CrossMark

The Factors Influencing China's Population Distribution and Spatial Heterogeneity: a Prefectural-Level Analysis using Geographically Weighted Regression

Zhibin Xu[1] · Anjiao Ouyang[1]

Abstract The study of population distribution and its influencing factors is one of the key fields within population geography. Most previous studies have used traditional global regression models to explore the factors influencing population distribution, but they neglect spatial heterogeneity. To overcome this weakness, we employed the Geographically Weighted Regression (GWR) model to identify spatially varying relationships between population density and potential influencing factors in mainland China. The results showed that road density, GDP, temperature and arable land proportion were the key factors influencing population distributions and that the influence of each factor varied in different regions. A lower road density significantly restricted population agglomeration in less developed regions such as Southwest China. The regression coefficients of GDP decreased from the more developed Southeast China to the less developed Northwest China. The regression coefficients of temperature were higher in southeastern coastal areas. Arable land proportion was a significant factor increasing population agglomeration in Xinjiang, Northwest China, but this relationship was weaker in other parts of China. We argue that regional population and development policy should be made according to the specific factors crucially influencing population distributions in various regions in order to promote orderly immigration and optimize population distribution.

Keywords Population density · Geographically weighted regression (GWR) · China · Influencing factors · Population geography

✉ Anjiao Ouyang
    ouy219@163.com

  Zhibin Xu
    xzb3232@qq.com; xuzhibin@zju.edu.cn

[1]   Zhejiang University, Zhejiang, China

🖄 Springer

Introduction

The spatial distribution of the human population is a classic and eternal topic in population geography. It was systematically studied by Friedrich Ratzel as early as 1890, and it has been one of the key fields within population geography for over a century (Hu 1935; Mera 1977; Bedford 1999; Beeson et al. 2001; Voss 2007; Feng and Li 2011; Matthews and Parker 2013; Zhu et al. 2016). In earlier studies, many spatial and statistical methods have been employed to analyze population distributions, such as population centers of gravity, Lorenz curves, spatial autocorrelation analysis and multivariate regression (Chi and Zhu 2007). Recent studies have focused on the spatial mechanisms and logic behind population distribution by integrating statistical and GIS spatial analyses (Zhu et al. 2016). Meanwhile, advanced GIS techniques are used to simulate population distribution in order to overcome spatiotemporal resolution limitations in population census data by using various data sources, such as remote sensing data (Lung et al. 2013), land use data (Luo and Wei 2006; Gallego 2010), and mobile phone data (Deville et al. 2014).

Population distribution in China is particularly important because it is the world's most populous country, having a vast territory and huge regional differences. China is undergoing the largest migration in human history, with the population flocking to developed areas in the east of the country (Fan 2008; Ye Liu et al. 2014). Revealing the cause of population distribution and determining the dominant factors that promote or inhibit population growth represent foundational undertakings that provide a scientific basis for creating population policies that promote orderly immigration and optimize population distribution.

The past two decades have witnessed the rapid development of GIS technology and the implementation of China's population census, both of which have provided technical and data support for quantitative research on population distribution and its influencing factors. Many excellent studies have been conducted (Table 1). Some researchers (Dong et al. 2002; Chen et al. 2007; Feng et al. 2008; Y. Fang et al. 2012; Ying Liu et al. 2015) have focused on natural influencing factors. They found that population density was positively correlated with water sources, precipitation, temperature and net primary productivity (NPP) but was negatively correlated with the relief degree of land surface (RDLS), elevation, and distance to rivers and coastlines. The other researchers (Song et al. 2007; Lv et al. 2009; J. Wang et al. 2012; Bai et al. 2015; L. Wang et al. 2015) have examined both natural and socioeconomic factors. Their findings indicated that the influence of socioeconomic factors cannot be ignored. Due to the complexity of these factors, researchers might choose different indicators in different study areas and scales. However, it is commonly believed that complex topography and harsh climate will hinder population agglomeration. Humans prefer to live in places with long development histories, advanced economies and good public services and infrastructure conditions. These findings contributed to the selection of influencing factors in our paper.

Most previous studies (Chen et al. 2007; Song et al. 2007; J. Wang et al. 2012) have used traditional global regression models to test the relationship between population distribution and influencing factors. The global regression model assumes that the relationships between the variables are homogenous. However, spatial dependences often are not homogenous across large geographical regions (Matthews and Parker

2013). For example, water resources are particularly important influencing factors for population distribution in arid areas, but may not be the main factor in humid areas. To illustrate these differences, some researchers (Bai et al. 2015; L. Wang et al. 2015) divided their study areas into several sub-districts then used traditional global regression model in each sub-district. This approach helps, but at the same time isolates each sub-district. The impact between samples in different sub-districts is neglected, especially geographically adjacent samples.

To overcome this weakness, we employed the Geographically Weighted Regression (GWR) model to explore the spatially varying relationships between human population distribution and potential influencing factors in mainland China at the prefectural level. The GWR model is a local spatial statistical technique for exploring spatial nonstationarity (Brunsdon et al. 1996, 1998) and is widely used in spatial statistical analyses of land and house prices (Dziauddin et al. 2015; S. Li et al. 2016), migration (Helbich and Leitner 2009; Jivraj et al. 2013; Villarraga et al. 2014), climate change and carbon emissions (Samson et al. 2011; S. Wang et al. 2014) and health and epidemiology (Black 2014). Thus, the GWR is expected to be used as a powerful tool for identifying the factors influencing China's human population distribution and spatial heterogeneity.

