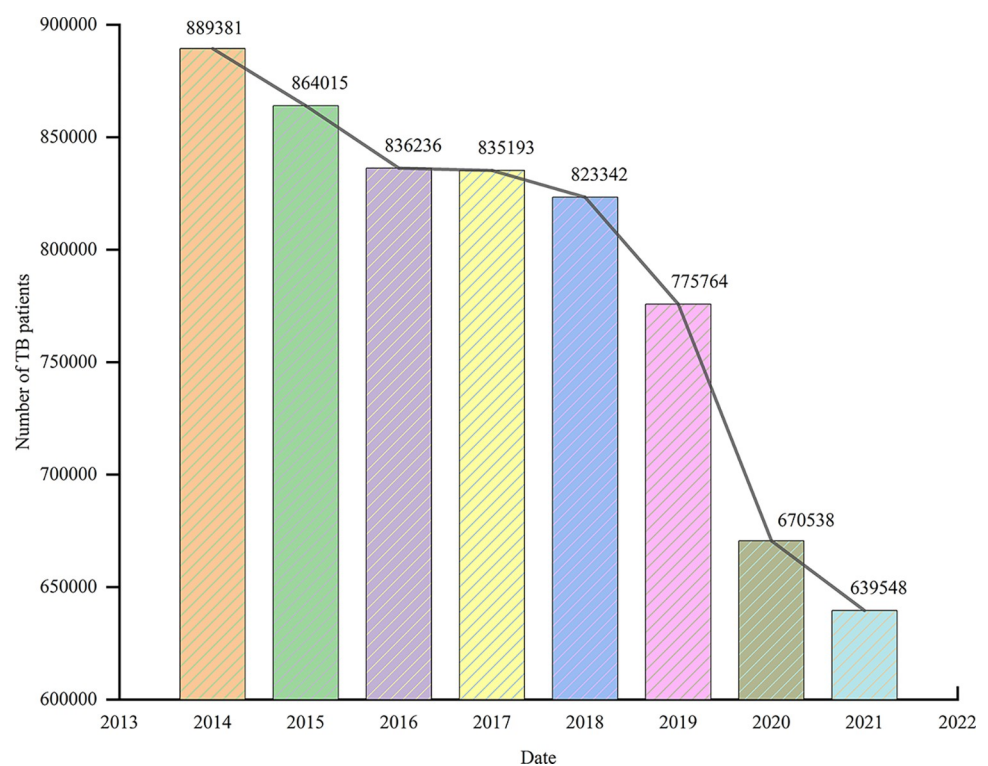


## 2.2 Data collection

Since the outbreak of COVID-19, the number of TB cases in mainland China has dropped significantly, as shown in Fig 2. One of the reasons is that mainland China was affected by COVID-19, and most of medical resources were occupied, while the control of TB was neglected. Another reason is some patients failed to go to the hospital in time, resulting in deviations in the statistics. Therefore, in order to ensure the authenticity and reliability of TB research data, this study excludes the data from 2019 to the present, analyzes the situation in the five years before the outbreak of COVID-19, and collects relevant data from different official websites. TB data of each province and city in China from January 2014 to December 2018 comes from Public Health Science Data Center (<https://www.phsciencedata.cn/>).

Because TB is an infectious disease that spreads through the respiratory tract, air quality will directly affect the incidence of TB [14, 27, 28]. In recent years, the awareness of environmental protection in various countries has continuously improved, and the importance of air quality has been repeatedly emphasized.  $PM_{2.5}$ ,  $SO_2$  and  $NO_2$ , as the main pollutants for judging the quality of air, can directly enter the lungs through respiratory tract and affect human health [55, 56, 58]. In severe cases, they may cause alveolar atrophy, pulmonary edema and other problems [59], which leads to the onset of TB. Therefore, this study selects the data of  $PM_{2.5}$ ,  $SO_2$  and  $NO_2$  from 2014 to 2018 as the factors affecting the relationship between air quality and TB.

However, TB is a disease that reflects socio-economic and environmental conditions [9], and the level of social life is also one of the important factors affecting TB. Per capita consumption expenditure, Average number of clinic visits, Per capita healthcare consumption expenditure, Per capita GDP and Passenger traffic volume are important factors to evaluate the level of social life in a region, which largely reflects the local economic level and development level.



**Fig 2. Trends in the number of TB patients in mainland China from 2014 to 2021.**

<https://doi.org/10.1371/journal.pone.0290978.g002>



Table 1. Variable list of influencing factors.

Dimension	Indicator	Source	Description
air quality	PM <sub>2.5</sub>	China Air Quality Online Monitoring and Analysis Platform ( <a href="https://www.aqistudy.cn/">https://www.aqistudy.cn/</a> )	Annual average concentration of PM <sub>2.5</sub> in mainland China from 2014 to 2018
	SO <sub>2</sub>	China Air Quality Online Monitoring and Analysis Platform ( <a href="https://www.aqistudy.cn/">https://www.aqistudy.cn/</a> )	Annual average concentration of SO <sub>2</sub> in mainland China from 2014 to 2018
	NO <sub>2</sub>	China Air Quality Online Monitoring and Analysis Platform ( <a href="https://www.aqistudy.cn/">https://www.aqistudy.cn/</a> )	Annual average concentration of NO <sub>2</sub> in mainland China from 2014 to 2018
social life	Per capita consumption expenditure	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018
	Average number of clinic visits	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018
	Per capita healthcare consumption expenditure	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018
	Per capita GDP	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018
	Passenger traffic volume	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018, including railways, water transport and civil aviation
medical level	Number of beds in medical institutions	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018, including hospitals, primary health institutions and professional public health institutions
	Number of health technicians per thousand population	China Statistical Yearbook ( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018, including urban and rural
	Number of medical institutions	China Statistical Yearbook( <a href="http://www.stats.gov.cn/tjsj/ndsj">http://www.stats.gov.cn/tjsj/ndsj</a> )	Annual data from 2014 to 2018, includes hospitals, primary health institutions and professional public health institutions

<https://doi.org/10.1371/journal.pone.0290978.t001>

Due to the differences in social life in different regions, the investment in healthcare costs during the early prevention period is also different, which has also led to a significant decline in the incidence of TB in some regions with higher social living standards [11, 14, 18, 29]. Therefore, this study selects data such as Per capita consumption expenditure, Average number of clinic visits, Per capita healthcare consumption expenditure, Per capita GDP and Passenger traffic volume from 2014 to 2018 as the influencing factors of social life and TB.

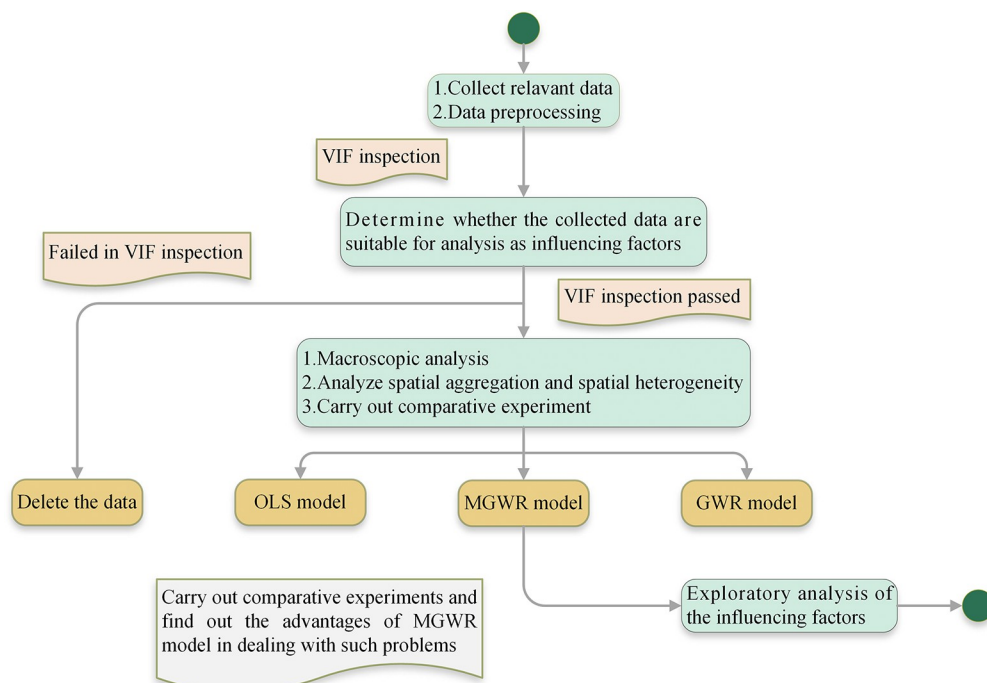
Moreover, TB is an infectious disease, and medical level will directly affect TB [14, 30, 31]. Number of beds in medical institutions, Number of health technicians per thousand population, and Number of medical institutions all reflect the medical level in a region, which in turn reflects the ability to detect and treat patients with TB in time, thereby reducing the spread of TB. Therefore, this study selects data such as Number of beds in medical institutions, Number of health technicians per thousand population, and Number of medical institutions from 2014 to 2018 as the factors influencing the relationship between medical level and TB, see Table 1.

### 2.3 Technical route

The technical route of this study is shown in Fig 3, and the important research contents and steps are briefly introduced as follows:

1. Collect relevant data. Including TB data in mainland China from 2014 to 2018, 11 influencing factors data, and map resource data.
2. Preprocess data. Including the VIF test on 11 influencing factors to judge whether they are suitable for analysis as impact factors, and linear transformation of each data to facilitate model calculation.
3. Macroscopic analysis of the distribution of TB incidence in mainland China from 2014 to 2018.





