

Quantifying the nonlinear factors of urban shrinkage: A case study in Heilongjiang Province, China

Xianjian Yi¹, Xinyue Gu^{2,*}, Qiru Zhong²

¹School of Jangho Architecture, Northeastern University, Shenyang, China

²Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China

* Corresponding Author Email: xinyue.gu@connet.polyu.hk

Abstract. With urbanization, urban shrinkage has become a more prevalent and severe global problem. Although many studies have explored it, most of them lack precision, and the complex nonlinear causal mechanisms of urban shrinkage are still poorly understood. Therefore, based on the background of the transformation of traditional industrial cities in China, this research selects Heilongjiang Province as a typical case study and quantify the spatial pattern of shrinking cities. Furthermore, from a nonlinear relational perspective, the complex mechanism of urban shrinkage is revealed by using LightGBM. The results show that the area experienced a large-scale comprehensive shrinkage and severe shrinkage in some areas, while also generating scattered population clusters. In addition, the economic and industrial structure are still the most influential factors for population change, and the industrial structure transformation is still the main obstacle for its development. The research results are conducive to elevating the awareness of urban shrinkage in industrial cities and providing some theoretical guidance for sustainable development.

Keywords: Urban shrinkage, Population loss, Non-linear relationship, LightGBM.

1. Introduction

Urban shrinkage caused by population migration has become an inevitable outcome of urbanization development in many countries since the 20th century. This phenomenon has become global and affects regions at a significant scale. The definition of urban shrinkage is diverse. It was first defined as population loss due to economic recession [1], and then developed into a multidimensional concept that goes beyond population loss. Population loss can lead to many socio-economic and urban development problems, such as economic recession, underutilization of existing urban infrastructure and low land use efficiency [2]. Therefore, it has become a serious challenge for urban development and has attracted great attention from urban planning scholars. Therefore, studying the factors that cause urban shrinkage can provide effective insights for policy making and urban management for sustainable urban development.

The most distinctive feature of shrinking cities is the decline in population, which has persistence and irreversibility. Therefore, most studies use population change as the criterion to identify urban shrinkage [3] and quantify the shrinkage level by the rate of population loss. Most of the current studies use panel data to analyze population, which cannot reflect the more accurate spatial distribution of population. However, with the development of urban data science, more and more scholars have used various high-precision grid data to measure the shrinkage pattern and quantify the shrinkage level accurately. For instance, NPP-VIIRS night light data was used to analyze urban population change and define urban shrinkage [4]. Likewise, using Land Scan and other data were used to explore urban dynamics from a spatial perspective of construction and economy [5].

The mechanism of urban shrinkage varies widely among different countries and regions, depending on various factors such as population, economy and society. For example, Europe and America regard globalization and deindustrialization as the main causes of shrinking cities, while Japan regards low fertility rate, aging and population structure change as the main factors [3]. However, the causes of urban shrinkage in different types of contexts may be more complex and deviate from the general causes in common contexts. For example, Ma et al. showed that Yokohama,

a representative traditional industrial city in Japan, shrank more due to the built environment attributes rather than population aging or low fertility rate [6]. Therefore, most studies will explore the related factors of shrinkage based on specific city type backgrounds. In the research on the driving factors, most of the current studies use linear models, such as OLS model [7] and GWR model [8], to study the impact of shrinking cities on social, economic and population factors. However, linear relationship studies may be difficult to reflect the complex mechanism of multi-factor influence on shrinking cities. In recent years, many urban studies have used machine learning methods to explore the causal mechanisms. Yang et al. explored how urban vitality and built environment attributes are related in a non-linear way [9]. Peng et al. used RF to analyze the nonlinear factors of Japanese shrinking cities. [10]. Their studies provided new research perspectives for analyzing the complex mechanism of urban shrinkage.