The rest of the study is structured as follows: We first introduce the data and methods. The results indicate that road density, GDP, temperature and arable land

Table 1 Summary of the factors influencing China's human population distribution

Studies	Main influencing factors	Main methods
Dong et al. (2002)	Number of inhabited sites, density of highway/railway/water systems, slope and land cover	Correlation analysis and multivariate analysis/simulation
Chen et al. (2007)	Arable land proportion, precipitation, elevation, temperature and distance from energy distribution location and main rivers	multiple linear regression analysis
Song et al. (2007)	Elevation, river density, slope, road density, roughness, precipitation, temperature	Correlation analysis and stepwise regression analysis
Feng et al. (2008)	Relief degree of land surface	Correlation analysis
Lv et al. (2009)	Climate factors, altitude, per capita GDP, transportation	Spatial autocorrelation analysis
Y. Fang et al. (2012)	Climate factors, topography factors and water system factors	Spatial statistic method and Lorenz curve
J. Wang et al. (2012)	Elevation, slope, annual average sunshine hours, annual rainfall, average temperature, land use and GDP	Spatial autocorrelation analysis and correlation analysis
Ying Liu et al. (2015)	Relief degree of land surface	Correlation analysis
Bai et al. (2015)	Relief degree of land surface, water system density, road system density and nighttime light index	Correlation analysis and stepwise regression analysis
L. Wang et al. (2015)	The density of initial population, hospital beds, telephones, transport routes and GDP	Partial least squares regression

proportion are identified as the key factors influencing population distribution and that the influence of each factor varies in different regions. We argue that regional population and development policies should be made according to the specific factors influencing the population distribution in each region. Finally, we discuss the meaning of our findings for policymaking as well as the limitations in this paper.

Data and Methods

Data

Population density, which is the population size divided by the land area in a region, is used as dependent variable. Due to China's large population and area, the administrative divisions, including the provincial, prefecture, county and township levels, are complicated. The prefecture level was chosen as the analysis scale of this study, because it shows more detailed differences compared to the provincial-level. In addition, data accessibility is another concern. 342 prefecture-level administrative districts in mainland China were studied, including 283 prefecture-level cities, 17 prefectures, 30 autonomous prefectures, 3 leagues, 4 municipalities, and 5 regions directly governed by provinces. Permanent population data were derived from the 2010 Population Census of the PRC. The spatial distribution of the population densities in mainland China in 2010 are presented in Fig. 1.

Twelve explanatory variables were selected (Table 2) according to the previous studies and on the basis of their availability (Chen et al. 2007; Song et al. 2007;
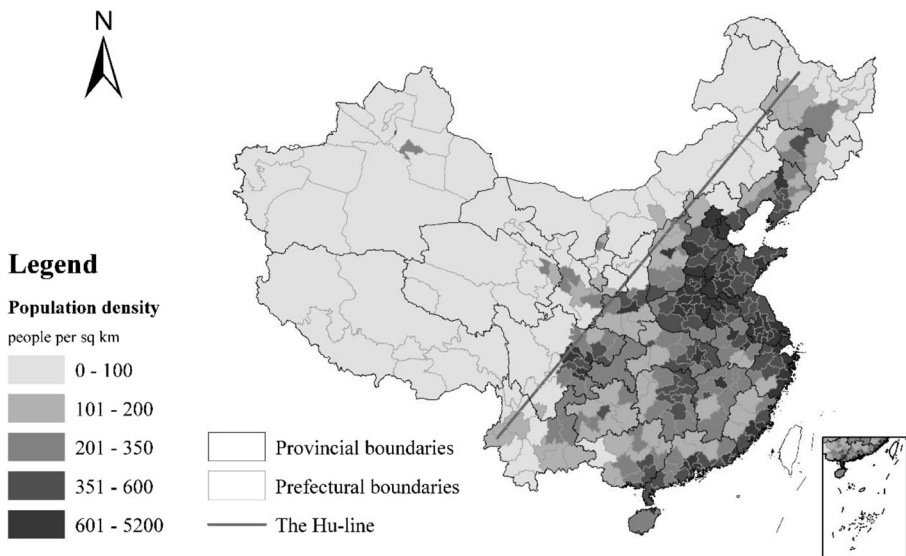

Fig. 1 Population densities of various prefectures in mainland China in 2010. Note: The Hu-line was imagined by the Chinese demographer Hu Huanyong, stretching from the city of Heihe to Tengchong County, in 1935. At that time, land to the east of this line contained 96% of the population, and land to the west contained the remaining 4%. This marks a striking difference in the distribution of China's population at that time, a pattern that still remains after nearly 70 years (Ge and Feng 2008)

Table 2 Potential factors influencing population density and their data sources

Category		Variables	Description	Data source
Natural factors	Topographical factors	Elevation	Mean of elevation	RESDC
		Terrain slope	Mean of terrain slope	
	Climate factors	Precipitation	Average annual precipitation	CNERN
		Temperature	Average annual temperature	
Socioeconomic factors	Development history factors	Arable land proportion	Arable land area divided by total land area	RESDC
	Economic factors	GDP	Gross Domestic Product	NBSC
		Per capita GDP	GDP divided by population	
		Per capita income	Average wage of staff and workers being on duty	
	Public services and infrastructure factors	LGGBE	Local government general budgetary expenditure	
		Educational resources	Number of full-time teachers	
		Medical and health services	Number of licensed doctors and assistant doctors	
		Road density	Mean of road density	OpenStreetMap