**Fig 3. Main technical route of this study.**

<https://doi.org/10.1371/journal.pone.0290978.g003>

4. Analyze the spatial distribution characteristics of TB by spatial autocorrelation methods and analyze spatial heterogeneity of TB by MGWR model.
5. Conduct comparative experiments. Verify the advantages of MGWR model in dealing with such problems through the comparison of OLS model, GWR model and MGWR model.
6. Carry out exploratory analysis of the influencing factors. Based on the coefficients and significance obtained from the analysis of MGWR model, targeted suggestions are put forward for the prevention and control of TB.

**2.3.1 Spatial autocorrelation.** In order to study the temporal and spatial distribution characteristics of TB, this study uses spatial autocorrelations method to explore the spatial distribution and spatial aggregation of TB incidence in mainland China in the five years before the outbreak of COVID-19, namely from 2014 to 2018. It can help us provide scientific decision-making support for the precise prevention and control of TB, taking mainland China as an example, when the authenticity and reliability of TB data has been reduced after the outbreak of COVID-19. And spatial autocorrelation refers to the potential dependence of some factors in a region, which can be roughly divided into two categories in terms of function: one is global spatial autocorrelation and the other is local spatial autocorrelation [32–35]. The global spatial autocorrelation method is used to test whether the whole region has correlation. The most widely used index is the global Moran's I index, which is often used to indicate the degree of correlation between each region and other regions, it is defined as follows:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) \sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$



Where  $n$  represents the number of statistical provinces and cities,  $x_i$  and  $x_j$  represent the annual incidence of TB in provinces  $i$  and  $j$ , respectively,  $\bar{x}$  represents the annual average incidence of 31 provinces and cities,  $w_{ij}$  is the spatial weight matrix, which is defined as follows:

$$w_{ij} = \begin{cases} 1, & i, j \text{ provinces and cities are adjacent or have a common edge} \\ 0, & i, j \text{ provinces and cities aren't adjacent or have no common edge} \end{cases} \quad (2)$$

The value range of Moran's  $I$  index is  $-1 < I < 1$ . When  $I > 0$ , it indicates that the study area is clustered between high-incidence areas and high-incidence areas, low-incidence areas and low-incidence areas, namely positive correlations. When  $I < 0$ , it indicates that the study area shows a mixed aggregation distribution of high-incidence areas and low-incidence areas, namely negative correlations. When  $I = 0$ , it indicates that the study area is randomly aggregated in high-incidence areas and low-incidence areas, namely no correlations.

The local spatial autocorrelation method is used to test whether the region have local aggregation when the whole region has correlation. This paper also uses the local Moran's  $I$  index for analysis, and its definition is shown in Formula 3, where the definitions of  $n$ ,  $x_i$ ,  $x_j$  and  $w_{ij}$  are the same as those defined in the global spatial autocorrelation method.

$$I_i = \frac{n(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \sum_{j=1, j \neq i}^n w_{ij}(x_j - \bar{x}) \quad (3)$$

Z test is often applied to the detection of the difference in the average value of large samples. It compares the Z score of the difference between the two averages with the specified theoretical Z score to determine whether the difference between the two averages is significant, and the relationship between the Z score and the significance of the difference is shown in Table 2. Through comparing the positive and negative values of Moran's  $I$  with the Z test value, four spatial distributions can be obtained: high-high (H-H) aggregation, high-low (H-L) aggregation, low-high (L-H) aggregation, low-low (L-L) aggregation [36].

**2.3.2 Multiscale Geographically Weighted Regression (MGWR) model.** MGWR model is one of the important models for analyzing spatial heterogeneity, which is often used in urban housing land prices, economic development, medical and health fields, but not often used in the exploration of disease influencing factors. MGWR model is an optimization model of GWR. On the one hand, it introduces the location characteristics of each sample point in space, and takes the distance between sample points as an important factor to define the weight, which can effectively reveal the spatial imbalance distribution. On the other hand, MGWR model further divides the research variables into global variables without significant spatial heterogeneity and local variables with significant spatial heterogeneity [37–41], and sets different bandwidths for each research variable, which can better explain the spatial effects of the research variables. Therefore, this paper uses MGWR model to explore the influencing factors of TB.

In order to better introduce MGWR model, this paper first introduces OLS model and GWR model. OLS model is often used to deal with the problem that a dependent variable is

**Table 2. Relationship between Z score and significance of difference.**

Z score	P value	Significance of difference
> 2.58	< 0.01	very significant
> 1.96	< 0.05	significant
≤ 1.96	≥ 0.05	not significant

<https://doi.org/10.1371/journal.pone.0290978.t002>



affected by multiple independent variables. The most common analytical model is Multiple Linear Regression Model, where  $y$  is the dependent variable, and  $x_1, x_2, \dots, x_i$  are independent variables, and they are linearly related. This study adopts this model for analysis, expressed as follows:

$$y = \beta_0 + \beta_i x_i + \epsilon_i \quad (4)$$

Where  $\beta_0$  is the constant,  $\beta_i$  is the regression coefficient of independent variable  $i$ ,  $x_i$  is the actual value of independent variable  $i$ , and  $\epsilon_i$  is the random error of independent variable  $i$ .

The difference between GWR model and ordinary linear regression model is that the spatial location information is added, and the regression coefficients of each area are analyzed separately [42–45]. First, the expression of GWR model is given as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^k \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad (5)$$

Where  $y_i$  is the predicted value of observation point  $i$ ,  $\beta_0(u_i, v_i)$  is the constant whose position  $(u_i, v_i)$  is the longitude and latitude of observation point  $i$ ,  $\beta_k(u_i, v_i)$  is the regression coefficient of observation point  $i$  to independent variable  $k$ ,  $x_{ik}$  is the actual value of observation point  $i$  to independent variable  $k$ ,  $\epsilon_i$  is the random error of observation point  $i$ .

The spatial weight matrix  $W(u_i, v_i)$  is the core of GWR model. The selection of the appropriate weight function is crucial to the accuracy of regression model, and the commonly used weight functions are Gaussian function and Bi-square function [46]. Among them, Gaussian function discusses all samples in the spatial scope, while Bi-square function only discusses the samples within the bandwidth range, which can avoid discussing some samples that have no impact on the regression coefficients. Therefore, when using GWR model, Bi-square function is often used as its spatial weight matrix, and it is expressed as follows:

$$w_{ij} = \begin{cases} \left(1 - \frac{d_{ij}^2}{b^2}\right)^2, & d_{ij} < b \\ 0, & d_{ij} \geq b \end{cases} \quad (6)$$

Where  $w_{ij}$  is the weight factor of fitting point  $i$  and sample point  $j$ ,  $d_{ij}$  is the distance of fitting point  $i$  and sample point  $j$ , and  $b$  is the bandwidth which represents the degree of non-negative attenuation between weight and distance.

When the bandwidth is selected, fixed function and adaptive function are often used. Fixed function is suitable for uniform distribution of sample points, and each sample point uses same bandwidth to analyze. Adaptive function is suitable for discrete distribution of sample points, and each sample point continuously adjusts the bandwidth  $b$  to make the model optimal [47]. Since the distribution of TB incidence is discrete, the adaptive function is used to analyze it.

MGWR model avoids the over-analysis of some independent variables in the GWR model, and divides the  $k$  independent variables in Formula (5) into  $p$  global variables without significant spatial heterogeneity and  $k-p$  local variables with significant spatial heterogeneity. It can be shown as:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i) x_{ik} + \sum_{k=p}^k \beta_k(u_i, v_i) x_{ik} + \epsilon_i \quad (7)$$



MGWR model can be regarded as a mixed model of GWR model and general linear regression model, which can judge spatial heterogeneity of each independent variable, speed up the calculation convergence speed and improve the fitting degree.

**2.3.3 Primary selection of influencing factors.** This paper builds a regression model based on the selected 11 influencing factors. In the establishment of MGWR model, some variables may have a certain degree of correlation with other variables, so these variables are defined as having multicollinearity. In order to ensure the rationality of statistical analysis, factors with multicollinearity should be eliminated. In this paper, variance inflation factor (VIF) test [48] is used to help eliminate collinearity problems, and some influencing factors with  $VIF > 10$  can be considered to have multicollinearity.

We conducted the VIF test on the selected variables in Table 1, including air quality, social life, and medical level. Table 3 shows that the VIF value of Per capita consumption expenditure is greater than 10, and this factor is considered to have multicollinearity, so it is no longer participate in subsequent modeling. The VIF of the other 10 influencing factors are all less than 10, and it can be considered that these influencing factors have no multicollinearity, so these factors are retained to continue to participate in subsequent modeling.