By reviewing the relevant literature, it is found that the identification of shrinking cities is mainly based on panel data, which lacks spatial accuracy and cannot accurately distinguish between registered and resident population. In terms of the influencing factors, most studies use OLS or geospatial analysis methods to analyze them, lacking the discussion of the nonlinear and complex causes, while machine learning is gradually applied to other urban issues. In this study, in order to better understand the shrinkage situation of industrial cities in the Chinese context, we use most of the regions in Heilongjiang Province, which has more traditional industrial cities, as case studies to investigate the impact of nonlinear factors of shrinking cities. The main contributions of this study are: (1) Based on high-precision grid unit population data, we investigate how population change varies spatially and how shrinking cities are distributed. (2) We reveal the nonlinear mechanism of shrinking cities. To conclude, the interrelationships and interactions between urban shrinkage and social, economic and built environment factors are revealed by this study. It also provides effective support for the planning and policy making of relevant government departments, and enables cities to develop sustainably and shrink wisely.

2. Methodology

2.1. Research Area and Data

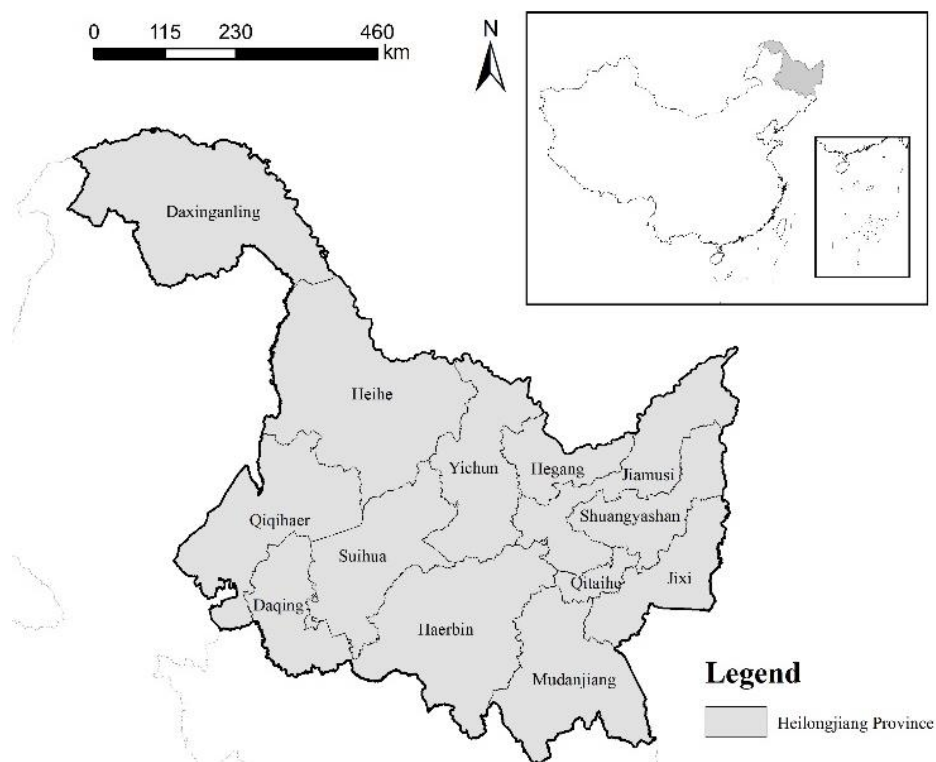


Fig. 1 Regions of Heilongjiang Province.

The study area, Heilongjiang Province, is located in the northernmost part of China. It has 12 prefecture-level cities and one regional administration. As one of the important provinces of China's old industrial base, Heilongjiang Province is a cluster of traditional heavy industrial cities. However, due to some external and internal factors, such as the obstruction of industrial transformation, excessive exploitation of energy and economic structure imbalance, it gradually lost its development advantages [11]. In recent years, their regional competitiveness has declined significantly, leading to large-scale urban shrinkage, which has caused many adverse phenomena, such as low and blind expansion of urban land and space utilization, aggravating the urban shrinkage process. But at the same time, the Chinese government attaches great importance to the development of the Northeast, by proposing the "Revitalizing the Northeast Strategy" to achieve high-quality urban development transformation [12]. In hence, conducting targeted research on the urban shrinkage problem in Heilongjiang Province is both urgent and essential, which can support the exploration of the transformation path of Northeast industrial cities. Since Jiagedaqi District belongs to Daxinganling Prefecture of Heilongjiang Province, but actually lies in Inner Mongolia Autonomous Region, considering the integrity of the research area, we temporarily select other sites except this area as research objects.