RESDC (Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences: http://www.resdc.cn); CNERN (The Chinese National Ecosystem Research Network: http://cnern.cern.ac.cn/en); NBSC (National Bureau of Statistics of China: http://www.stats.gov.cn/english); OpenStreetMap: https://www.openstreetmap.org

Lv et al. 2009; L. Wang et al. 2015). Elevation and terrain slope were selected to indicate topographical factors. Similarly, precipitation and temperature were chosen to indicate climatic factors. As an agriculture-dominated country throughout history, China has witnessed a population agglomeration process accompanied by the process of land reclamation; accordingly, arable land proportion was chosen to indicate the development history. Gross Domestic Product (GDP), per capita GDP and per capita income are used to indicate both the size and level of the economy. The public services and infrastructure in China are mainly provided by the government. Therefore, the local government general budgetary expenditure (LGGBE) was chosen to indicate the general level of public services and infrastructure. In addition, educational resources, medical and health services and road density were considered separately. Both the dependent and independent variables were normalized using the z-scores method to allow for the comparison of corresponding normalized values among different datasets.

Methods

The Pearson's correlation coefficients between the dependent variable and independent variables were calculated. The OLS (ordinary least squares) regression was then conducted to investigate the influence of the 12 independent variables. The stepwise method was conducted in IBM SPSS Statistics 19.0. Considering the classical theory of demographic geography and the results of previous empirical research, the model's suggestions were checked and modified. Finally, the GWR model was used in ESRI ArcGIS 9.3 to explore the spatially varying relationships between population density and the potential influencing factors. As mentioned above, the GWR model improves on the OLS regression model by taking spatial structure into account (Brunsdon et al. 1996, 1998; Fotheringham et al. 1998). The model can be expressed as follows.

$$y_i = \beta_0(\mu_i, v_i) + \sum_k \beta_k(\mu_i, v_i)x_{ik} + \varepsilon_i \tag{1}$$

where.

y_i	is the population density of the ith region;
(u_i, v_i)	is the central spatial coordinates of the ith region;
$\beta_0(u_i, v_i)$	is the local estimated intercept for ith region;
$\beta_k(u_i, v_i)$	is the effect of kth independent variables for ith region;
x_{ik}	is the kth explanatory variable associated with β_k;
ε_i	is a random component assumed to be independently and identically distributed.

Bandwidth is an important parameter of the GWR model. In this paper, the AICc method is adopted to determine the optimal bandwidth (Fotheringham et al. 2002). The Akaike Information Criterion (AIC) (Akaike 1974) was used to compare OLS with GWR. If the AIC is more than 3 units smaller, the smaller one is significantly improved the model fit. Moran's I is used to examine the spatial autocorrelation of the standardized residuals. If Moran's I is close to zero, it means that the residual is randomly distributed and indicates that the model fits well.

Table 3 The Pearson's correlation coefficients between the dependent variable and 12 independent variables

No.	Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Population density													
2	Elevation	−0.390**												
3	Terrain slope	−0.331**	0.673**											
4	Precipitation	0.293**	−0.397**	0.070										
5	Temperature	0.425**	−0.591**	−0.153**	0.822**									
6	Arable land proportion	0.330**	−0.583**	−0.619**	−0.045	0.255**								
7	GDP	0.657**	−0.294**	−0.244**	0.120*	0.197**	0.203**							
8	Per capita GDP	0.343**	−0.235**	−0.322**	−0.098	0.014	0.016	0.515**						
9	Personal income	0.326**	0.269**	0.035	−0.126*	−0.144**	−0.154**	0.543**	0.559**					
10	LGGBE	0.543**	−0.180**	−0.130*	0.074	0.121*	0.127*	0.913**	0.327**	0.485**				
11	Educational resources	0.406**	−0.340**	−0.208**	0.159**	0.296**	0.375**	0.705**	0.076	0.134*	0.690**			
12	Medical and health services	0.511**	−0.311**	−0.203**	0.099	0.191**	0.273**	0.891**	0.298**	0.382**	0.878**	0.852**		
13	Road density	0.852**	−0.410**	−0.337**	0.313**	0.435**	0.312**	0.642**	0.474**	0.390**	0.469**	0.358**	0.473**	

* Sig. <0.05 ** Sig. < 0.01 (two-tailed)

Results

All 12 independent variables were significantly correlated with population density (Table 3). Of these, the correlation coefficients for elevation and terrain slope were negative, while those of the other 10 factors were positive. Those factors with correlation coefficients larger than 0.4 were all socioeconomic factors, such as the road density, GDP, per capita income, LGGBE, and medical and health services. In addition, the correlation coefficients among the independent variables were significant, and some were even larger than 0.8, indicating that multicollinearity should be considered in the next step. We tested for multicollinearity utilizing the variance inflation factor (VIF) (Menard 2002).

The stepwise OLS tests suggested that road density, LGGBE, temperature and arable land proportion were the key factors influencing population distribution. As the classical theory of population migration and many related empirical studies have shown, the major causes of migration are economic (Ravenstein 1885, 1889; Lee 1966; Grigg 1977). In addition, GDP, which is the most basic and commonly used economic indicator in empirical studies (Fan 2005; C. Liu et al. 2007; L. Li and Clarke 2012; Jiao et al. 2016), should be emphatically investigated in this paper. On the other hand, the Pearson correlation between GDP and LGGBE was 0.913, which means that these two indicators contain much of the same information. To prevent a serious model estimate deviation caused by multicollinearity, we eventually replaced LGGBE with GDP. Other selected independent variables were also checked. For example, the road density has the strongest Pearson correlation with population density in this paper. Some studies have also shown that improved transport facilities can reduce travel coasts and bring convenience to residents (Kotavaara et al. 2011; T. Li et al. 2012). So, it plays a very important role in population agglomeration. Finally, road density, GDP, temperature and arable land proportion were chosen as the 4 key factors influencing population distribution. The highest VIF value was 2.10, which was well below the common cut-off point of 10 (Menard 2002), indicating that multicollinearity was not biasing the OLS estimations.