**2.3.4 Comparative experiment.** In this paper, OLS model and GWR model are selected to carry out comparative experiments to verify the performance of MGWR model. The OLS model is realized by SPSS24.0 software, the GWR model is realized by GWR4 software, and the MGWR model is realized by MGWR2.2 software. Among them, many criteria are used to judge the fitting effect of statistical model. The most common one is AIC (Akaike Information Criterion), which is used as a standard to measure the fitting effect of statistical model and is defined as follows:

$$AIC = 2n + n \ln \left( \frac{SSR}{n} \right) \quad (8)$$

$$SSR = \sum_{i=1}^n (x - \bar{x})^2 \quad (9)$$

Where  $n$  is the number of statistical sample points,  $x$  is the predicted data of observation point,  $\bar{x}$  is the actual value of observation point, and SSR is residual sum of squares. Among them, the lower the SSR and AIC, the smaller the prediction deviation of the model and the better the fitting effect. Meanwhile, use Mean Absolute Error MAE, Mean Square Error MSE

**Table 3. VIF test of influencing factors.**

Dimension	Indicator	VIF
air quality	PM <sub>2.5</sub>	3.540
	SO <sub>2</sub>	1.916
	NO <sub>2</sub>	2.856
social life	Per capita consumption expenditure	16.097
	Average number of clinic visits	4.948
	Per capita healthcare consumption expenditure	5.009
	Per capita GDP	8.082
	Passenger traffic volume	3.443
medical level	Number of beds in medical institutions	8.790
	Number of health technicians per thousand population	2.501
	Number of medical institutions	6.612

<https://doi.org/10.1371/journal.pone.0290978.t003>



and Mean Absolute Percentage Error MAPE to compare the error between the predicted value and the actual value, which are defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (10)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (11)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \quad (12)$$

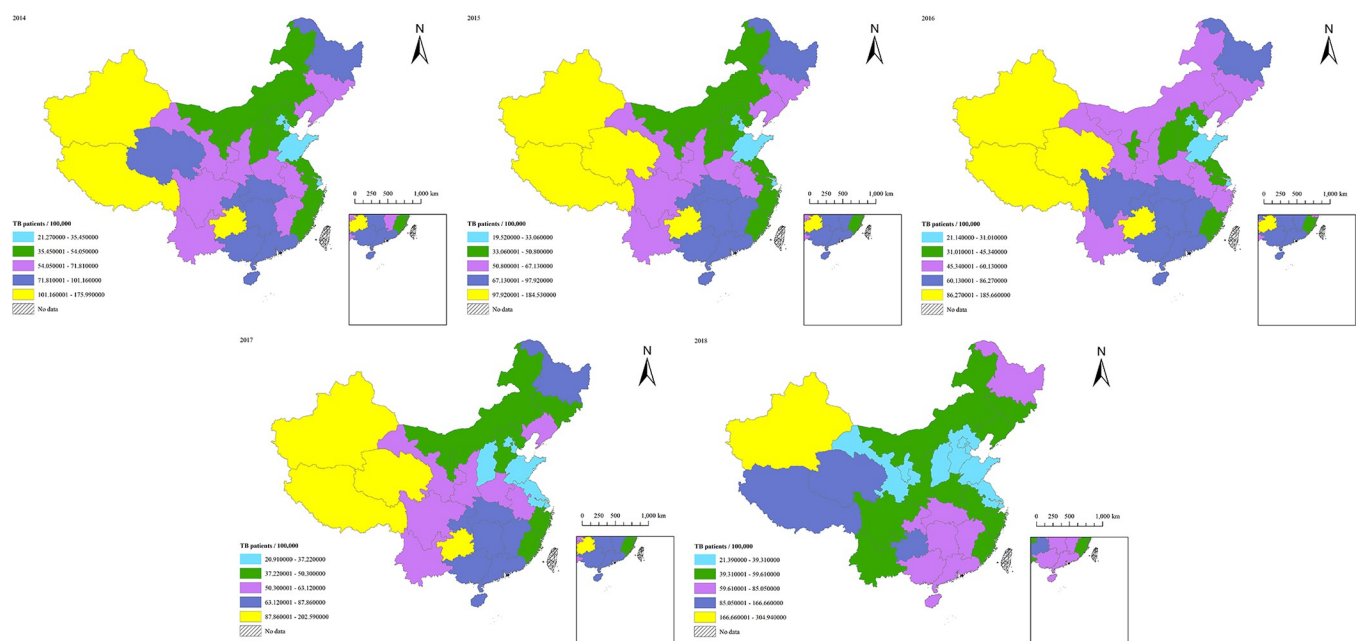
Where  $\hat{y}_i$  is the predicted value of observation point  $i$ , and  $y_i$  is the actual value of observation point  $i$ .

### 3. Experiments and results

#### 3.1 Macroscopic analysis and spatial autocorrelation experiment of TB

The spatial distribution of TB incidence in mainland China from 2014 to 2018 is shown in Fig 4. The areas with the highest annual average incidence of TB were Xinjiang (211.81/100,000), Tibet (153/100,000), Guizhou (125.89/100,000), Qinghai (125.69/100,000) and the areas with the lowest annual average incidence of TB were Tianjin (20.85/100,000), Shanghai (27.25/100,000), Shandong (31.49/100,000) and Beijing (32.09/100,000). The overall incidence showed a downward trend from 2014 to 2018, but the incidence in Qinghai, Tibet and Xinjiang increased from 101.16/100,000, 147.99/100,000, 175.99/100,000 to 140.33/100,000, 166.66/100,000, 304.94/100,000, respectively.

The data of TB incidence in mainland China are associated with the vector map of mainland China through ArcGIS10.8, and the results of global spatial autocorrelation are shown in



**Fig 4. Distribution of TB incidence in mainland China from 2014 to 2018.**

<https://doi.org/10.1371/journal.pone.0290978.g004>



**Table 4. Spatial aggregation analysis results of TB in mainland China from 2014 to 2018.**

year	Moran's I	Z score	P value
2014	0.443014	4.235262	< 0.01
2015	0.454716	4.341855	< 0.01
2016	0.486561	4.662733	< 0.01
2017	0.499934	4.887245	< 0.01
2018	0.384110	4.519545	< 0.01

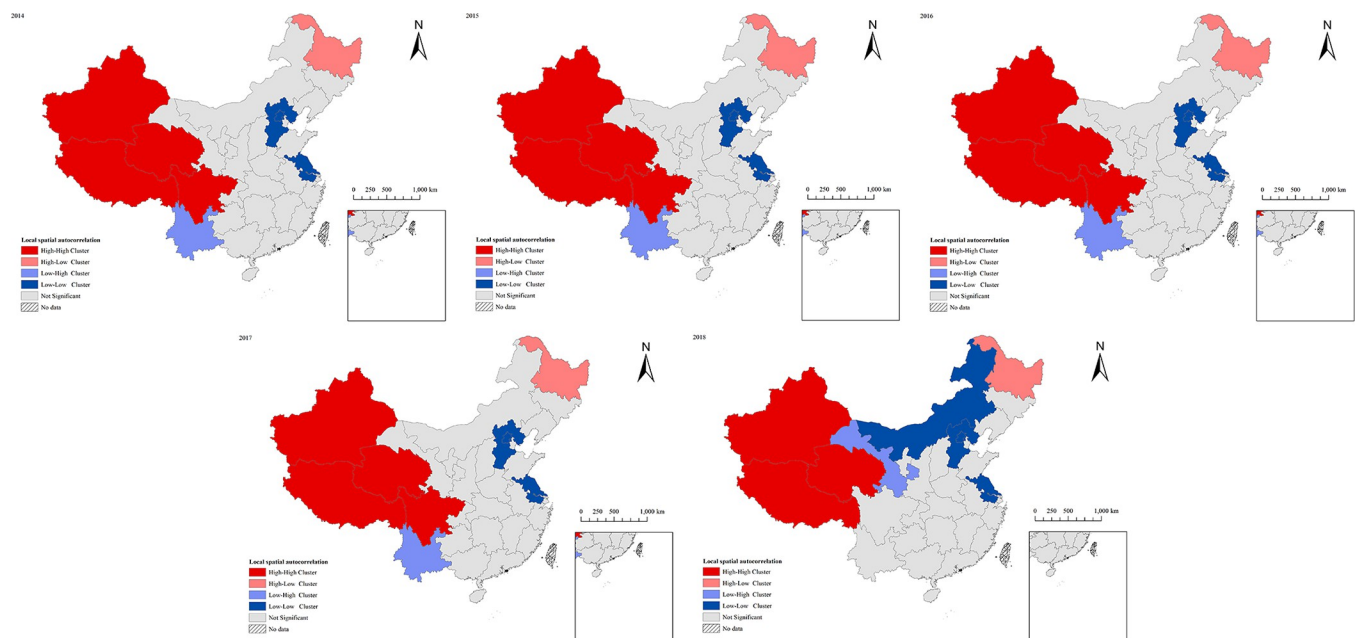
<https://doi.org/10.1371/journal.pone.0290978.t004>

**Table 4.** The values of Moran's I from 2014 to 2018 are all greater than 0, and the P values are all less than 0.01, indicating that TB incidence has spatial aggregation in the study area. Local spatial autocorrelation analysis shows that there are four spatial distributions of TB in mainland China, as shown in Fig 5. The H-H clusters from 2014 to 2017 are concentrated in Xinjiang, Tibet, Qinghai and Sichuan, the H-L clusters are concentrated in Heilongjiang, the L-H clusters are concentrated in Yunnan, and the L-L clusters are concentrated in Beijing, Hebei and Jiangsu. Moreover, the H-H clusters in 2018 are concentrated in Xinjiang, Tibet and Qinghai, the H-L clusters are still concentrated in Heilongjiang, the L-H clusters are concentrated in Gansu, and the L-L clusters are concentrated in Beijing, Hebei, Jiangsu and Inner Mongolia. It indicates the presence of local clusters of TB within the study area.

### 3.2 Experiments and comparative experiments of MGWR analysis

The selected influencing factors are analyzed through OLS model, and the analysis results are shown in Table 5. At this point, Formula 4 is transformed into:

$$y = 125.19818 + 0.44310x_1 + 1.01356x_2 + 0.03326x_3 - 5.74165x_4 - 0.02633x_5 - 0.00057x_6 - 0.00008x_7 + 0.29768x_8 + 10.07590x_9 - 0.00072x_{10}$$

**Fig 5. Distribution of local clusters of TB in mainland China from 2014 to 2018.**

<https://doi.org/10.1371/journal.pone.0290978.g005>



Table 5. Parameters of OLS model.

Dimension	Indicator	Coefficient
air quality	Constant	125.19818
	PM <sub>2.5</sub>	0.44310
	SO <sub>2</sub>	1.01356
	NO <sub>2</sub>	0.3326
social life	Average number of clinic visits	-5.74165
	Per capita healthcare consumption expenditure	-0.02633
	Per capita GDP	-0.00057
	Passenger traffic volume	-0.00008
medical level	Number of beds in medical institutions	0.29768
	Number of health technicians per thousand population	10.07590
	Number of medical institutions	-0.00072

<https://doi.org/10.1371/journal.pone.0290978.t005>

Where  $x_1$  represents the actual value of PM<sub>2.5</sub>,  $x_2$  represents the actual value of SO<sub>2</sub>,  $x_3$  represents the actual value of NO<sub>2</sub>,  $x_4$  represents the actual value of Average number of clinic visits,  $x_5$  represents the actual value of Per capita healthcare consumption expenditure, and  $x_6$  represents the actual value of Per capita GDP,  $x_7$  represents the actual value of Passenger traffic volume,  $x_8$  represents the actual value of Number of beds in medical institutions,  $x_9$  represents the actual value of Number of health technicians per thousand population,  $x_{10}$  represents the actual value of Number of medical institutions.

