Journal of Geography and Cartography (2020) Volume 3 Issue 1: Special Issue - Man-land research by analyzing and mapping geographic phenomena using cartographic methods in China
doi:10.24294/jgc.v3i1.1305

ORIGINAL RESEARCH ARTICLE

Study on the distribution pattern and influencing factors of shrinking cities in Northeast China based on the random forest model

Guanghua Yan[1,2*], Xi Chen[2,3], Yun Zhang[4]

[1] *School of Public Administration at Jilin University, Changchun 130012, Jilin Province, China. E-mail: GhuaYan11@163.com*

[2] *School of Geographical Sciences, Changchun Normal University, Changchun 130032, Jilin Province, China.*

[3] *Nanjing Institute of Geography & Limnology Chinese Academy of Sciences / Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, Jiangsu Province, China.*

[4] *National Marine Environmental Monitoring Center, Dalian 116023, Liaoning Province, China.*

ABSTRACT

Based on the population change data of 2005–2009, 2010–2014, 2015–2019 and 2005–2019, the shrinking cities in Northeast China are determined to analyze their spatial distribution pattern. And the influencing factors and effects of shrinking cities in Northeast China are explored by using multiple linear regression method and random forest regression method. The results show that: 1) In space, the shrinking cities in Northeast China are mainly distributed in the "land edge" areas represented by Changbai Mountain, Sanjiang Plain, Xiaoxing'an Mountain and Daxing'an Mountain. In terms of time, the contraction center shows an obvious trend of moving northward, while the opposite expansion center shows a trend of moving southward, and the shrinking cities gather further; 2) in the study of influencing factors, the results of multiple linear regression and random forest regression show that socio-economic factors play a major role in the formation of shrinking cities; 3) the precision of random forest regression is higher than that of multiple linear regression. The results show that per capita GDP has the greatest impact on the contraction intensity, followed by the unemployment rate, science and education expenses and the average wage of on-the-job workers. Among the four influencing factors, only the unemployment rate promotes the contraction, and the other three influencing factors inhibit the formation of shrinking cities to various degrees.

Keywords: Shrinking Cities; Northeast China; Demographic Changes; Linear Regression

ARTICLE INFO

Received: 3 June 2020
Accepted: 30 July 2020
Available online: 6 August 2020

1.Introduction

In the history of random forest, the cases of urban decline and extinction are not rare, and the reasons are mostly from wars, diseases and disasters. Since modern times, the western world once ushered in the wave of high-speed urbanization. However, after World War II, especially in the 1960s and 1970s, due to the influence of many factors such as suburbanization, deindustrialization, post socialist society, globalization and population development transformation, many cities and regions that carried out urbanization earlier experienced serious population loss[1]. In western academic circles, this urban phenomenon characterized by population reduction and economic recession is called urban shrinkage, and the cities experiencing urban shrinkage are called shrinking cities. The German government funded project "shrinking city research" found that in the past 50 years, 450 urban areas with a population of more than 1 million in the world have lost 1/10 of their total urban population[2]. Such a large-scale and substantial reduction of urban population is unprecedented. It is not caused by war, disease and disasters as people generally know. It occurs in the era of prosperity and peace[3]. In order to deal with the ill effects of urban contraction, foreign academic circles have already carried out research on urban contraction[4], which mainly focuses on the definition[5], classification[6], comparison[7], quantitative analysis[8], causes[9], countermeasures and suggestions of shrinking cities[10,11]. In China, due to the late start of China's urbanization and the rapid urbanization after the reform and opening up, the domestic academic circles mainly focused on growth and expansion[12]. However, with China's social and economic development entering the adjustment period, the decline of fertility rate and population aging, coupled with the proposal of "new urbanization" and other strategies, issues other than "growth" have attracted attention[13]. In this context, the phenomenon of urban contraction has gradually attracted the attention of domestic scholars[14,15]. The research shows that China's cities are also facing a more profound problem of urban contraction. Urban contraction has also occurred not only in the central and western regions, but also in the Yangtze River Delta, Pearl River Delta, Beijing Tianjin Hebei and other areas with better socio-economic conditions, in which the northeast is the hardest hit area of urban contraction[16,17].

At present, the domestic academic circles have a deep understanding of the phenomenon of urban contraction, but most studies still focus on the identification and judgment of shrinking cities, and the methods adopted simply rely on the construction of index system, without combining with the background of regional social and economic development[18]. Some scholars have made quantitative analysis on the deep causes of the formation of shrinking cities[19–21]. However, most of these scholars use linear analysis models with many restrictions and poor explanatory power in the face of complex nonlinear problems. The nonlinear stochastic forest model can overcome these shortcomings. The research shows that the random forest model has high prediction accuracy. It can process high-dimensional and multicollinearity data, and is convenient to calculate the nonlinear effect of variables[22].

Taking the northeast region, which is typical in the wave of urban contraction, as an example, this paper puts forward the judgment method of shrinking cities under the background of large changes in the overall population of the region, and analyzes the spatial distribution pattern of shrinking cities in the northeast region by using the method of spatial statistics. Multiple linear regression method and random forest regression method are used to analyze the influencing factors and their effects on the formation of shrinking cities in this area, in order to provide scientific basis and decision support for the research and management of shrinking cities.