This study selects 2015 and 2020 as two time points for analysis, to investigate the characteristics and influential factors of shrinking cities. Since traditional population census and other statistical data are difficult to analyze the shrinkage pattern more accurately from the spatial distribution and local perspective, we obtain high-precision grid data on population, economy, society and other aspects related to urban shrinkage from diversified open data platforms to explore the relation, which are summarized in Table 1. To ensure the data accuracy, all data are cleaned and filtered, and then normalized by Python to eliminate the dimensional difference between data and improve the consistency and interpretability of data.

Table 1. Details of multi-source urban data.

Type	Name	Description	Resolution	Data Source
Population	PD	The population density (PD) is reflected by the number of people in each grid unit. The change rate can be generated by further calculation.	1km	
Society	BR	The birth rate (BR) is indicated by the number of 0-1 year old population in each grid unit, reflecting the impact of age structure.	100m	WorldPop (https://hub.worldpop.org/)
	AR	The aging rate (AR) is indicated by the number of elderly population over 65 years old in each grid unit, reflecting the impact of age structure.	100m	
	PA	The degree of population agglomeration (PA) in the region is calculated by the local Moran index	1km	
	HF	Human Footprint (HF) reflects the range of human activities in a specific year, to a certain extent, reflecting the human vitality.	1km	[13]
Economy	gdp	The GDP output in each grid unit reflects the accurate spatial distribution of economic output.	1km	[14]
	PPI	The proportion of the primary industry (PPI) in the regional GDP \times grid GDP reflects its industrial structure effect.	1km	China Statistical Yearbook
	PSI	The proportion of the secondary industry (PSI) in the regional GDP \times grid GDP indicates its industrial structure impact.	1km	(http://www.stats.gov.cn/)
	PTI	The proportion of the tertiary industry (PTI) in the regional GDP \times grid GDP indicates its industrial structure impact.	1km	[14]
Built Environment	cityland	It reflects the effect of the urban built-up land area.	1km	[15]
	GAIA	Global artificial impervious areas (GAIA) reflect the degree of land development and utilization.	30m	[16]
	POI	Point of interests (POI) indicate the impact of urban basic service facilities.	/	https://www.openstreetmap.org
	RD	Road density (RD) shows the effect of accessibility.	/	p.org

2.2. Measurement and Spatial Pattern Analysis of Population Shrinkage

To determine the distribution of shrinkage, this study performs a quantitative analysis of the spatial distribution of population change rate in unit grids. Then, we use the Local Moran's I test to further analyze the pattern and degree of shrinkage, as shown in Eq. (1). The local Moran's I index is a way to measure how spatial units relate to their adjacent units. It identifies spatial clusters and outliers, evaluates the differences and similarities between adjacent grids, and finally generates local indicators of spatial association (LISA) [17]. Usually, using local Moran's I will generate four types of clusters/outliers (HH, HL, LH, LL). They could show the different relationships between the population change rate of the grid and its surrounding grids with the average value. In hence, to verify the aggregation and dispersion effects of population change within the region, we can analyze the spatial pattern of population change in each grid using Local Moran's I statistics.

$$I_i = \frac{Z_i}{S^2} \sum_{j \neq i}^n w_{ij} Z_j \quad (1)$$

2.3. Nonlinear Relationship Analysis Based on LightGBM

In this study, We use python to build a LightGBM model, which is an efficient machine learning framework based on gradient boosting trees with the features of low memory usage, fast calculation speed and high precision [18]. We use it to investigate the nonlinear relation between population change and its related indicators. We also compare it with linear models to test its validity. Considering the impact of multicollinearity, we use correlation matrix and variance inflation factor (VIF) to detect multicollinearity, and ensure that the VIF values of all variables are less than 10 to verify the validity of the variables, and avoid the skewed or misleading results that may be caused by multicollinearity. Then, we use the LightGBM model and the simultaneously built regression model to compare and evaluate the model interpretability by R^2 and RMSE. The closer the value of R^2 is to 1, the stronger the explanatory power of the model, and the better the fitting degree. Correspondingly, the model is more accurate when the RMSE value is smaller, which means the deviation between the predicted and observed values is smaller. In addition, we build two data sets for the LightGBM model: training set (80%) and test set (20%). Then, we use LightGBM model to generate feature importance scores, which explain how important each variable is in the model.