Compared with OLS model, the core point of GWR model is to add spatial location information and introduce weight matrix, and the results are shown in Table 6. For specific parameters, see S4 Dataset.

The bandwidth of MGWR model reflects the effect of the influencing factors on the model. The bandwidth is larger, the spatial heterogeneity of the influencing factors is less obvious, and vice versa [49]. As shown in Table 7, the bandwidth of SO<sub>2</sub> in air quality factors is 46, and the spatial heterogeneity is not obvious. However, the bandwidths of PM<sub>2.5</sub> and NO<sub>2</sub> are 151 and 137, respectively, which are basically equal to the total number of samples, so they are global variables. Among the social life factors, the bandwidths of Average number of clinic visits, Per capita healthcare consumption expenditure, Per capita GDP and Passenger traffic volume are 46, 46, 67 and 46, respectively, with little spatial heterogeneity. In the medical level factors, the bandwidths of Number of health technicians per thousand population and Number of medical

Table 6. Parameters of GWR model.

Dimension	Indicator	Mean	Standard Deviation
air quality	Constant	-0.287	0.152
	PM <sub>2.5</sub>	0.023	0.303
	SO <sub>2</sub>	0.207	0.318
	NO <sub>2</sub>	-0.068	0.215
social life	Average number of clinic visits	-0.319	0.252
	Per capita healthcare consumption expenditure	-0.083	0.279
	Per capita GDP	0.130	0.387
	Passenger traffic volume	-0.068	0.295
medical level	Number of beds in medical institutions	0.175	0.225
	Number of health technicians per thousand population	-0.054	0.355
	Number of medical institutions	-0.305	0.489

<https://doi.org/10.1371/journal.pone.0290978.t006>



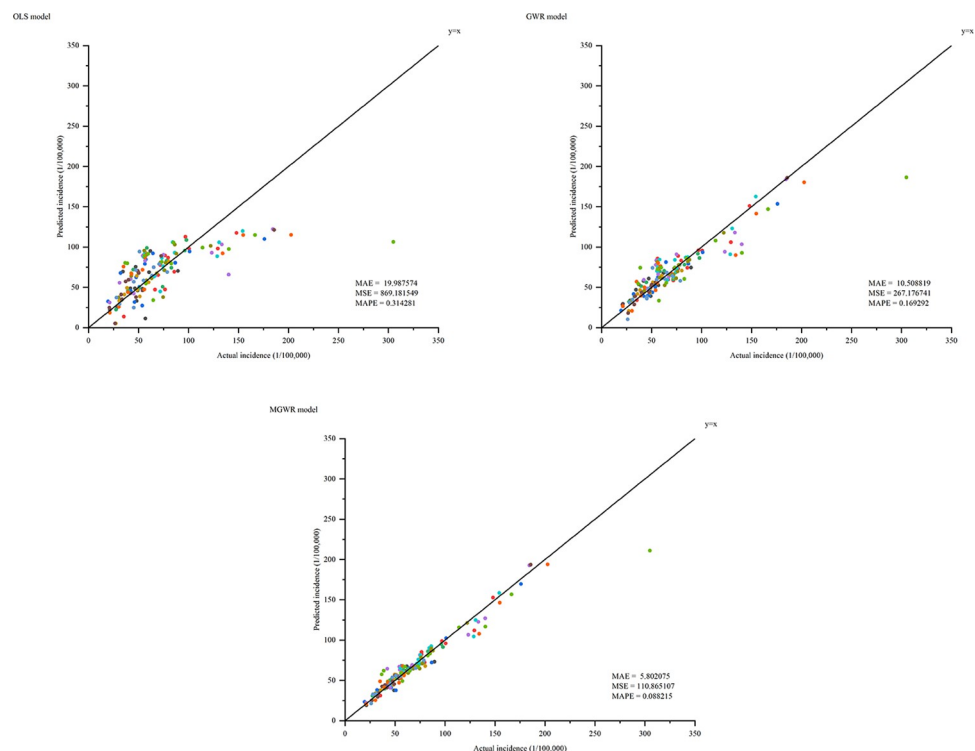
Table 7. Parameters of MGWR model.

Dimension	Indicator	bandwidth	Adjust t value
air quality	Constant	46	2.492
	PM <sub>2.5</sub>	151	2.029
	SO <sub>2</sub>	46	2.605
	NO <sub>2</sub>	137	2.149
social life	Average number of clinic visits	46	2.564
	Per capita healthcare consumption expenditure	46	2.659
	Per capita GDP	67	2.363
	Passenger traffic volume	46	2.574
medical level	Number of beds in medical institutions	122	2.100
	Number of health technicians per thousand population	56	2.528
	Number of medical institutions	56	2.410

<https://doi.org/10.1371/journal.pone.0290978.t007>

institutions are 56 and 56, respectively, which shows that the spatial heterogeneity is not obvious. However, the bandwidth of Number of beds in medical institutions is 122, which is almost equal to the total sample size, so it is a global variable. According to the t-test table, the adjusted t-values are all greater than 1.812, which indicates that the selected impact factors are significantly credible at the 95% confidence level.

As shown in Fig 6, the MAE values of OLS, GWR and MGWR models are 19.987574, 10.508819, 5.802075, the MSE values are 869.181549, 267.176741, 110.865107 and the MAPE values are 0.314281, 0.169292 and 0.088215, respectively. The MAE, MSE and MAPE of MGWR model are better than the other two models, which are closer to the actual incidence.



**Fig 6. Model comparison results.**

<https://doi.org/10.1371/journal.pone.0290978.g006>



Table 8. Comparison of model fitting indicators.

indicator	OLS	GWR	MGWR
$R^2$	0.466	0.836	0.932
Adjust $R^2$	0.429	0.797	0.910
Residual sum of squares	134723	41412	17184
AIC	1512.84	1360.13	1198.89

<https://doi.org/10.1371/journal.pone.0290978.t008>

As shown in Table 8, the MGWR model has a smaller AIC value and residual sum of squares than the other two models. Among them, the  $R^2$  and the adjusted  $R^2$  are 0.932 and 0.910, respectively, indicating that the fitting effect of the MGWR model is better than the other two models.

#### 4. Results and discussion of MGWR analysis

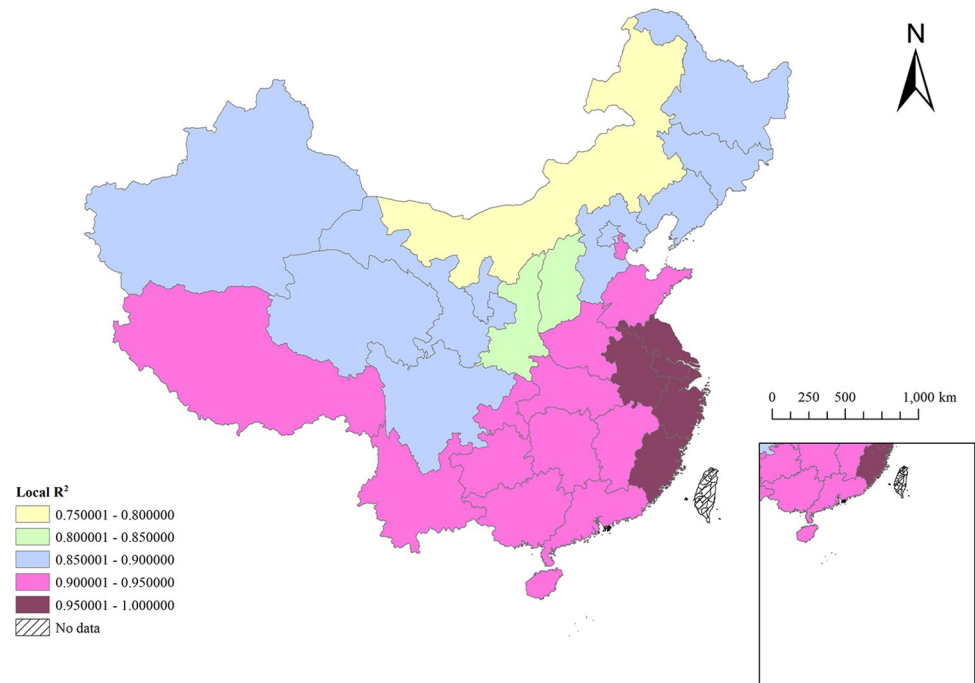
As COVID-19 swept the world in 2019, each country struggled to deal with patients infected with COVID-19, which occupied most of medical resources. At the same time, some patients with TB failed to go to the hospital in time, resulting in a sharp drop in the number of TB cases. In order to ensure the objectivity and reliability of the analyzed data, this period is excluded in this study. This paper explores the temporal and spatial distribution of TB incidence in mainland China in the five years before the outbreak of COVID-19, and analyzes the degree of impact on TB incidence from three levels: air quality, social life, and medical level. The quantitative contribution of each influencing factor is calculated through the MGWR model, which provides a decision-making reference for the prevention and treatment of TB in other countries.

As shown in Fig 4, the areas with high incidence are mainly concentrated in the Southwestern China and the Central of China from 2014 to 2018, while the incidence of TB in East China is relatively low. The overall incidence shows a trend of high in the west and low in the east [50], with large local differences and obvious aggregation distribution in some areas. For example, the incidence in the surrounding areas of Qinghai and Guizhou is generally high and shows a positive aggregation distribution, while the incidence in the surrounding areas of Beijing and other places is low, showing a reverse aggregation distribution. Relatively backward economic development, and insufficient prevention and control of infectious diseases are the main reasons for high incidence of TB in Qinghai, Tibet, Xinjiang [18, 51, 52].