2. Data sources and research methods

2.1 Data sources

Urban growth and contraction is a long gradual process, which requires a long time to observe the growth or contraction tendency of the city. Therefore, the time span should not be too short when observing[23,24]. Therefore, this study takes 5 years as a time unit interval, and selects 2005–2009, 2010–2014 and 2015–2019 as the basic observation time unit respectively. The time span of 5a is proper. On the one hand, the change of urban growth or contraction characteristics are obvious. On the other hand, it can also avoid the risk of increasing the difficulty and inaccuracy of data acquisition due to a long time span.

This study takes the prefecture level administrative regions in Northeast China as the basic research unit. Considering the adjustment of administrative divisions in some regions from 2005 to 2019, taking the administrative divisions in 2019 as the benchmark, 40 cities (including prefecture level cities, regions, autonomous prefectures and leagues) in Northeast China are finally determined as samples, including 14 in Liaoning Province, 9 in

Jilin Province (including prefecture level cities and autonomous prefectures), 13 in Heilongjiang Province (including prefecture level cities and regions), 4 in Mongolia (including prefecture level cities and League) (**Figure 1**). The socio-economic data comes from the statistical yearbooks and statistical bulletins of provinces and cities in each year. The nighttime lighting data comes from the National Geophysical Data Center of the United States (https://www.ngdc.noaa.gov); the land use data, DEM data, meteorological data and NDVI data used in this paper are from the Resource and Environment Science and Data Center (http://www.resdc.cn).

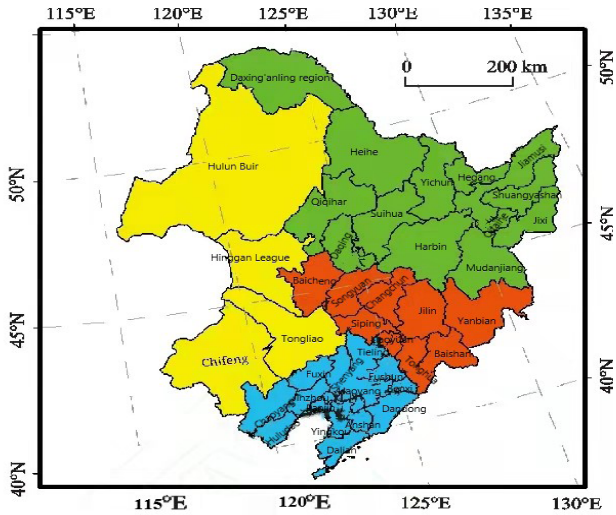

Figure 1. Sketch of study area.

2.2 Judgment method of shrinking cities

At present, there is no unified theory on the definition of shrinking cities at home and abroad[25,26]. Although there are still differences in the quantitative judgment of shrinking Cities in the world, the judgment of shrinking Cities based on population reduction has been widely recognized by the academic circles.

Combined with the background of China's population growth and population loss in Northeast China in the past ten years, this study defines the shrinking cities in Northeast China as: the population growth rate in a period of time (5 years in a short period and 15 years in a long period) is lower than that in Northeast China. The calculation formula is as follows:

$$D = \Delta p - \Delta P \quad (1)$$

In the formula, D is the difference between the population change rate of a single city and the population change rate of the whole northeast in this period. Δp is the population change rate of a single city during this period. ΔP is the population change rate of the whole Northeast during this period. The population change rate is calculated by the total registered residence population at the end of the year. To a certain extent, formula (1) can also reduce the deviation of the total household population at the end of the year. Based on the formula (1), the intensity of the urban population is calculated:

$$I = \frac{D}{|\Delta P|} \quad (2)$$

In the formula, I is the contraction intensity of a single city in this period. When $I < 0$, it means that the city is a shrinking city. The smaller the value is, the greater the population contraction degree of the city is.

2.3 Analysis method of shrinking urban distribution pattern

The standard deviation ellipse can be used to reveal the spatial distribution characteristics of geographical objects on the overall level and reflect the trend direction of geographical objects[27]. This study uses the standard deviation ellipse analysis to explore the centers of three shrinking cities and expanding cities in a short period of time, so as to obtain the moving path of population contraction and expansion.

Spatial autocorrelation analysis refers to the statistical correlation between certain attribute values of geographical elements in different geographical spatial distribution locations. The closer the distance is, the greater the correlation between the two values is[28]. This study uses spatial autocorrelation analysis to explore the spatial agglomeration characteristics of shrinking cities and expanding cities in four periods.

2.4 Analysis method of influencing factors of urban contraction

2.4.1 Selection of influencing factors

This study takes 5 years as a time unit. The selection of socio-economic factors adopts the starting year data, and fully considers the lag of urban contraction. In addition, this study attempts to take into account natural and human factors, and selects 24 influencing factors in combination with relevant literature (**Table 1**)[29–30].