3. Results

3.1. Spatial Pattern of Population Shrinkage in Heilongjiang Province

As shown in Fig. 2(a), this study calculated the population change rate in 475,283 grids, with a valid range from 0.40 to 9.30. Among them, 176,349 grids experienced shrinkage, accounting for 37.1%. Thirteen regions underwent large-scale shrinkage, and the most severe cities were Yichun and Jixi (in the eastern and northern parts of Heilongjiang Province), where a large area lost more than 30% of the population. Then We further examined the spatial pattern of population change using Local Moran's I. As Fig. 2(b) indicates, the northern and eastern parts of Heilongjiang Province had more LL clusters in the Daxinganling, Heihe and Jixi regions, while other regions were mainly dominated by HH clusters. The former may be due to the low population change rate caused by being located in the natural forest area, while the latter was consistent with Figure 2(a) and caused by population shrinkage. In addition, we also observed some LH and HL outliers. There were more HL outliers in the northern (Jiamusi region), southwestern (Daqing region) and eastern (Jixi region) parts of Heilongjiang Province, and the more severe the shrinkage, the more HL outliers there were. The LH outliers were mainly distributed around the LL clusters, and both reflected a certain population agglomeration effect caused by shrinkage. Among them, the LH outliers in Jiamusi and Daqing regions were mainly distributed around HH clusters, while those in Jixi region were mainly within LL clusters. In general, the population change in Heilongjiang Province showed the synchronicity of

population shrinkage and agglomeration, presenting a pattern of comprehensive shrinkage, regional severe shrinkage, and scattered population agglomeration.

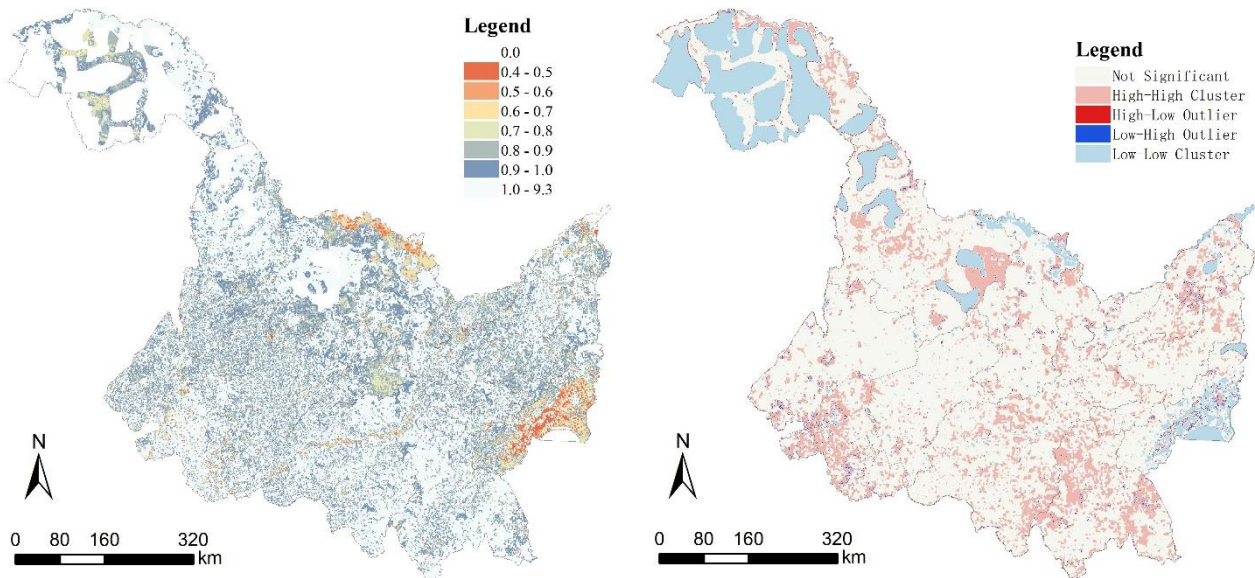


Fig. 2 Spatial patterns of (a) population loss and (b) LISA in Heilongjiang Province.