For the MGWR model in this study, the local  $R^2$  in Fig 7 is the goodness of fit of the selected influencing factors to TB incidence in each region. The results show that the local  $R^2$  fitting effect of each region is good, including 96.8% of the study areas with a goodness of fit exceeding 0.80, 58.1% of the study areas with a goodness of fit exceeding 0.90, and 16.1% of the study areas with a goodness of fit exceeding 0.95, which prove that the selected influencing factors have sufficient explanatory power for TB incidence. The larger  $R^2$  in East China and the Central of China indicates that the selected influencing factors have the best explanatory power in this area. However, dry climate, sparse vegetation, and excessive dust lead to an increase in TB [53, 54], which makes the model fit very low in North China.

TB is a respiratory infectious disease, and air quality directly affects its incidence [55–57]. The regression coefficients and significance of constant terms and air quality factors of the MGWR model constructed in this paper are shown in Fig 8. Among them, the trend of the coefficient of constant term is basically consistent with the trend of TB incidence in mainland China from 2014 to 2018. The eastern regions of China all show a downward trend in different degrees, while the western regions also have a small upward trend. Among the selected air





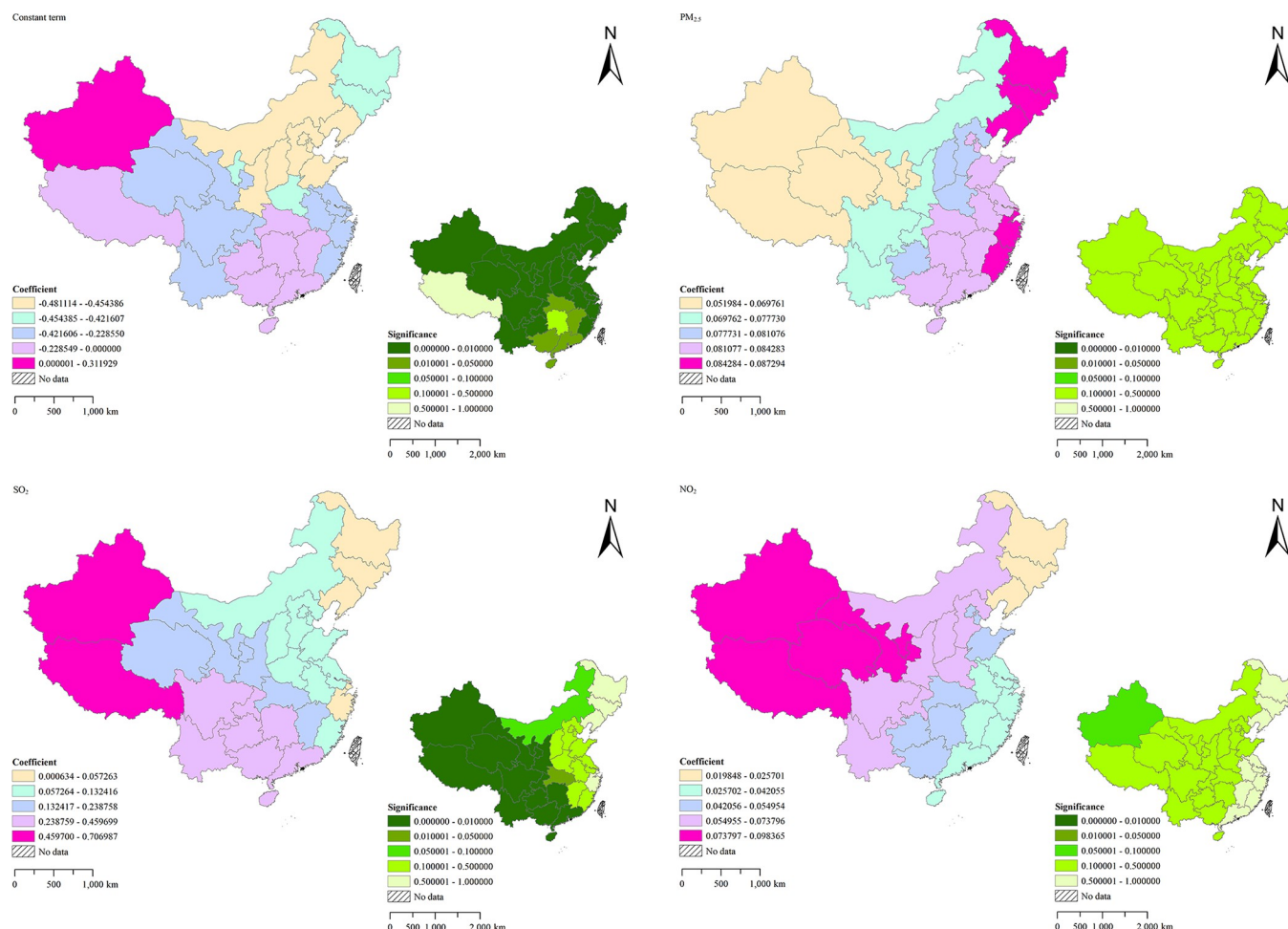
**Fig 7. Spatial distribution of local  $R^2$  of MGWR model.**

<https://doi.org/10.1371/journal.pone.0290978.g007>

quality factors,  $PM_{2.5}$  shows a significant positive correlation trend to TB incidence, and the correlation degree gradually increases from west to east. As a particle with a diameter of less than 2.5 microns, it can directly enter the lungs through respiratory tract and affect human health [58], leading to an increase in TB incidence. Moreover,  $SO_2$  shows a significant positive correlation trend to TB incidence, and the impact on the western regions is greater. As a colorless and irritating gas,  $SO_2$  has an irritating effect on the respiratory tract, and long-term exposure to high concentrations may cause lung damage, thereby increasing the possibility of TB [55, 56]. An increase in the concentration of  $NO_2$  also leads to the same problem. According to the research, long-term exposure to high concentrations of  $NO_2$  will lead to alveolar atrophy, pulmonary edema and other problems [59], which leads to the onset of TB. The regression coefficient of  $NO_2$  obtained by the MGWR model also proves this view. Although the absolute value of the regression coefficient is relatively low, the problem of  $NO_2$  emission can't be ignored. Every country should still pay attention to this problem and achieve effective prevention and control of TB.

Air quality is not the only aspect that affects the spread of TB, social living standard also have a great impact on TB from many different levels [10, 11, 60]. The regression coefficients and significance of social life factors are shown in Fig 9. Among them, Average number of clinic visits has a significant negative correlation trend with TB incidence. Due to the relatively low awareness of prevention, the influence of this factor is more obvious in the western regions of China, and the confidence is generally high. Early diagnosis and timely treatment are important methods to effectively control TB, and delay diagnosis may increase infectivity and worsen the condition [61]. Therefore, especially under COVID-19, regular hospital check-ups are important for the prevention of TB. In addition, Per capita healthcare consumption expenditure and TB incidence also show a significant negative correlation trend, which have a greater impact on the Southern China and Southwestern China, and the confidence is generally high. Due to the high medical expenses of TB patients [62], it is important to appropriately



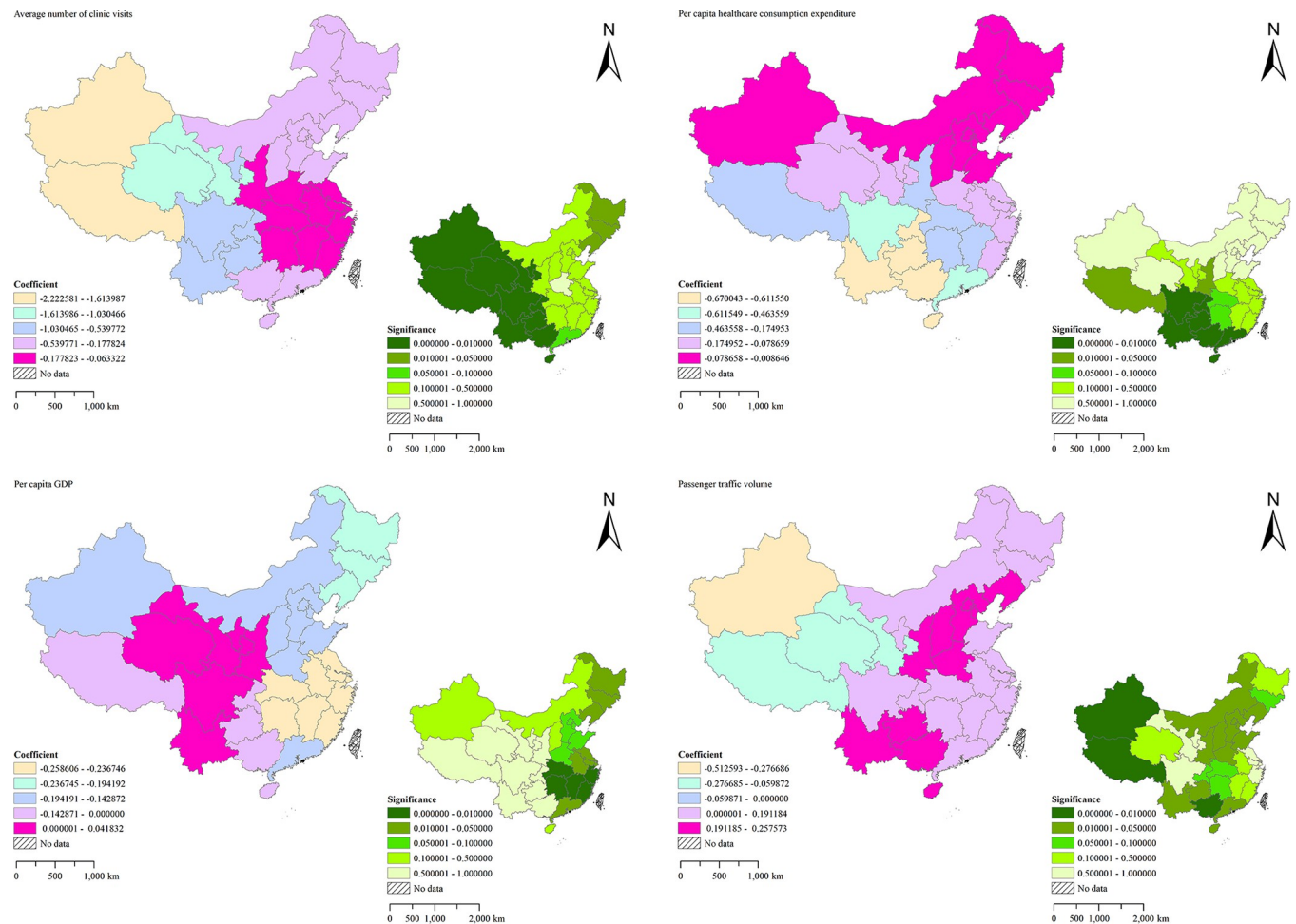


**Fig 8. Spatial distribution of regression coefficients and significance of TB in mainland China (constant term and air quality factors).**