Table 1. Influencing factors of shrinking cities

Influencing factors	Variable	Data description
Total population at the end of the year /person	X_1	Total population at the end of the starting year of the unit period
Birth rate/‰	X_2	Annual birth rate at the beginning of unit period
Proportion of population over 60/%	X_3	Proportion of population over 60 years old in the starting year of unit period
Population density/(person/km2)	X_4	Total population at the end of the starting year of the unit period / administrative area
Night light index	X_5	Night light index of starting year in unit period
Per capita GDP/10000 CNY	X_6	GDP in the starting year of unit period / total population at the end of the starting year of unit period
Investment in fixed assets/10000 CNY	X_7	Investment in fixed assets in the initial year of the unit period
Science education cost/10000CNY	X_8	Initial annual science education expenses in the unit period
Student teacher ratio/%	X_9	Number of primary and secondary school students in the starting year of the unit period / total number of primary school teachers in the starting year of the unit period
Unemployment rate/%	X_{10}	Initial annual unemployment rate in unit period
Per capita number of medical and health beds/piece	X_{11}	Number of beds in medical and health institutions in the starting year of the unit period / total population at the end of the starting year of the unit period
Number of buses/trams per capita/vehicle	X_{12}	Number of buses / trams per capita in the starting year of the unit period
Greening rate of built-up area/%	X_{13}	Greening rate of built-up area in the initial year of unit period
Paved road area per capita/km2	X_{14}	Road pavement area (length) in the starting year of unit period / total population at the end of the starting year of unit period
Average salary of on-the-job employees/CNY	X_{15}	Average wage of on-the-job employees in the starting year of the unit period
Waste utilization index	X_{16}	Average value of comprehensive utilization rate of industrial solid waste, urban domestic sewage treatment rate and harmless treatment rate of domestic waste in the initial year of the bit period
Shape fluctuation	X_{17}	Average topographic relief (grid space ratio is 90 m, and the window is 0.9 km)
Average slope/(°)	X_{18}	Average slope (grid spatial resolution of 90 m)
Average altitude/m	X_{19}	Average altitude (grid spatial resolution of 90 m)
Annual average temperature/°C	X_{20}	Average temperature (using spatial interpolation method)
Annual average precipitation/mm	X_{21}	Average precipitation (using spatial interpolation method)
Cultivated land index	X_{22}	Cultivated land index in the initial year of unit period (grid spatial resolution of 1 km)
Artificial surface index	X_{23}	Artificial surface index in the initial year of unit period (grid spatial resolution of 1 km)
NDVI	X_{24}	Normalized vegetation index (grid spatial resolution of 1 km) for the starting year of the unit period

2.4.2 Regression model and test

Taking the influencing factors in **Table 1** as independent variables and the calculated shrinkage intensity as dependent variables, multiple linear regression and random forest regression methods are used for fitting verification to explore the influencing factors of urban shrinkage in Northeast China.

In multiple linear regression, in order to eliminate the influence of irrelevant or low correlation influencing factors on the regression results, it is necessary to conduct correlation analysis between each independent variable and dependent variable[31]. During the construction of the regression model, the correlation between the respective variables is tested by the variance expansion coefficient VIF. The larger the VIF is, the more serious multicollinearity between the independent variable and other independent variables is. The result of the model is not significant, so the model should be reconstructed.

Random forest model has randomness in the selection of samples and features. It has obvious and unique advantages in algorithm, good tolerance for noise. It is not easy to over fit. Now it is widely used in classification and regression problems[31–33]. In this study, the sklearn module is called on the Python platform for random forest regression. After many debugging, the parameter n_ estimators = 800, criterion = MSE is selected, other parameters are default.

In this study, judgment coefficient R^2 and mean square error MSE are used to evaluate the accuracy of regression model. Judgment coefficient R^2 is the fitting degree of fitting value after regression to observed value, and its value range is [0,1]. The closer R^2 is to 1, the better the fitting effect of the regression model is. The mean square error MSE is the average value of the sum of squares of residuals, which is one of the most commonly used loss functions in regression. The clearer MSE is, the smaller the model error is.

3. Spatial distribution pattern of shrinking cities in Northeast China

3.1 Overall distribution pattern

Calculate the shrinkage intensity of each city in each period according to formulas 1 and 2, and draw the polar coordinate map corresponding to the four periods (**Figure 2**), and visualize the shrinkage cities ($I < 0$) in each period on the map (**Figure 3**). As can be seen from **Figure 2**, Shenyang, Dalian, Daqing and Changchun were prominent in the four periods, while Jixi, Yichun and Daxinganling areas were at a low level for a long time. **Figure 3** shows that the number of shrinking cities was 24 from 2005 to 2009 and 26 from 2010 to 2014, and the spatial distribution of shrinking cities in the two periods is similar, that is, they are distributed in the east of Northeast China. The northern and western regions roughly correspond to areas with high terrain and large topographic relief such as Changbai Mountain area, Xiaoxing'an Mountain area and Daxing'an Mountain area, while cities dominated by population growth show an "inverted triangle". It is distributed along Harbin Dalian railway with relatively flat terrain and good traffic conditions and in central and southern Liaoning. From 2015 to 2019, the number of shrinking cities was 22, 13 cities in Heilongjiang Province were identified as shrinking cities, and only Changchun and Jilin in Jilin Province were not identified as shrinking cities. During this period, shrinking cities are mainly distributed in the North and East. For three cities in a short period, **Figure 4** shows that only Huludao, Panjin, Yingkou, Dalian, Changchun and Daqing have not experienced contraction. Most shrinking cities are located on the "land edge" of Northeast China. In the eastern Changbai Mountain