The GWR analysis showed that the bandwidth was 1205 km, the adjusted R2 was 0.788 (larger than OLS's 0.753), the AICc was 459 (lower than OLS's 500) and the standardized residual of Moran's I index was 0.010 (nearly zero), indicating that the GWR had an improved model fit. The regression coefficients from the GWR analysis are summarized in Table 4, and the visualized maps are shown in Fig. 2. The solid white areas indicate that the parameter estimated with a t-value with a significance less than 95%. The shaded areas display where the spatially varying effects were significant. It was found that the influence of each factor on the population distribution varied in different regions. The mean regression coefficient between the population and road density was the largest, and the regional coefficients reached the highest in the Southwest but were lower in Northeast China (Fig. 2a). The regression coefficients between population density and GDP decreased from Southeast to Northwest China (Fig. 2b). The regression coefficients between population density and temperature were higher in southeastern coastal areas but were lower in inland area of China (Fig. 2c). The mean regression coefficient between population density and the arable land proportion was the lowest, and the coefficients in Xinjiang (short for Xinjiang Uyghur Autonomous Region) in Northwest China were most significant (Fig. 2d).

Table 4 Comparison of parameter estimates from global (OLS) and local (GWR) models

a) Global model (OLS)		Estimate	P-value	VIF		
Road density		0.671	0.000	2.103		
GDP		0.198	0.000	1.725		
Temperature		0.078	0.010	1.276		
Arable land proportion		0.061	0.034	1.130		
Adjusted R-Square	0.753					
Akaike Information Criterion	500					
b) Local model (GWR)		Minimum	Lower quartile	Median	Upper quartile	Maximum	Std. Dev.
Road density		0.386	0.507	0.579	0.687	0.868	0.098
GDP		-0.255	0.162	0.197	0.257	0.426	0.096
Temperature		-0.067	0.072	0.150	0.192	0.430	0.102
Arable land proportion		0.049	0.110	0.142	0.161	0.309	0.045
Adjusted R-Square	0.788						
Akaike Information Criterion	459						
Moran's I of Standard Residuals	0.010						

The GWR results of local r-square values are mapped in Fig. 3, ranging from 0.717 to 0.960. Hence, this model fits well in most parts of the country.

Discussion and Conclusion

Explaining Spatially Varying Population Densities

The abovementioned results showed that road density, GDP, temperature and arable land proportion were the 4 key factors influencing population distribution, and that socioeconomic factors influenced population density more significantly than natural ones. This result can also be obtained by traditional global regression method and is roughly the same as many other studies (Lv et al. 2009; L. Wang et al. 2015). However, few studies have been done to estimate the spatial heterogeneity of these influences.

Complex terrain and a lack of transportation infrastructure in Southwest China seriously restrict the economic development and population agglomeration in this region. Therefore, it was found that the regression coefficients between population and road density were the highest in these regions. In contrast, population density was not very sensitive to road density in Northern and Eastern China due to better transportation infrastructure. Y. Zhang and Ren (2012); Tang et al. (2015) have also found that in mountainous areas in China, the better the transportation infrastructure
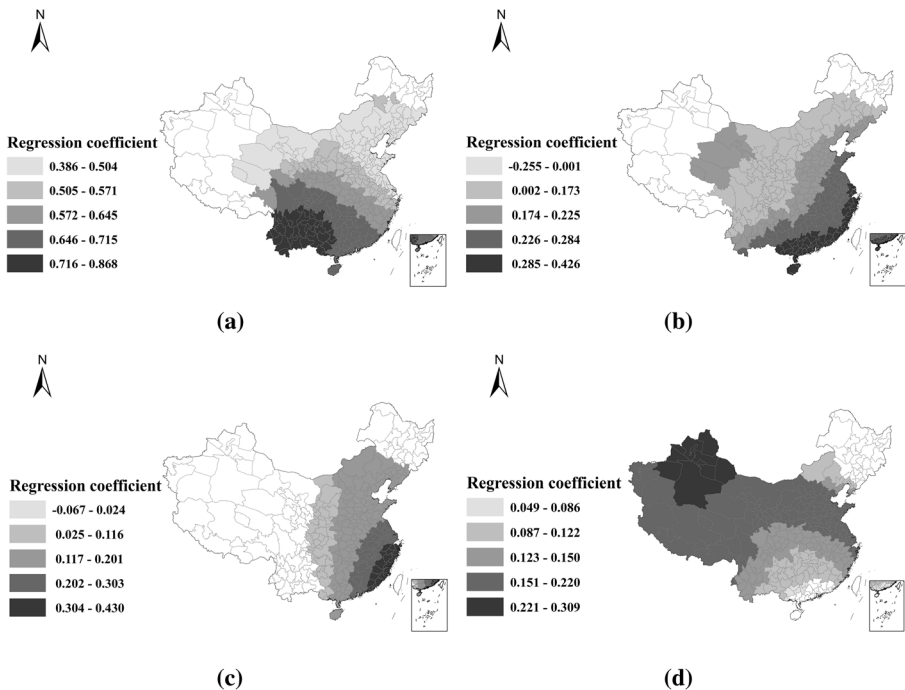