3.2. Nonlinear Relationship between Urban Shrinkage and Corresponding Variables.

Then, we used LightGBM to discover the relationship between urban population change and corresponding indicators. As shown in Table 2. and Fig. 3, this study used the correlation matrix and VIF to test whether there was mutual influence between variables. According to the results, we observed that the VIF values of all variables were less than 5, so we assumed that the whole data set did not have a serious multicollinearity problem.

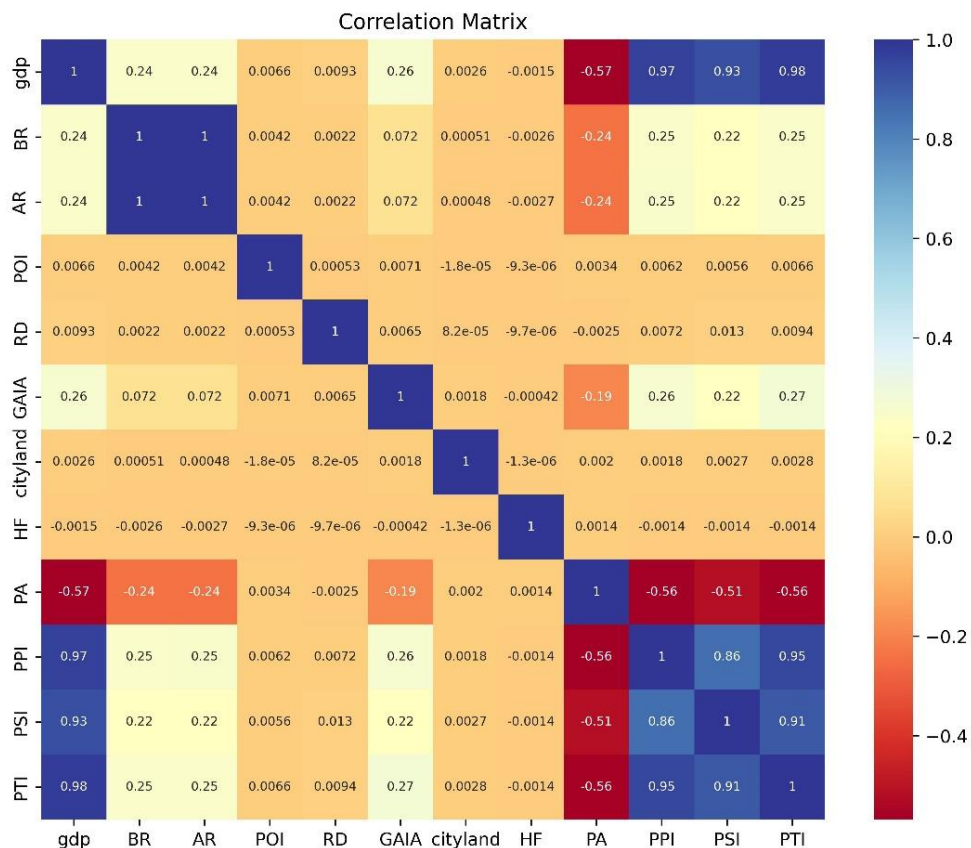


Fig. 3 Correlation matrix of indicators affecting the population change in Heilongjiang Province.

Table 2. VIF value of indicators affecting the population change in Heilongjiang Province.

Number	Feature	VIF
1	AR	0.999934
2	BR	0.999934
3	gdp	0.992786
4	PTI	0.981024
5	PPI	0.974799
6	PSI	0.951890
7	PA	0.406027
8	GAIA	0.160131
9	HF	0.000470
10	RD	0.000387
11	POI	0.000195
12	cityland	0.000035

In addition, as shown in Table 3, LightGBM can be effectively used to study the relationship between corresponding variables and population change. The model evaluation results showed that both types of models could effectively explain urban shrinkage. Among them, the R^2 of LightGBM model on the training set and test set were 0.9980 and 0.9973, respectively, indicating that the model had a good fit to the data. Moreover, a low prediction error of the model was indicated by the RMSE of LightGBM model, which were 0.0049 and 0.0057 on the training and test set, respectively. The R^2 of linear regression model on the training set was 0.9848, and the RMSE was 0.0137. If the model has high R^2 and low RMSE on both the training set and test set, it indicates that the model is neither underfitting nor overfitting and has a good model interpretability. Compared to linear regression model, LightGBM regression model had a better fit and a lower bias, which indicated that the nonlinear model could better account for the factors of shrinkage than the linear model. This also reflected the complexity of the path causes of urban shrinkage.