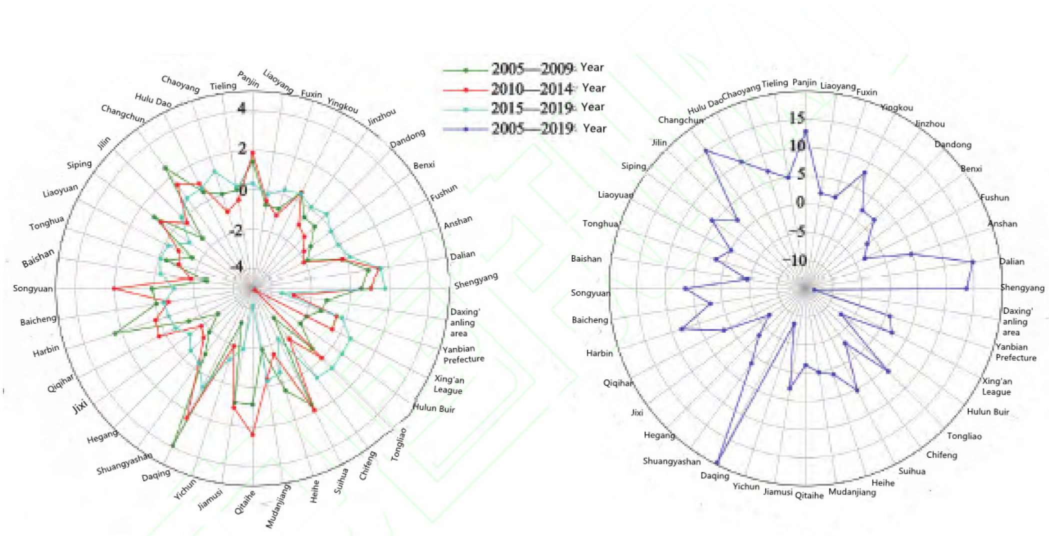

Figure 2. Polar map of shrinkage intensity I in Northeast China.

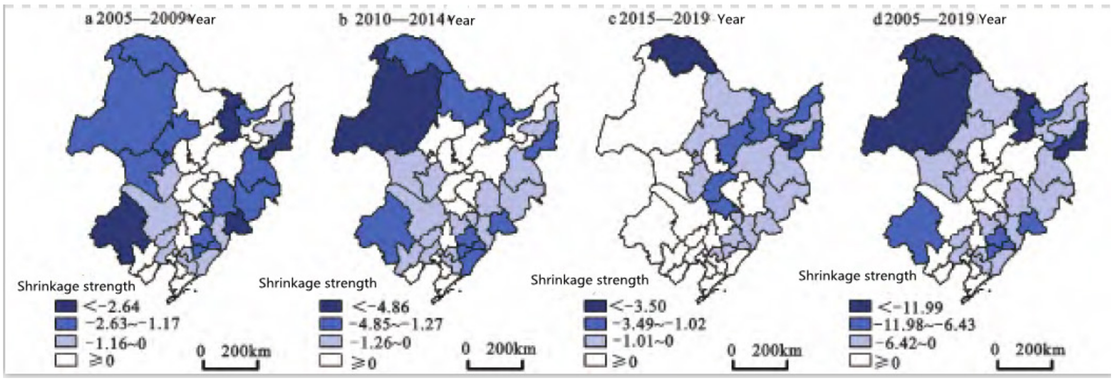

Figure 3. Shrinking cities in Northeast China in 2005-2019.

area, the number of urban contraction is generally large. The large topographic relief and low level of traffic accessibility may be the reasons for its contraction.

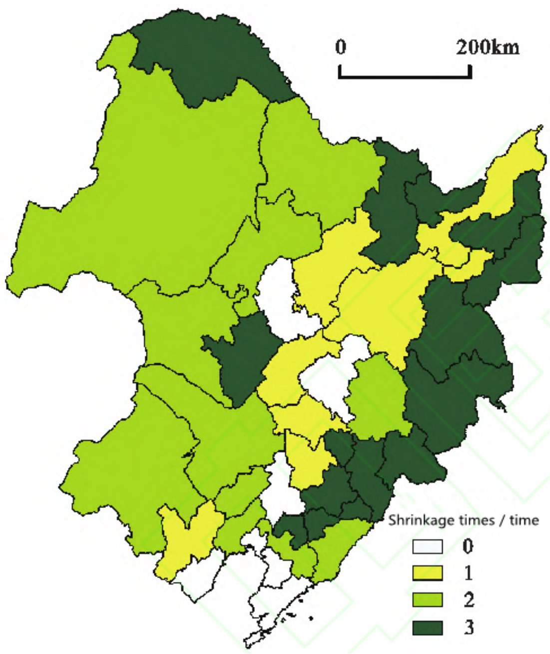

Figure 4. Urban shrinkage in Northeast China in 2005-2009, 2010-2014 and 2015-2019.