Fig. 2 The GWR regression coefficients between population density and road density **a**, GDP **b**, temperature **c**, and arable land proportion **d**. Note: Significant areas at 95%
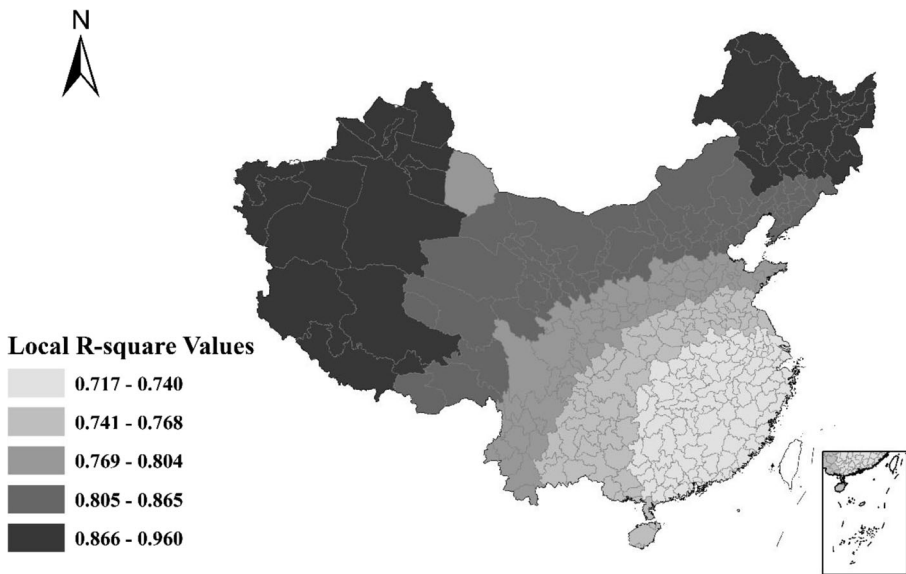
Local R-square Values

- 0.717 - 0.740
- 0.741 - 0.768
- 0.769 - 0.804
- 0.805 - 0.865
- 0.866 - 0.960

Fig. 3 Local r-square values of GWR model

was, the less it influenced population distribution. That is, the influence of road density on population density is marginally decreasing.

The regression coefficients of GDP were almost positive and were higher in southeastern coastal areas than inland. Southeastern coastal areas are some of the more developed areas in China, where there are more employment opportunities, higher wages and better living standards, attracting a large amount of migration from the interior. This result conformed to Ravenstein's classic population migration theory and was also consistent with many scholars' research results, for example, the results of Fan (2002) documenting the rising population in the eastern region as a result of population growth due to migration and in response to the widening economic gap between coastal and interior China.

The regression coefficients of temperature were almost positive and decreased from southeast to northwest. This finding was consistent with Samson et al. (2011), who developed the first global index to predict impacts of climate change on human population density using a GWR model, and found that human population density tended to be positively related to average annual temperature in the high latitudes of the world. However, their research scope was global and the model bandwidth was nearly 3 times that of ours, and therefore did not show the detailed differences that were apparent from our model. In addition, as mentioned above, China has been an agriculture-dominated country throughout history. Agricultural production efficiency is an important factor in population agglomeration. Southeast China has a subtropical monsoon climate zone, and the abundant rainfall and sufficient heat there are favorable to agricultural production. Therefore, the influence of temperature is significant. Northwest China is arid, with an annual precipitation almost under 500 mm. The higher the temperature, the greater the evaporation, which aggravates drought, reducing the suitability for human habitation. In Northeast China, it is too cold for humans to live, but there are abundant forest resources. To exploit these resources, some forestry cities like Yichun have been established and have attracted a certain amount of

population influx (J. Zhao 1992). Southwestern China is hot and humid. Malaria, which is transmitted through mosquito bites, is strongly determined by temperature (X. Zhao et al. 2014) and to some extent hinders the immigration of Han Chinese and affects the population distribution patterns of local ethnic minorities (Cang 2004). Thus, the effects of temperature are not very significant in these areas.

The significant regression coefficient between population density and arable land proportion in Xinjiang (in the extreme northwest) might be due to the Xinjiang Production and Construction Crops (XPCC) project, a special frontier migration project aimed at cultivating wilderness and guarding the frontier (Cappelletti 2015). The XPCC has reclaimed more than 10,000 km^2 of fertile lands since 1954, and now has 176 farming and herding regiments with 2.5 million people (almost 10% of the population of Xinjiang). Therefore, population distribution is highly correlated with the proportion of arable land in Xinjiang.

Implications for Population Policy Making

One of the Chinese central government's long-term population development strategies is to promote orderly migration, reasonable population distributions, and maintain a balance between population, social development, resources and environmental protection (State council of the PRC 2000). The Hu line still remains, and high population density areas are concentrated mainly in the North China Plain, the Yangtze River Delta, the Pearl River Delta and Sichuan Basin. We argue there is no "one size fits all" policy. Regional population and development policies should be made according to the specific factors influencing the population distribution in each region. Our results provide some useful wisdom:

First, different regions have different driving factors that promote or restrict population agglomeration. Therefore, suitable regional population development policies should be made according to different local conditions. For example, poor transport infrastructure conditions are one of the main limiting factors for population agglomeration in southwest. Therefore, accelerating the construction of transport infrastructure is a key factor for enhancing population agglomeration in these areas.