Table 3. Statistical results of LightGBM and linear regression models.

models	R^2	RMSE
Linear Regression	0.9848	0.0137
LightGBM Regression Train	0.9980	0.0049
LightGBM Regression Test	0.9973	0.0057

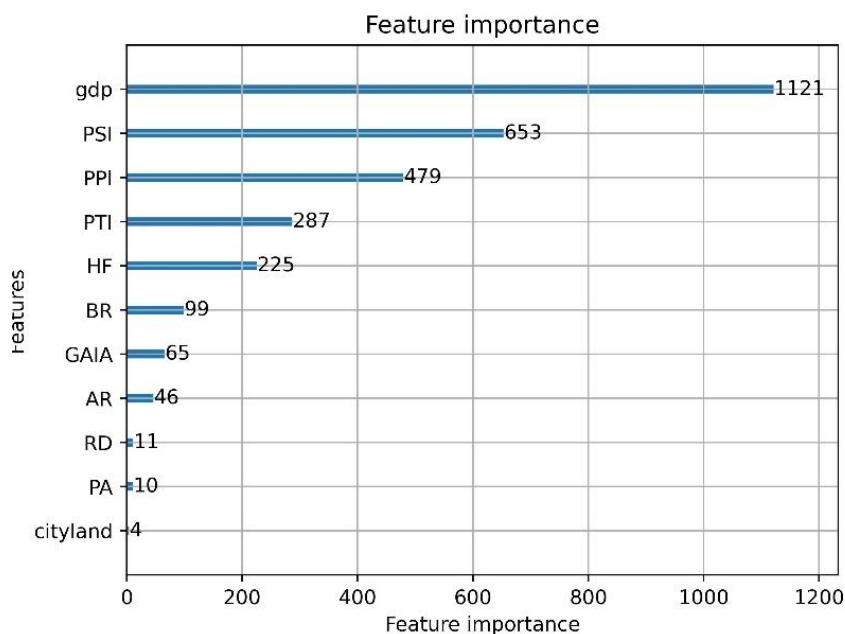


Fig. 4 Feature importance of indicators affecting the population change in Heilongjiang Province.

Afterwards, we further discussed the relationship by using the feature importance plot (Fig. 4). Generally speaking, gdp had the highest feature importance, far exceeding other variables. In addition, HF and the three indicators of industrial structure all exceeded 200, indicating that these variables had a certain relationship with population change. At the same time, BR, GAIA, AR had slightly weaker scores, all below 100, which showed that population structure and existing built environment had a weak impact on population change. Meanwhile, other variables, RD, PA, cityland had a weak relation with urban shrinkage, with scores all below 15, indicating that urban built-up area, accessibility and population agglomeration had little impact on population change. The POI data was too small to be identified by the model due to its low importance.

To sum up, we concluded that economic-related variables, namely economic output and industrial structure, had the greatest influence on shrinking cities, which showed the economic structural imbalance, reflecting the population outflow triggered by the challenges faced in the process of industrial transformation. Among them, the proportion of second industry output decreased, while the corresponding proportions of other two industries increased, reflecting the dilemma faced by industrial cities in industrial transformation and innovative development, resulting in a large number of people moving to low-value-added industries, which hindered economic development and correspondingly caused population loss. In addition, the low birth rate and high aging rate, as well as human activities, also showed relatively low development expectations and urban vitality. The existing built environment had little impact on population change, which may be due to the relatively early development and economic stagnation in the northeast region, resulting in slower construction speed, inertia and stability of spatial order and built environment, which led to the paradox of urban environment expansion and urban shrinkage.

4. Discussion and Conclusion

As urban shrinkage is an inevitable path of urban transformation in the later stages of urban development, exploring the impact mechanisms of urban shrinkage is very important for the correct formulation of relevant urban development management policies. This study selected Heilongjiang Province, an old industrial base, as the research object to analyze the population shrinkage in Heilongjiang from 2015 to 2020 and discussed the relationship between relevant factors and urban shrinkage from the perspectives of spatial impact distribution and related influencing factors of population change, which is conducive to providing new analysis and decision-making perspectives for future planning and economic development.