3.2 Analysis of spatial agglomeration characteristics

Taking the shrinkage intensity of three short-term shrinking cities ($I < 0$) and expanding cities ($I \geq 0$) as the weight, the standard deviation ellipse and its center of shrinking cities and expanding cities are obtained (**Table 2**). From 2005 to 2009, the shrinking city center was located in Dehui, Changchun City, and the expanding city center was located in Nong'an County, Changchun City. From 2010 to 2014, the shrinking city center was located in Mongolian Autonomous County of Qian Gorlos, Songyuan City, and the expanding city center was located in the jurisdiction of Changchun City. From 2015 to 2019, the shrinking city center was located in Tonghe County, Harbin City, and the expanding city center was located in Zhangwu County, Fuxin City. The shrinking urban ellipse and its center move northward obviously, and the expanding urban ellipse and its center move southward obviously. Generally speaking, in these three periods, the spatial transfer direction of urban contraction is northward and to the north of Heilongjiang Province; while the transfer direction of expansion is southward and to the Bohai Bay.

The standard deviation ellipse reveals the temporal and spatial distribution characteristics of shrinking cities from a certain level, but it does not fully show the agglomeration law among the research units. In order to fully reveal the spatial correlation of the shrinkage intensity of cities in Northeast China in various periods, this study uses the spatial autocorrelation method for research and analysis. The results show that from 2005 to 2009, 2010 to 2014 and 2015 to 2019, Moran's I were 0.132, 0.234, 0.241 and 0.184, and all passed the 0.1 significance test, which shows that there is a significant spatial positive correlation between urban population growth and reduction in Northeast China, and there is a trend of further strengthening.

Further local spatial autocorrelation calculation

Table 2. Parameters of standard deviational ellipse of shrinkage intensity I

Period	Types	Standard deviation ellipse parameter			
		Central coordinates	X-axis length / km	Y-axis length / km	Rotation angle
2005–2009	Shrinking City	44°37′N，125°55′E	933.55	718.78	43°50′
	Expanding City	44°40′N，125°32′E	873.21	332.56	22°49′
2010–2014	Shrinking City	44°45′N，124°52′E	1 007.42	817.51	179°27′
	Expanding City	44°11′N，125°23′E	1 002.49	356.46	27°43′
2015–2019	Shrinking City	46°2′N，128°45′E	686.51	417.75	47°8′
	Expanding City	42°37′N，122°10′E	788.28	394.90	171°48′

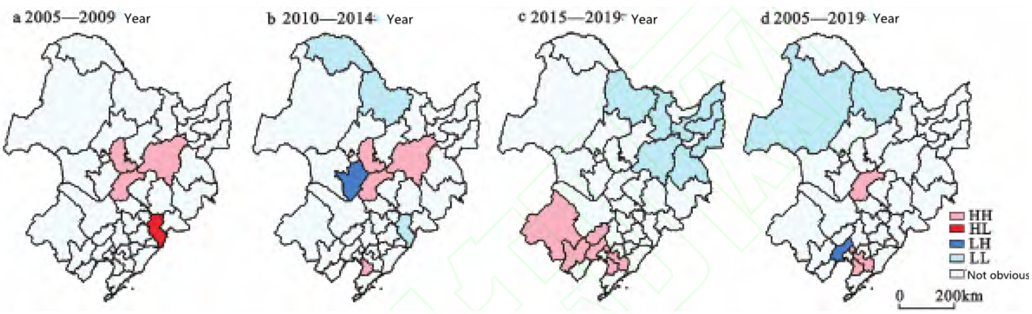

Figure 5. Spatial agglomeration characteristics of shrinking cities in Northeast China.

is carried out to obtain the LISA map of shrinking cities in Northeast China (**Figure 5**). On the map, the distribution characteristics of shrinking cities in Northeast China are divided into five levels in each period, that is, insignificant, “High-High” aggregation area (HH area), “High-Low” aggregation area (HL area), “Low-High” aggregation area (LH area) and “Low-Low” aggregation area (LL area). The results show that from 2005 to 2009, the areas with Harbin, Changchun and Daqing as the core formed HH area, that is population growth gathering area, while Tonghua formed HL area, that is contraction surrounding growth area. From 2010 to 2014, HH expanded to the coastal area of Liaoning with Yingkou as the core. LL area is mainly distributed in the northwest of Heilongjiang Province with Daxinganling area and Heihe River as the core. From 2015 to 2019, there is an obvious north-south differentiation, which is mainly manifested in that the southern area with the central and Western Liaoning Province and Chifeng City as the core is HH area, that is, population growth agglomeration area. And the northern and northeast areas with Sanjiang Plain as the core are LL area, that is, population contraction agglomeration area. From 2005 to 2019, HH district was mainly distributed in the central part of Jilin Province with Changchun as the core and the central and southern part of Liaoning Province with Yingkou as the core, while LL district was distributed in the northern part with Hulunbuir and Heihe as the core.