Second, the regression coefficients of GDP were higher in the southeastern coastal areas than inland. This means that, even if the speed of economic growth in the coastal and inland areas were equivalent, the attraction of migrants to coastal areas will be higher than that of the inland areas. That is, the population will continue migrating to coastal areas inertially. To reverse the trend of people flocking to the eastern coastal cities and build a population-balanced society, China must promote the Great Western Development Strategy and speed up the economic development of Midwest China.

Third, food transport has historically cost a great deal; therefore, local land reclamation is an important measure to promote the development and prosperity of a region. However, this relationship has been gradually weakened with the development of food production, storage and transport technology. Our results show that, except in Northwest China, the regression coefficients of arable land proportion are lower. This reminds us to take a new look at the relationship between agricultural development policies and population agglomeration at present and in the future.

We also found an interesting phenomenon: As the capital city, Beijing's GDP ranks second among Chinese cities, only behind Shanghai. However, Beijing's GDP regression coefficient was not particularly high (Fig. 2b). As mentioned above, the GWR model performs local regressions considering not only the region itself but also the

surrounding regions, and the closer the distance the greater the weight. Unlike urban agglomerations in the Yangtze River Delta and Pearl River Delta, the cities in Beijing-Tianjin-Hebei region are polarized (Sun and Yuan 2014). Except for Beijing and Tianjin, other cities in Hebei province are relatively small. Therefore, Beijing's value is lowered by its surrounding cities. Currently, Beijing is suffering a series of "big city diseases", such as population over-expansion, high housing prices, traffic congestion and so on (Xinhua 2015). Our results, from one aspect, demonstrate the importance of regionally coordinated development and support the execution of the Beijing-Tianjin-Hebei coordinated development strategy.[1]

Study Limitations

While this study has several strengths, it is not without limitations. The GWR model is a useful technique for exploring the spatially varying relationships between variables. We successfully revealed the spatial heterogeneity of our variables. However, we cannot indicate that there is strict causality between independent variables and dependent variable by using this method alone. For example, one could ask whether population agglomeration caused the increase of the road density, or the increase of the road density caused the population agglomeration. Though we cited a number of previous researchers' achievements and tried to supplement the explanation of our results, more work is needed to reveal the mechanisms of these influencing factors' effects on population distribution.

Moreover, there are some limitations in the data. Concerning data accessibility, prefecture level units are used in our final analysis. However, population distribution is scale-dependent. The results may be a little different at different scales. Additionally, there are potential concerns about the selection of independent variables. For example, we cannot directly quantify the level of public services and infrastructure, so we elected to use LGGBE for instead.

Third, in a country like China where the government has strong control over resource allocation and macro-economic dynamics, the effects of policy on population distribution cannot be ignored. For example, the household registration system restricts population migration (C. Fang 1995; Y. Wang 2013), and the family planning policy actually varies between regions and ethnic groups (Xu 2010; Z. Zhang 2016; J. Zhang 2017). In addition, some historical "movements" or "national development strategies" may also have had a great influence on population distribution, including the "Up to the Mountains and Down to the Countryside Movement" (Glassman 1978) and the "Third Front Movement" (Meyskens 2015) during the Maoist era, a large number of people and factories moved to Midwest China. During the Chinese Economic Reform in the post-Mao era, people returned to coastal areas. However, it was very difficult to quantify the effects of policy in this paper. A detailed analysis about these issues will be a focus in our further studies.

---

[1] According to this strategy, Beijing will transfer its non-capital functions, such as general manufacturing, downtown wholesale markets as well as some educational and medical services, to Tianjin and Hebei; guide the immigrant population out of Beijing and ease the big city disease; and promote regional coordinated development.

References

Akaike, H. (1974). A new look at the statistical model identification. *IRRR Transactions on Automatic Control, 19*, 716–723. doi:10.1007/978-1-4612-1694-0_16.

Bai, Z., Wang, J., Yang, Y., & Sun, J. (2015). Characterizing spatial patterns of population distribution at township level across the 25 provinces in China. *ACTA GEOGRAPHICA SINICA*(08), 1229–1242.

Bedford, R. (1999). End of an era? Population geography in New Zealand at the turn of the century. *New Zealand Geographer, 55*(2), 8–24. doi:10.1111/j.1745-7939.1999.tb00539.x.

Beeson, P. E., DeJong, D. N., & Troesken, W. (2001). Population growth in U.S. counties, 1840–1990. *Regional Science and Urban Economics, 31*(6), 669–699. doi:10.1016/s0166-0462(01)00065-5.

Black, N. C. (2014). An ecological approach to understanding adult obesity prevalence in the united states: A county-level analysis using geographically weighted regression. *Applied Spatial Analysis and Policy, 7*(3), 283–299. doi:10.1007/s12061-014-9108-0.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis, 28*(4), 281–298. doi:10.1111/j.1538-4632.1996.tb00936.x.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*(3), 431–443. doi:10.1111/1467-9884.00145.

Cang, M. (2004). The influence of malaria to national constitution and distribution in the southwestern frontier. *Journal of the Central University for Nationalities* (02), 37-41.

Cappelletti, A. (2015). Developing the land and the people: Social development issues in Xinjiang Uyghur autonomous region (1999–2009). *East Asia, 32*(2), 137–171. doi:10.1007/s12140-015-9240-8.

Chen, M., Xu, C., & Wang, R. (2007). Key natural impacting factors of China's human population distribution. *Population and Environment, 28*(3), 187–200. doi:10.1007/s11111-007-0041-x.