<https://doi.org/10.1371/journal.pone.0290978.g008>

increase the investment in healthcare costs during the early prevention period for the prevention of TB. However, the influence of Per capita GDP on TB incidence is polarized. Some central regions of China have a significant positive correlation trend, accounting for 19.35% of the study areas. According to the research, people in economically underdeveloped areas are more likely to suffer from TB [1, 18, 29], because these areas often overemphasize regional development but neglect the control of TB. Aiming at the problems in these areas, the suggestion put forward in this paper is that these areas should focus on the development of economic construction and disease control at the same time. In other regions of China, Per capita GDP and the incidence of TB show a negative correlation trend, accounting for 80.65% of the study areas. With the development of regional economy, these regions have strengthened the control of TB, providing an example for other countries that economic growth drives disease prevention. The effect of Passenger traffic volume on TB incidence also shows a significant polarization phenomenon, and the positive correlation trend is significant in the North China, Northeastern China, the Central of China, Southern China and East China, accounting for 87.10% of the study areas. Due to the relatively developed economy, the improvement of transportation convenience, and the continuous increase in the migration of floating population, the possibility of contacting other TB patients on the transport has also increased, resulting in an upward trend in the incidence of TB in these regions [63]. However, the negative





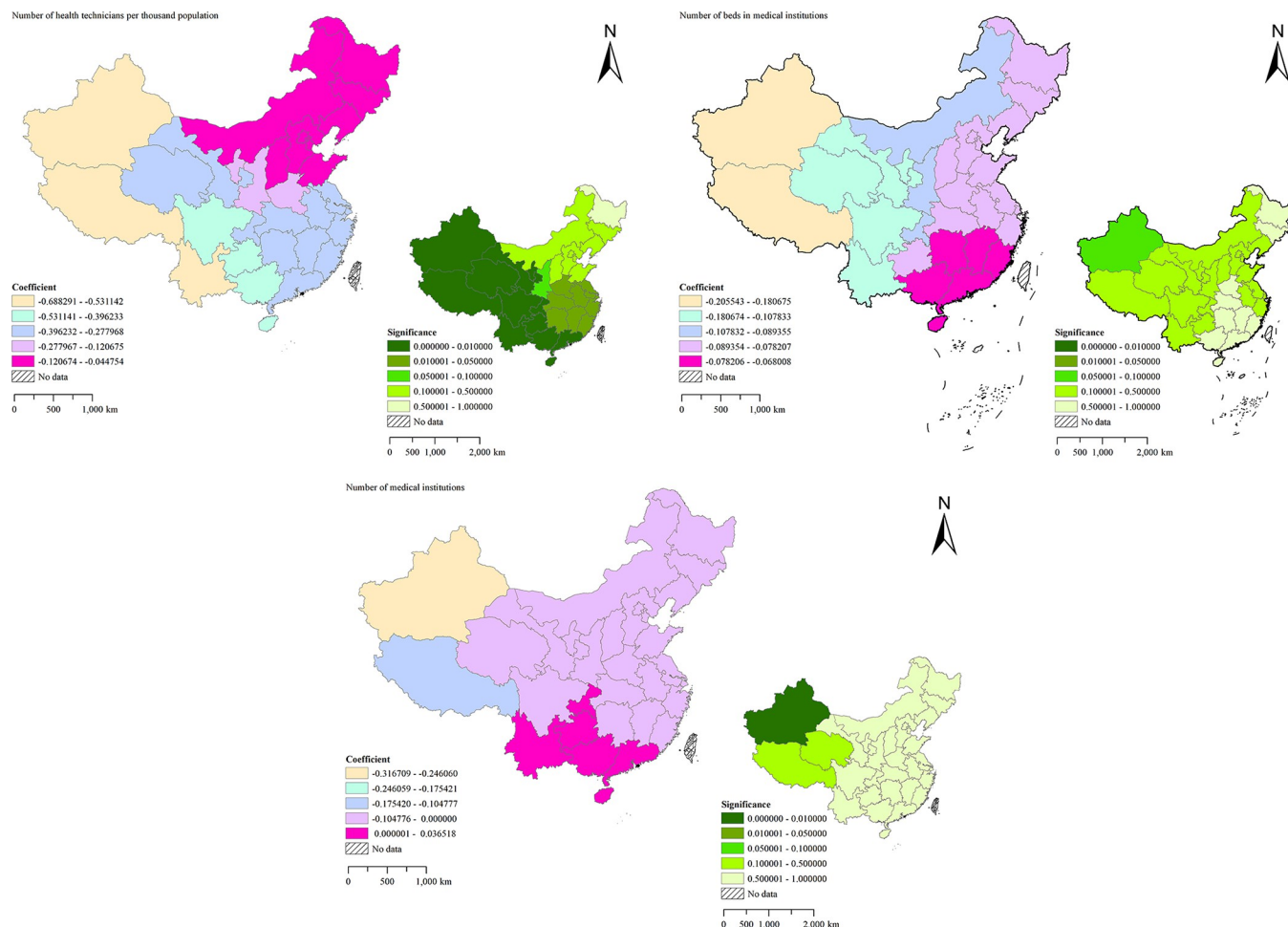
**Fig 9. Spatial distribution of regression coefficients and significance of TB in mainland China (social life factors).**

<https://doi.org/10.1371/journal.pone.0290978.g009>

correlation trend between Passenger traffic volume and the incidence of TB in the Northeastern China and Southwestern China, accounting for 12.90% of the study areas. Compared with other areas, these areas have lower population density and weaker economic strength, resulting in the migration of indigenous people to economically developed regions, such as the eastern regions of China. With the improvement of traffic convenience, the population flow in the western regions of China is more obvious. The indigenous people in the western regions of China migrate to other regions, resulting to a downward trend in the incidence of TB, while an upward trend in other regions, which also explains the phenomenon why the eastern regions of China is positively correlated.

The medical level is also crucial to the prevention and treatment of TB [64]. The regression coefficients and significance of medical level factors are shown in Fig 10. Among them, Number of health technicians per thousand population and TB incidence show a significant negative correlation trend. The Southwestern China and Northwestern China are significantly affected by this factor, and the overall confidence in these regions is relatively high, which effectively proves the importance of this factor to the prevention and control of TB in these regions. Therefore, an appropriate increase in medical technicians can effectively reduce the incidence of TB [51]. Furthermore, Number of beds in medical institutions and TB incidence also show a significant negative correlation trend, which has a greater impact on the Northwestern China and Southwestern





**Fig 10. Spatial distribution of regression coefficients and significance of TB in mainland China (medical level factors).**

<https://doi.org/10.1371/journal.pone.0290978.g010>

China. With the gradual increase in hospital investment and the number of beds, the situation that once shortened the hospitalization days of TB patients no longer appeared, resulting in more and more effective treatment of TB today [61]. Although the absolute value of the overall regression coefficient is low, appropriately increasing the number of beds in medical institutions can still effectively reduce the number of TB patients. However, the impact of Number of medical institutions on the incidence of TB is polarized. This factor has a significant positive correlation trend with TB in Southern China, accounting for 19.35% of the study areas. Due to the dense population and extremely fast population growth rate in these areas, there are also more medical institutions in these areas, but these reasons also make it more difficult to control the spread of TB. The number of medical institutions in other regions of China has a significant negative correlation trend with TB incidence, accounting for 80.65% of the study areas. Abundant medical resources enable the control of TB to be well managed [65]. Therefore, targeted increase of medical institutions can more effectively prevent the spread of TB.

## 5. Conclusion

Due to the decline in the authenticity and reliability of TB under the COVID-19, this paper uses spatial autocorrelation methods to analyze the spatial distribution of TB incidence in



mainland China from 2014 to 2018, and introduces MGWR model to analyze its spatial heterogeneity. Using the collected data on air quality, social life and medical level, the potential factors affecting the transmission of TB are analyzed. Moreover, through the comparative experiment of OLS model and GWR model, the superiority of MGWR model in dealing with such problems is verified, and it provides decision-making references for China and other countries in the prevention and control of TB under the impact of major infectious diseases such as COVID-19. The following are the main findings of this study:

1. The overall incidence of TB in mainland China shows a downward trend from 2014 to 2018, and the distribution of TB cases in the study areas shows spatial aggregation and spatial heterogeneity during this period.
2. Among the air quality factors,  $PM_{2.5}$ ,  $SO_2$  and  $NO_2$  are positively correlated with TB incidence. Among the social life factors, Average number of clinic visits, Per capita healthcare consumption expenditure, Per capita GDP and Passenger traffic volume are generally negatively correlated with TB incidence. Among the medical level factors, Number of health technicians per thousand population, Number of beds in medical institutions and Number of medical institutions generally have a negative effect. Therefore, improving air quality, promoting national economic development, and increasing investment in medical resources can effectively control the spread of TB.
3. The MGWR model introduces spatial position characteristics of each sample point, and divides them into global variables and local variables according to the effect of each influencing factor on the model. Therefore, it makes the influencing factor have better spatial-temporal analysis ability, and its overall fitting effect is significantly better than OLS model and GWR model.