4. Analysis on influencing factors of urban contraction in Northeast China

4.1 Verification of multiple linear regression and random forest regression

Through the correlation analysis between the selected influencing factors and the shrinkage intensity of each city in each period, 16 of the 24 influencing factors passed the 0.01 significance test, of which 13 had a large correlation with the shrinking city (r ≥ 0.3). Take these 13 influencing factors as independent variables for preliminary linear regression, and eliminate the influencing factors with high VIF (VIF ≥ 10). For the remaining 10 influencing factors, the

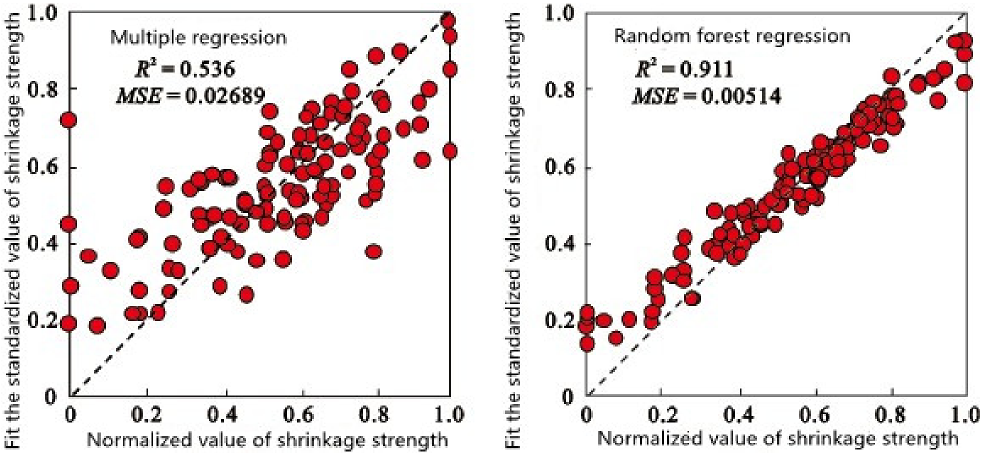

Figure 6. Fitted scatter plots of multiple linear regression and random forest regression.

second multiple linear regression was carried out again. The absolute values of the coefficients obtained by the second multiple linear regression are unemployment rate (−0.304), per capita GDP (0.266), science and education expenses (0.206), cultivated land index (0.195), average wage of on-the-job employees (0.173) and fixed asset investment (0.111), average altitude (−0.107), total population at the end of the year (0.057), topographic relief (0.024) and greening rate of built-up area (0.015).

Input the above 10 influencing factors into the multiple linear regression model and input all influencing factors into the random forest model to calculate the R2 and MSE with multiple linear regression and random forest regression respectively. Then we obtain the corresponding scatter diagram for the comparison between fitting values and observed values (**Figure 6**). The results show that R2 of multiple linear regression is 0.536 and MSE is 0.02689; R2 of random forest regression is 0.911 and MSE is 0.00514. The error of random forest regression is much less than that of multiple linear regression. Random forest has higher accuracy in this study and its results are more trustworthy.

4.2 Importance ranking and analysis of influencing factors

The importance ranking of influencing factors under random forest regression is obtained by using the IncMSE method. IncMSE is the average reduction value of precision. This method follows the principle of control variables. It assigns random values to each variable for many times and carries out prediction and fitting on the original model, and then calculates the MSE between the fitting results and the observation results. The more MSE increases, the more important the influencing factor is. The ranking results of IncMSE show that the top four influencing factors are per capita GDP (17.619%), unemployment rate (13.179%), science and education expenses (11.003%) and average wage of on-the-job employees (9.251%). However, the artificial surface index and cultivated land index ranked No. 5 are only 5.843% and 5.690%, which fully shows that the top four influencing factors are the leading factors of urban contraction. In addition, the total population ranked No. 18 at the end of the year, and IncMSE was 1.378%, which shows that the existing population has little impact on urban contraction. It can be found from the data that Qiqihar City, Anshan City, Tongliao City, Hulunbuir and Dandong Cities with large basic population are also frequently affected by urban contraction.

Based on the results of IncMSE, the random forest regression model is used to further analyze the impact of the top four influencing factors on shrinking cities. The standardized values of each influencing factor are input into the model and the average fitting results are fed back. The results are shown in **Figure 7**. As can be seen from **Figure 7**, only the unemployment rate has played a role in promoting the formation of shrinking cities, and in a period, the unemployment rate ≥ 0.44 has the strongest role in promoting the formation of shrinking cities. In addition to the unemployment rate, the other three influencing factors have played a role in inhibiting the formation of shrinking cities. In a period, when per capita GDP ≥ 0.6, science and education expenses ≥ 0.35 and the average wage of on-the-job employees ≥ 0.84, the inhibitory effect on the formation of shrinking cities reaches the maximum and remains unchanged.

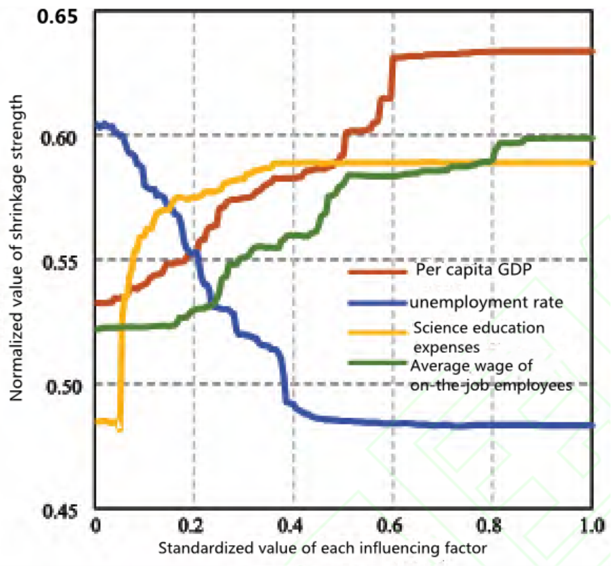

Figure 7. Impact of top 4 influencing factors on shrinkage intensity.