Chi, G., & Zhu, J. (2007). Spatial regression models for demographic analysis. *Population Research and Policy Review, 27*(1), 17–42. doi:10.1007/s11113-007-9051-8.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., et al. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America, 111*(45), 15888–15893. doi:10.1073/pnas.1408439111.

Dong, C., Liu, J., Zhao, R., & Wang, G. (2002). A discussion on correlation of geographical parameter with spatial population distribution. *Remote Sensing Information, 04*, 61–64.

Dziauddin, M. F., Powe, N., & Alvanides, S. (2015). Estimating the effects of light rail transit (LRT) system on residential property values using geographically weighted regression (GWR). *Applied Spatial Analysis and Policy, 8*(1), 1–25. doi:10.1007/s12061-014-9117-z.

Fan, C. C. (2002). Population change and regional development in China: Insights based on the 2000 census. *Post-Soviet Geography and Economics, 43*(6), 425–442. doi:10.2747/1538-7216.43.6.425.

Fan, C. C. (2005). Modeling interprovincial migration in China, 1985-2000. *Eurasian Geography and Economics, 46*(3), 165–184. doi:10.2747/1538-7216.46.3.165.

Fan, C. C. (2008). *China on the move: Migration, the State, and the household*. London and New York: Routledge.

Fang, C. (1995). The causes, trends and policies of population migration. *Chinese Journal of Population Science, 06*, 8–16.

Fang, Y., Ouyang, Z., Zheng, H., Xiao, Y., Niu, J., Chen, S., et al. (2012). Natural forming causes of China population distribution. *Chinese Journal Appied Ecology, 12*, 3488–3495.

Feng, Z., & Li, P. (2011). Review of population geography in the past century. *Progress in Geography, 30*(2), 131–140.

Feng, Z., Tang, Y., Yang, Y., & Zhang, D. (2008). Relief degree of land surface and its influence on population distribution in China. *Journal of Geographical Sciences, 18*(2), 237–246. doi:10.1007/s11442-008-0237-8.

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

Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment & Planning A, 30*(11), 1905–1927. doi:10.1068/a301905.

Gallego, F. J. (2010). A population density grid of the European Union. *Population and Environment, 31*(6), 460–473. doi:10.1007/s11111-010-0108-y.

Ge, M., & Feng, Z. (2008). Research on China's population distribution in 2000 based on GIS method: Compared with Hu Huanyong's study in 1935. *Population Research* (01), 51-57.

Glassman, J. (1978). Up to the mountains and down to the villages: The transfer of youth from urban to rural China. *American Political Science Review, 72*(04), 1448–1449. doi:10.2307/1954614.

Grigg, D. B. (1977). E. G. Ravenstein and the laws of migration. *Journal of Historical Geography, 3*(1), 41–54. doi:10.1016/0305-7488(77)90143-8.

Helbich, M., & Leitner, M. (2009). Spatial analysis of the urban-to-rural migration determinants in the Viennese metropolitan area. A transition from suburbia to Postsuburbia? *Applied Spatial Analysis and Policy, 2*(3), 237–260. doi:10.1007/s12061-009-9026-8.

Hu, H. (1935). Distribution of China's population. *ACTA GEOGRAPHICA SINICA* (02), 33–74.

Jiao, J., Wang, J., Jin, F., & Du, C. (2016). Understanding relationship between accessibility and economic growth: A case study from China 1990-2010. *Chinese Geographical Science, 26*(6), 803–816. doi:10.1007/s11769-016-0831-0.

Jivraj, S., Brown, M., & Finney, N. (2013). Modelling spatial variation in the determinants of neighbourhood family migration in england with geographically weighted regression. *Applied Spatial Analysis and Policy, 6*(4), 285–304. doi:10.1007/s12061-013-9087-6.

Kotavaara, O., Antikainen, H., & Rusanen, J. (2011). Population change and accessibility by road and rail networks: GIS and statistical approach to Finland 1970-2007. *Journal of Transport Geography, 19*(4), 926–935. doi:10.1016/j.jtrangeo.2010.10.013.

Lee, E. S. (1966). A theory of migration. *Demography, 3*(1), 47. doi:10.2307/2060063.

Li, T., Cao, X., & Huang, X. (2012). The relationship between spatial structure of accessiblity and population change in Pearl River Delta. *Geographical Research* (09), 1661–1672.

Li, L., & Clarke, K. C. (2012). Cartograms showing China's population and wealth distribution. *Journal of Maps, 8*(3), 320–323. doi:10.1080/17445647.2012.722792.

Li, S., Ye, X., Lee, J., Gong, J., & Qin, C. (2016). *Spatiotemporal analysis of housing prices in*. China: A big data perspective. Applied Spatial Analysis and Policy. doi:10.1007/s12061-016-9185-3.

Liu, Y., Deng, W., & Song, X.-q. (2015). Relief degree of land surface and population distribution of mountainous areas in China. *Journal of Mountain Science, 12*(2), 518–532. doi:10.1007/s11629-013-2937-5.

Liu, C., Otsubo, K., Wang, Q., Ichinose, T., & Ishimura, S. (2007). Spatial and temporal changes of floating population in China between 1990 and 2000. *Chinese Geographical Science, 17*(2), 99–109. doi:10.1007/s11769-007-0099-5.

Liu, Y., Stillwell, J., Shen, J., & Daras, K. (2014). Interprovincial migration, regional development and State policy in China, 1985–2010. *Applied Spatial Analysis and Policy, 7*(1), 47–70. doi:10.1007/s12061-014-9102-6.