However, this paper also has some limitations. In terms of data, if we can collect real and reliable data on the incidence of TB under the COVID-19 and compare it with this paper, we can provide more reliable reference opinions for the prevention and control of TB, and there may still be some influencing factors that have not been put into the model for discussion, resulting in a reduction in the explanatory power of the model. Follow-up research can further add other influencing factors and research data to make the model more perfect.

## Supporting information

**S1 Dataset. TB data and selected influencing factors data from 2014 to 2018.**  
(XLSX)

**S2 Dataset. Monthly incidence data of TB in 2014–2018.**  
(XLSX)

**S3 Dataset. MGWR coefficient table.**  
(XLSX)

**S4 Dataset. GWR coefficient table.**  
(XLSX)

## Acknowledgments

Throughout the writing of this dissertation, I have received a great deal of support and assistance.



I would first like to thank my supervisor, Prof. Hong Fan, whose expertise was invaluable in formulating the research questions and methodology. Thank you for your valuable guidance throughout my studies. You provided me with the right direction and brought my work to a higher level.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear.

Finally, I could not have completed this dissertation without the support of my friends, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

## Author Contributions

**Conceptualization:** Zhipeng Ma.

**Data curation:** Hong Fan.

**Formal analysis:** Zhipeng Ma.

**Funding acquisition:** Hong Fan.

**Validation:** Zhipeng Ma.

**Visualization:** Zhipeng Ma.

**Writing – original draft:** Zhipeng Ma.

**Writing – review & editing:** Zhipeng Ma.

## References

1. Ng IC, Wen TH, Wang JY, Fang CT. Spatial dependency of tuberculosis incidence in Taiwan. *PLoS One*. 2012; 7(11): e50740. <https://doi.org/10.1371/journal.pone.0050740> PMID: 23226371
2. Alene KA, Wangdi K, Clements ACA. Impact of the COVID-19 Pandemic on Tuberculosis Control: An Overview. *Trop Med Infect Dis*. 2020; 5(3): 123. <https://doi.org/10.3390/tropicalmed5030123> PMID: 32722014
3. Alemu A, Bitew ZW, Seid G, Diriba G, Gashu E, Berhe N, et al. Tuberculosis in individuals who recovered from COVID-19: A systematic review of case reports. *PLoS One*. 2022; 17(11): e0277807. <https://doi.org/10.1371/journal.pone.0277807> PMID: 36441785
4. Wang W, Guo W, Cai J, Guo W, Liu R, Liu X, et al. Epidemiological characteristics of tuberculosis and effects of meteorological factors and air pollutants on tuberculosis in Shijiazhuang, China: A distribution lag non-linear analysis. *Environ Res*. 2021; 195: 110310. <https://doi.org/10.1016/j.envres.2020.110310> PMID: 33098820
5. Yang S, Gao Y, Luo W, Liu L, Lei Y, Zhang X. Spatiotemporal Distribution of Tuberculosis during Urbanization in the New Urban Area of Nanchang City, China, 2010–2018. *Int J Environ Res Public Health*. 2019; 16(22): 4395. <https://doi.org/10.3390/ijerph16224395> PMID: 31717896
6. Hwang SS, Kang S, Lee JY, Lee JS, Kim HJ, Han SK, et al. Impact of outdoor air pollution on the incidence of tuberculosis in the Seoul metropolitan area, South Korea. *Korean J Intern Med*. 2014; 29(2): 183–90. <https://doi.org/10.3904/kjim.2014.29.2.183> PMID: 24648801
7. Ehrlich R, Akugizibwe P, Siegfried N, Rees D. The association between silica exposure, silicosis and tuberculosis: a systematic review and meta-analysis. *BMC Public Health*. 2021; 21(1): 953. <https://doi.org/10.1186/s12889-021-10711-1> PMID: 34016067
8. Li Z, Mao X, Liu Q, Song H, Ji Y, Xu D, et al. Long-term effect of exposure to ambient air pollution on the risk of active tuberculosis. *Int J Infect Dis*. 2019; 87: 177–184. <https://doi.org/10.1016/j.ijid.2019.07.027> PMID: 31374344
9. Im C, Kim Y. Spatial pattern of tuberculosis (TB) and related socio-environmental factors in South Korea, 2008–2016. *PLoS One*. 2021; 16(8): e0255727. <https://doi.org/10.1371/journal.pone.0255727> PMID: 34352032
10. Vargas MH, Furuya ME, Pérez-Guzmán C. Effect of altitude on the frequency of pulmonary tuberculosis. *Int J Tuberc Lung Dis*. 2004; 8(11): 1321–4. PMID: 15581199



11. Maciel EL, Pan W, Dietze R, Peres RL, Vinhas SA, Ribeiro FK, et al. Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil. *Int J Tuberc Lung Dis*. 2010; 14(11): 1395–402. PMID: [20937178](#)
12. Harling G, Ehrlich R, Myer L. The social epidemiology of tuberculosis in South Africa: a multilevel analysis. *Soc Sci Med*. 2008; 66(2): 492–505. <https://doi.org/10.1016/j.socscimed.2007.08.026> PMID: [17920743](#)
13. Memish ZA, Bamgboye EA, Abuljadayel N, Smadi H, Abouzeid MS, Al Hakeem RF. Incidence of and risk factors associated with pulmonary and extra-pulmonary tuberculosis in Saudi Arabia (2010–2011). *PLoS One*. 2014; 9(5): e95654. <https://doi.org/10.1371/journal.pone.0095654> PMID: [24824783](#)
14. Li XX, Wang LX, Zhang J, Liu YX, Zhang H, Jiang SW, et al. Exploration of ecological factors related to the spatial heterogeneity of tuberculosis prevalence in P. R. China. *Glob Health Action*. 2014; 7: 23620. <https://doi.org/10.3402/gha.v7.23620> PMID: [24929074](#)
15. Sun W, Gong J, Zhou J, Zhao Y, Tan J, Ibrahim AN, et al. A spatial, social and environmental study of tuberculosis in China using statistical and GIS technology. *Int J Environ Res Public Health*. 2015; 12(2): 1425–48. <https://doi.org/10.3390/ijerph120201425> PMID: [25633032](#)
16. Chen J, Qiu Y, Yang R, Li L, Hou J, Lu K, et al. The characteristics of spatial-temporal distribution and cluster of tuberculosis in Yunnan Province, China, 2005–2018. *BMC Public Health*. 2019; 19(1): 1715. <https://doi.org/10.1186/s12889-019-7993-5> PMID: [31864329](#)
17. Yang DL, Li W, Pan MH, Su HX, Li YN, Tang MY, et al. Spatial analysis and influencing factors of pulmonary tuberculosis among students in Nanning, during 2012–2018. *PLoS One*. 2022; 17(5): e0268472. <https://doi.org/10.1371/journal.pone.0268472> PMID: [35609085](#)
18. Wubuli A, Xue F, Jiang D, Yao X, Upur H, Wushouer Q. Socio-Demographic Predictors and Distribution of Pulmonary Tuberculosis (TB) in Xinjiang, China: A Spatial Analysis. *PLoS One*. 2015; 10(12): e0144010. <https://doi.org/10.1371/journal.pone.0144010> PMID: [26641642](#)
19. Magalhães MA, Medronho RA. Spatial analysis of Tuberculosis in Rio de Janeiro in the period from 2005 to 2008 and associated socioeconomic factors using micro data and global spatial regression models. *Cien Saude Colet*. 2017; 22(3): 831–840. <https://doi.org/10.1590/1413-81232017223.24132015> PMID: [28300991](#)
20. Ponticelli A, Sturkenboom MC, Simonetti A, Ortolani R, Malerba M, Sanduzzi A. Deprivation, immigration and tuberculosis incidence in Naples, 1996–2000. *Eur J Epidemiol*. 2005; 20(8): 729–34. <https://doi.org/10.1007/s10654-005-0615-9> PMID: [16151887](#)
21. Harling G, Castro MC. A spatial analysis of social and economic determinants of tuberculosis in Brazil. *Health Place*. 2014; 25: 56–67. <https://doi.org/10.1016/j.healthplace.2013.10.008> PMID: [24269879](#)
22. Erazo C, Pereira SM, Costa Mda C, Evangelista-Filho D, Braga JU, Barreto ML. Tuberculosis and living conditions in Salvador, Brazil: a spatial analysis. *Rev Panam Salud Publica*. 2014; 36(1): 24–30. PMID: [25211674](#)
23. Helbich M, Brunauer W, Vaz E, Nijkamp P. Spatial heterogeneity in hedonic house price models: The case of Austria. *Urban Studies*. 2014; 51(2): 390–411.
24. Wen HZ, Jin YL, Zhang L. Spatial heterogeneity in implicit housing prices: Evidence from Hangzhou, China. *International Journal of Strategic Property Management*. 2017; 21(1): 15–28.
25. Cui WC, Li JY, Xu WT, Güneralp B. Industrial electricity consumption and economic growth: A spatio-temporal analysis across prefecture-level cities in China from 1999 to 2014. *Energy*. 2021; 222: 119932.
26. Forati AM, Ghose R, Mantsch JR. Examining Opioid Overdose Deaths across Communities Defined by Racial Composition: a Multiscale Geographically Weighted Regression Approach. *J Urban Health*. 2021; 98(4): 551–562. <https://doi.org/10.1007/s11524-021-00554-x> PMID: [34231120](#)
27. Smith GS, Schoenbach VJ, Richardson DB, Gammon MD. Particulate air pollution and susceptibility to the development of pulmonary tuberculosis disease in North Carolina: an ecological study. *Int J Environ Health Res*. 2014; 24(2): 103–12. <https://doi.org/10.1080/09603123.2013.800959> PMID: [24387197](#)
28. Yang J, Zhang M, Chen Y, Ma L, Yadikaer R, Lu Y, et al. A study on the relationship between air pollution and pulmonary tuberculosis based on the general additive model in Wulumuqi, China. *Int J Infect Dis*. 2020; 96: 42–47. <https://doi.org/10.1016/j.ijid.2020.03.032> PMID: [32200108](#)
29. Rodrigues NCP, Andrade MKN, O'Dwyer G, Flynn M, Braga JU, Almeida AS, et al. Distribution of pulmonary tuberculosis in Rio de Janeiro (Brazil): a spatial analysis. *Cien Saude Colet*. 2017; 22(12): 4125–4134. <https://doi.org/10.1590/1413-812320172212.0143016> PMID: [29267729](#)
30. Collins JJ. The contribution of medical measures to the decline of mortality from respiratory tuberculosis: an age-period-cohort model. *Demography*. 1982; 19(3): 409–27. PMID: [6749567](#)