The per capita GDP reflects the urban development level and economic benefits under certain time and conditions, and is highly related to the living standard to a certain extent. In the random forest regression, the per capita GDP is the leading factor in the formation of shrinking cities, that is, Dalian, Shenyang, Changchun, Harbin, Daqing and other cities with high per capita GDP are less impacted by urban contraction. The employment rate can reflect a city's labor capacity and economic development. Good employment is an important factor to ensure social stability, while the unemployment rate is the reverse reaction of the employment rate. High unemployment rate will lead to violence and social unrest. When a city has high unemployment. The labor force tends to leave the city for other opportunities. In some resource-based cities in Northeast China, the imbalance of industrial structure leads to the problem of labor force allocation, which further leads to the emergence of a large number of unemployed people, which is the main reason why some resource-based cities in the north of Heilongjiang Province have been plagued by urban contraction for a long time. Science education is the potential source of urban and economic development. In terms of strength, the investment in science and education can develop science and technology and promote social progress. For cities, it can attract talents and drives the development of relevant industries. Because the four central cities of Dalian, Shenyang, Changchun and Harbin have a large investment in science and education and have more high-level colleges and Universities, they have a strong attraction for talents. Compared with other cities, they can slow down population loss by virtue of their excellent scientific and educational advantages in the wave of contraction. The average wage of on-the-job employees is the monetary wage per employee of each unit in a certain period of time, which can fully reflect the average social income level. Due to the low average income level in Northeast China, it is less attractive to talents. At the same time, many enterprises lack a reasonable and perfect reward mechanism, Lower wages have become an important reason for the outflow of talents in Northeast China.

5. Conclusion and discussion

This study analyzes the spatial distribution of shrinking cities in Northeast China from 2005 to 2009, 2010 to 2014, 2015 to 2019 and 2005 to 2019, and introduces multiple linear regression and random forest regression methods to analyze the influencing factors and effects of shrinking cities in Northeast China. The main conclusions are as follows: 1) In space, the shrinking cities in Northeast China are mainly distributed in the "land edge" areas represented by Changbai Mountain, Sanjiang Plain, Xiaoxing'an Mountain and Daxing'an Mountain. The main characteristics of these areas are poor traffic accessibility and large topographic relief, especially in Heilongjiang Province. The urban contraction phenomenon is weak in areas with better traffic conditions and gentle terrain, such as central and southern Liaoning and Harbin, Changsha. In terms of time, the contraction center in Northeast China shows a trend of moving north, while the expansion center moves south, and the spatial autocorrelation between urban contraction and expansion is further strengthened with time. 2) The correlation analysis between the shrinkage intensity of each city in each period and the selected influencing factors show that the correlation coefficients of each influencing factor are quite different. In the multiple linear regression method, the unemployment rate has the strongest impact on the contraction intensity, followed by per capita GDP, science and education expenses, cultivated land index, average wage of on-the-job employees and so on. The random forest regression method has a larger R^2 and smaller MSE on the sample set, which has good fitting effect and more practical significance. Through the ranking of the importance of influencing factors by IncMSE, the ranking results are close to the results obtained by multiple linear regression. The top four are per capita GDP, unemployment rate, science and education expenses and average wage of on-the-job employees. Although the weight ranking of each influencing factor obtained by the two regression methods is not completely consistent, it is not difficult to find that socio-economic factors are the main factors determining whether the city shrinks or not. 3) The further calculation of impact under the random forest regression method can provide a clear direction for the control of urban contraction. The results show that among the top four influencing factors, only the unemployment rate plays a role in promoting the contraction of cities, and the promotion effect reaches the strongest and remains unchanged when its standardized value is ≥ 0.44 in a period. Combined with the ranking results of the importance of influencing factors, it can be seen that controlling the unemployment rate can effectively alleviate the current situation of urban contraction to a certain extent. In addition to the unemployment rate, the other three influencing factors have inhibited the formation of shrinking cities. In a period, when the per capita GDP ≥ 0.6, science and education expenses ≥ 0.35 and the average wage of on-the-job employees ≥ 0.84, the inhibition effect on the formation of shrinking cities reaches the maximum and remains unchanged.