Lung, T., Lübker, T., Ngochoch, J. K., & Schaab, G. (2013). Human population distribution modelling at regional level using very high resolution satellite imagery. *Applied Geography, 41*, 36–45. doi:10.1016/j.apgeog.2013.03.002.

Luo, J., & Wei, Y. (2006). Population distribution and spatial structure in transitional Chinese cities: A study of Nanjing. *Eurasian Geography and Economics, 47*(5), 585–603. doi:10.2747/1538-7216.47.5.585.

Lv, C., Fan, J., & Sun, W. (2009). Population distribution and the influencing factors based on ESDA. *Economic Geography* (11), 1797–1802, doi:10.15957/j.cnki.jjdl.2009.11.003.

Matthews, S. A., & Parker, D. M. (2013). Progress in spatial demography. *Demographic Research, 28*, 271–312. doi:10.4054/DemRes.2013.28.10.

Menard, S. (2002). *Applied logistic regression analysis*. London: Sage.

Mera, K. (1977). The changing pattern of population distribution in Japan and its implications for developing countries. *Habitat International, 2*(5–6), 455–479. doi:10.1016/0197-3975(77)90020-0.

Meyskens, C. (2015). Third front railroads and industrial modernity in late Maoist China. *Twentieth-Century China, 40*(3), 238–260. doi:10.1179/1521538515z.00000000068.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London, 48*(2), 167. doi:10.2307/2979181.

Ravenstein, E. G. (1889). The laws of migration. *Journal of the Royal Statistical Society, 52*(2), 241. doi:10.2307/2979333.

Samson, J., Berteaux, D., McGill, B. J., & Humphries, M. M. (2011). Geographic disparities and moral hazards in the predicted impacts of climate change on human populations. *Global Ecology and Biogeography, 20*(4), 532–544. doi:10.1111/j.1466-8238.2010.00632.x.

Song, G., Li, Z., Bao, Y., Lü, H., Gao, J., Wang, H., et al. (2007). Spatial distribution regularity and influence factors of population density in the LRGR. *Chinese Science Bulletin, 52*(S2), 90–97. doi:10.1007/s11434-007-7027-z.

State council of the PRC (2000). China's Population and Development in the 21st Century. http://news.xinhuanet.com/zhengfu/2002-11/18/content_633168.htm.

Sun, J., & Yuan, Q. (2014). Strategic comparison of joint development of Beijing, Tianjin and Hebei. *Comparative Economic and Social Systems* (05), 1-11.

Tang, N., Wei, D., Lv, Y., Wang, Z., & Long, D. (2015). Factors affecting population distribution in Qinba mountain and zoning development research: A case study of Ankang city in Shaanxi. *Northwest Population Journal* (01), 111–116.

Villarraga, H. G., Sabater, A., & Módenes, J. A. (2014). Modelling the spatial nature of household residential mobility within municipalities in Colombia. *Applied Spatial Analysis and Policy, 7*(3), 203–223. doi:10.1007/s12061-014-9101-7.

Voss, P. R. (2007). Demography as a spatial social Science. *Population Research and Policy Review, 26*(5–6), 457–476. doi:10.1007/s11113-007-9047-4.

Wang, Y. (2013). On changes of China's contemporary population migration policy: Based on the 1950-1990s era. *Journal of Renmin University of China, 05*, 103–111.

Wang, S., Fang, C., Ma, H., Wang, Y., & Qin, J. (2014). Spatial differences and multi-mechanism of carbon footprint based on GWR model in provincial China. *Journal of Geographical Sciences, 24*(4), 612–630. doi:10.1007/s11442-014-1109-z.

Wang, L., Feng, Z., & Yang, Y. (2015). The change in population density from 2000 to 2010 and its influencing factors in China at the county scale. *Journal of Geographical Sciences, 25*(4), 485–496. doi:10.1007/s11442-015-1181-z.

Wang, J., Yang, X., & Shi, R. (2012). Spatial distribution of the population in Shandong province at multi-scales. *Progress in Geography* (02), 176–182.

Xinhua. (2015). *Coordinated momentum needed for area surrounding capital.* http://www.chinadaily.com.cn/m/tianjin2012/2015-01/27/content_19432334.htm.

Xu, J. (2010). Analysis of Chinese population policies' effect. Jilin University.

Zhang, Z. (2016). *Study on the diversities of Chinese fertility policy.* Jilin University.

Zhang, J. (2017). The evolution of China's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives, 31*(1), 141–160. doi:10.1257/jep.31.1.141.

Zhang, Y., & Ren, Z. (2012). Factors affecting population distribution in mountainous areas: Geographically weighted regression using data from Bijie. *Population Research* (04), 53–63.

Zhao, J. (1992). Population and sustainable development of social-economic-natural complex ecosystem: System analysis and regulation strategies of the population of yichun city. *Acta Ecologica Sinica, 01*, 77–83.

Zhao, X., Chen, F., Feng, Z., Li, X., & Zhou, X.-H. (2014). Characterizing the effect of temperature fluctuation on the incidence of malaria: An epidemiological study in south-west China using the varying coefficient distributed lag non-linear model. *Malaria Journal, 13*(1), 192. doi:10.1186/1475-2875-13-192.

Zhu, Y., Ding, J., Wang, G., Shen, J., Lin, L., & Ke, W. (2016). Population geography in China since the 1980s: Forging the links between population studies and human geography. *Journal of Geographical Sciences, 26*(8), 1133–1158. doi:10.1007/s11442-016-1319-7.