31. Kiwuwa MS, Charles K, Harriet MK. Patient and health service delay in pulmonary tuberculosis patients attending a referral hospital: a cross-sectional study. *BMC Public Health*. 2005; 5: 122. <https://doi.org/10.1186/1471-2458-5-122> PMID: 16307685
32. Sokal RR, Oden NL. Spatial autocorrelation in biology: 1. Methodology. *Biological journal of the Linnean Society*, 1978, 10(2): 199–228.
33. Koenig WD. Spatial autocorrelation of ecological phenomena. *Trends Ecol Evol*. 1999; 14(1): 22–26. [https://doi.org/10.1016/S0169-5347\(98\)01533-X](https://doi.org/10.1016/S0169-5347(98)01533-X) PMID: 10234243.
34. Boots B, Tiefelsdorf M. Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems*, 2000, 2(4): 319–348.
35. Fotheringham AS. “The Problem of Spatial Autocorrelation” and Local Spatial Statistics. *Geographical Analysis*, 2009, 41(4): 398–403.
36. Huo XN, Zhang WW, Sun DF, Li H, Zhou LD, Li BG. Spatial pattern analysis of heavy metals in Beijing agricultural soils based on spatial autocorrelation statistics. *Int J Environ Res Public Health*. 2011; 8(6): 2074–89. <https://doi.org/10.3390/ijerph8062074> PMID: 21776217
37. Mei CL, Wang N, Zhang WX. Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A*, 2006, 38(3): 587–598.
38. Fotheringham AS, Yang WB, Kang W. Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 2017, 107(6): 1247–1265.
39. Yu HC, Fotheringham AS, Li ZQ, Oshan T, Kang W, Wolf LJ. Inference in multiscale geographically weighted regression. *Geographical Analysis*, 2020, 52(1): 87–106.
40. Fotheringham AS, Yue H, Li ZQ. Examining the influences of air quality in China’s cities using multi-scale geographically weighted regression. *Transactions in GIS*, 2019, 23(6): 1444–1464.
41. Mansour S, Al Kindi A, Al-Said A, Al-Said A, Atkinson P. Sociodemographic determinants of COVID-19 incidence rates in Oman: Geospatial modelling using multiscale geographically weighted regression (MGWR). *Sustain Cities Soc*. 2021; 65: 102627. <https://doi.org/10.1016/j.scs.2020.102627> PMID: 33288993
42. Brunsdon C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 1996, 28(4): 281–298.
43. Brunsdon C, Fotheringham S, Charlton M. Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 1998, 47(3): 431–443.
44. Mennis J. Mapping the results of geographically weighted regression. *The Cartographic Journal*, 2006, 43(2): 171–179.
45. Brunsdon C, Fotheringham AS, Charlton M. Some notes on parametric significance tests for geographically weighted regression. *Journal of regional science*, 1999, 39(3): 497–524.
46. Zhao F, Park N. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transportation Research Record*, 2004, 1879(1): 99–107.
47. Guo L, Ma ZH, Zhang LJ. Comparison of bandwidth selection in application of geographically weighted regression: a case study. *Canadian Journal of Forest Research*, 2008, 38(9): 2526–2534.
48. Lin DY, Dean P.F, Lyle H.U. VIF Regression: A Fast Regression Algorithm for Large Data. *Journal of the American Statistical Association*, 2011, 106(493): 232–247.
49. Yuan WP, Sun H, Chen Y, Xia XC. Spatio-Temporal Evolution and Spatial Heterogeneity of Influencing Factors of SO<sub>2</sub> Emissions in Chinese Cities: Fresh Evidence from MGWR. *Sustainability*, 2021, 13(21): 12059.
50. Zhang Y, Wang XL, Feng T, Fang CZ. Analysis of spatial-temporal distribution and influencing factors of pulmonary tuberculosis in China, during 2008–2015. *Epidemiol Infect*. 2018; 147: e25. <https://doi.org/10.1017/S0950268818002765> PMID: 30303057
51. Du J, Pang Y, Liu Y, Mi F, Xu S, Li L. Survey of tuberculosis hospitals in China: current status and challenges. *PLoS One*. 2014; 9(11): e111945. <https://doi.org/10.1371/journal.pone.0111945> PMID: 25365259
52. Rao HX, Zhang X, Zhao L, Yu J, Ren W, Zhang XL, et al. Spatial transmission and meteorological determinants of tuberculosis incidence in Qinghai Province, China: a spatial clustering panel analysis. *Infect Dis Poverty*. 2016; 5(1): 45. <https://doi.org/10.1186/s40249-016-0139-4> PMID: 27251154
53. Yang R, Liu M, Jiang H, Zhang Y, Yin J, Li Q, et al. The epidemiology of pulmonary tuberculosis in children in Mainland China, 2009–2015. *Arch Dis Child*. 2020; 105(4): 319–325. <https://doi.org/10.1136/archdischild-2019-317635> PMID: 31771943
54. Zhu S, Wu Y, Wang Q, Gao L, Chen L, Zeng F, et al. Long-term exposure to ambient air pollution and greenness in relation to pulmonary tuberculosis in China: A nationwide modelling study. *Environ Res*. 2022; 214: 114100. <https://doi.org/10.1016/j.envres.2022.114100> PMID: 35985487



55. Dimala CA, Kadia BM. A systematic review and meta-analysis on the association between ambient air pollution and pulmonary tuberculosis. *Sci Rep*. 2022; 12(1): 11282. <https://doi.org/10.1038/s41598-022-15443-9> PMID: 35788679
56. Zhang CY, Zhang A. Climate and air pollution alter incidence of tuberculosis in Beijing, China. *Ann Epidemiol*. 2019; 37: 71–76. <https://doi.org/10.1016/j.annepidem.2019.07.003> PMID: 31473123
57. Lai TC, Chiang CY, Wu CF, Yang SL, Liu DP, Chan CC, et al. Ambient air pollution and risk of tuberculosis: a cohort study. *Occup Environ Med*. 2016; 73(1): 56–61. <https://doi.org/10.1136/oemed-2015-102995> PMID: 26514394
58. Wang YS, Chang LC, Chang FJ. Explore Regional PM2.5 Features and Compositions Causing Health Effects in Taiwan. *Environ Manage*. 2021; 67(1): 176–191. <https://doi.org/10.1007/s00267-020-01391-5> PMID: 33201258
59. Mustafa MG, Tierney DF. Biochemical and metabolic changes in the lung with oxygen, ozone, and nitrogen dioxide toxicity. *Am Rev Respir Dis*. 1978; 118(6): 1061–90. <https://doi.org/10.1164/arrd.1978.118.6.1061> PMID: 369411
60. Ortblad KF, Salomon JA, Bärnighausen T, Atun R. Stopping tuberculosis: a biosocial model for sustainable development. *Lancet*. 2015; 386(10010): 2354–62. [https://doi.org/10.1016/S0140-6736\(15\)00324-4](https://doi.org/10.1016/S0140-6736(15)00324-4) PMID: 26515678
61. Ukwaja KN, Alobu I, Nweke CO, Onyenwe EC. Healthcare-seeking behavior, treatment delays and its determinants among pulmonary tuberculosis patients in rural Nigeria: a cross-sectional study. *BMC Health Serv Res*. 2013; 13: 25. <https://doi.org/10.1186/1472-6963-13-25> PMID: 23327613
62. Yadav J, John D, Allarakha S, Menon GR. Rising healthcare expenditure on tuberculosis: Can India achieve the End TB goal? *Trop Med Int Health*. 2021; 26(10): 1256–1275. <https://doi.org/10.1111/tmi.13648> PMID: 34192385
63. Edelson PJ, Phipers M. TB transmission on public transportation: a review of published studies and recommendations for contact tracing. *Travel Med Infect Dis*. 2011; 9(1): 27–31. <https://doi.org/10.1016/j.tmaid.2010.11.001> PMID: 21167784
64. Mathema B, Andrews JR, Cohen T, Borgdorff MW, Behr M, Glynn JR, et al. Drivers of Tuberculosis Transmission. *J Infect Dis*. 2017; 216(suppl\_6): S644–S653. <https://doi.org/10.1093/infdis/jix354> PMID: 29112745
65. Go U, Park M, Kim UN, Lee S, Han S, Lee J, et al. Tuberculosis prevention and care in Korea: Evolution of policy and practice. *J Clin Tuberc Other Mycobact Dis*. 2018; 11: 28–36. <https://doi.org/10.1016/j.jctube.2018.04.006> PMID: 31720389