At present, there are still great differences in the quantitative determination of shrinking cities in the academic circles. Taking the population change in the northeast as a reference, this paper determines and identifies the shrinking cities in the northeast from the perspective of relative shrinkage, and further compares the influencing factors of shrinking cities by using multiple linear regression and random forest regression, it provides a novel perspective for the study of shrinking cities. However, compared with the traditional shrinkage determination method, the relative shrinkage determination method in this paper may only be suitable for the Northeast where most cities in the whole region have population loss. In addition, due to the availability of data, the selection of influencing factors in this paper is not comprehensive, and there are still many potential factors affecting urban population change that have not been selected. Based on the above problems, we can continue to improve the theory of quantitative identification and judgment of shrinking cities in combination with the situation of different regions in the future, and find the potential factors for the formation of shrinking cities Based on different cases.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Allagst K, Wiechimann T, Martinez-Fernandez C. Shrinking cities: International perspectives and policy implications. New York: Routledge; 2014.
2. Huang H. Smart shrinkage: Planning measures for urban decay and its practice in US. Journal of Urban and Regional Planning 2011; 4(3): 157–168.
3. Oswalt P. Shrinking cities: International research. Berlin: Hage Cantz; 2006.
4. Oswalt P. Shrinking cities: Interventions. Berlin: Hage Cantz; 2006.
5. Oswalt P, Rieniets T, Schirmel H, *et al.* Atlas of shrinking cities. Berlin: Hage Cantz; 2006.
6. Wolff M. Understanding the role of centralization processes for cities-evidence from a spatial perspective of urban Europe 1990-2010. Cities 2018; 75(5):

20–29.
7. Zingale N, Riemann D. Coping with shrinkage in Germany and the United States: A cross-cultural comparative approach toward sustainable cities. Urban Design International 2013; 18(1): 90–98.
8. Mallach A, Haase A, Hattori K. The shrinking city in comparative perspective: Contrasting dynamics and responses to urban shrinkage. Cities 2017; 69(9): 102–108.
9. Hartt D. How cities shrink: Complex pathways to population decline. Cities 2018; 75(5): 38–49.
10. Martinez-Fernandez C, Weyman T, Fol S, *et al.* Shrinking cities in Australia, Japan, Europe and the USA: From a global process to local policy responses. Progress in Planning 2016; 105(4): 1–48.
11. Hattori K, Kaido K, Matsuyuki M. The development of urban shrinkage discourse and policy response in Japan. Cities 2017; 69(9): 124–132.
12. Zhu X, Zhang Q, Sun P. Effects of urbanization on spatiotemporal distribution of precipitations in Beijing and its related causes. Acta Geographica Sinica 2018; 73(11): 38–56.
13. Wang L, Wu Ri, Li W. Spatial temporal patterns of population aging on China's urban agglomerations. Acta geographica Sinica 2017; 72(6): 1001–1016.
14. Long Y, Wu K, Wang J. Shrinking cities in China. Modern Urban Research 2015; (9): 14–19.
15. Mao Q, Long Y, Wu K. Temporal and spatial evolution of population density and spatial pattern of urbanization in China from 2000 to 2010. Urban Planning 2015; 39(2): 38–43.
16. Li X, Du Zi, Li X. The spatial distribution and mechanism of city shrinkage in the Pearl River Delta. Modern Urban Research 2015; (9): 36–43.
17. Liu Z, Qi W, Wang X, *et al.* A literature research on population shrinking. World Regional Studies 2019; 28(1): 13–23.
18. Yang Z, Yang D. Exploring shrinking areas in China availing of city development index. Human Geography 2019; 34(4): 63–72.
19. Zhang Mi, Xiao H. Spatial pattern characteristics and mechanism of urban contraction in Northeast China. Urban Problems 2020; (1): 33–42.
20. Yang M. Identification of Chinese urban shrinkage and its causes based on night light data. Hebei Academic Journal 2020; 40(2): 130–136.
21. Zhang Y, Wang Q, Fu Y, *et al.* Identification of shrinking cities at precision level in China and its driving forces. Journal of Geo-mathematics 2020; 45(2): 15–19.
22. Zheng L, Xu J, Wang X. Application of random forests algorithm in researches on wetlands. Wetland Science 2019; 17(1): 16–24.
23. Guo Y, Li L. Change in the negative externality of the shrinking cities in China. Scientia Geographica Sinica 2019; 39(1): 52–60.
24. Hoekveld J. Time-space relations and the differences between shrinking regions. Built Environment 2012; 38(2): 179–195.
25. Hollander JB, Németh J. The bounds of smart decline: A foundational theory for planning shrinking cities. Housing Policy Debate 2011; 21(3): 349–367.
26. Ma Z, Li C, Zhang J, *et al.* Urban shrinkage in developed countries and its implications for China. Human Geography 2016; 31(2): 13–17.
27. Yu C, Hu D, Cao S, *et al.* The spatial characteristics and changes of ISP-LST of Beijing in recent 30 years. Geographical Research 2019; 38(9): 2346–2356.
28. Anselin L, Bera A K, Florax R, *et al.* Simple diagnostic tests for spatial dependence. Regional Science and Urban Economics 1996; 26(1): 77–104.
29. Lin X, Yang J, Zhang X, *et al.* Measuring shrinking cities and influencing factors in urban China: Perspective of population and economy. Human Geography 2017; 32(1): 82–89.
30. Chen X, Zhang W, Zhang H. The relations of urban spatial expansion and economic growth in China: A case study of 261 precision level cities. Science Geographica Sinica 2016; 36(8): 1141–1147.
31. Wang C, Kan K, Zeng Y, *et al.* Population distribution pattern and influencing factors in Tibet based on random forest model. Acta Geographica Sinica 2019; 74 (4): 664–680.
32. Breiman L. Random forests. Machine Learning 2001; 45(1): 5–32.
33. Cutler A, Cutler D R, Stevens JR. Random forests. Machine Learning 2004; 45(1): 157–176.