We measured urban shrinkage by using population change as the indicator. We first analyzed the spatial pattern and agglomeration and dispersion effects of population change. In addition, we used LightGBM model to investigate how influencing factors of urban shrinkage relate to the occurrence and causal mechanism of regional shrinkage of industrial cities in Northeast China. The results are as follows: (1) Heilongjiang experienced a large area of comprehensive shrinkage and partial area of severe shrinkage, as well as scattered population agglomeration. Among them, Jixi and Yichun had the most serious shrinkage problems. (2) The main cause of urban shrinkage in Heilongjiang was still economic development. The economic output and industrial structure transformation of Heilongjiang were still the main factors affecting shrinkage, while the corresponding population structure, human activity intensity and scope had a certain impact, while the corresponding built facilities, agglomeration degree and other factors had only a weak impact on shrinkage.

However, due to objective conditions, this study also has some limitations. First, it only compared the feature importance of different factors, without reflecting their positive or negative effects on population shrinkage. In the future, nonlinear analysis can be performed on individual indicators and urban shrinkage separately to further refine their impact mechanisms. In addition, we only discussed the transformation of industrial cities from the perspective of industrial structure, without other more targeted data and analysis perspectives, owing to the limitations of data sources and accuracy, which can also be a possible direction for future refined research.

This study enriched the explanation of the factors leading to shrinking cities from the perspective of nonlinear relation, and also provided a comprehensive overview and insights for subsequent rational development planning and policy formulation, which is helpful to achieve sustainable development and wise shrinkage of shrinking cities.

References

- [1] Häussermann, H., & Siebel, W. Die Schrumpfende Stadt und die Stadtsoziologie [The shrinking city and urban sociology]. In *Soziologische Stadtforschung*. 1988.
- [2] Bartholomae F, Woon Nam C, Schoenberg A. Urban shrinkage and resurgence in Germany. *Urban Studies* 2017; 54(12): 2701–2718.
- [3] Döringer S, Uchiyama Y, Penker M, Kohsaka R. A meta-analysis of shrinking cities in Europe and Japan. Towards an integrative research agenda. *European Planning Studies* 2020; 28(9): 1693–1712.
- [4] Jiang Z, Zhai W, Meng X, Long Y. Identifying Shrinking Cities with NPP-VIIRS Nightlight Data in China. *Journal of Urban Planning and Development* 2020; 146(4): 04020034.
- [5] Lang W, Deng J, Li X. Identification of “Growth” and “Shrinkage” Pattern and Planning Strategies for Shrinking Cities Based on a Spatial Perspective of the Pearl River Delta Region. *Journal of Urban Planning and Development* 2020; 146(4): 05020020.
- [6] Ma, S., Kumakoshi, Y., Koizumi, H., & Yoshimura, Y. Determining the association of the built environment and socioeconomic attributes with urban shrinking in Yokohama City. *Cities* 2022; 120: 103474.
- [7] Zhang Y, Fu Y, Kong X, Zhang F. Prefecture-level city shrinkage on the regional dimension in China: Spatiotemporal change and internal relations. *Sustainable Cities and Society* 2019; 47: 101490.
- [8] Shan J, Liu Y, Kong X, Liu Y, Wang Y. Identifying City Shrinkage in Population and City Activity in the Middle Reaches of the Yangtze River, China. *Journal of Urban Planning and Development* 2020; 146(3): 04020027.
- [9] Yang J, Cao J, Zhou Y. Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. *Transportation Research Part A: Policy and Practice* 2021; 144: 74–88.
- [10] Peng W, Wu Z, Duan J, et al. Identifying and quantizing the non-linear correlates of city shrinkage in Japan. *Cities* 2023; 137: 104292.
- [11] Sun H, Zhou G, Liu Y, Fu H. Relationship between the degree and spatial pattern of urban shrinkage based on nighttime light data: a case study of the old industrial area in Northeast China. *International Journal of Urban Sciences* 2023; 0(0): 1–25.
- [12] Jiang Y, Chen Z, Sun P. Urban Shrinkage and Urban Vitality Correlation Research in the Three Northeastern Provinces of China. *International Journal of Environmental Research and Public Health* 2022; 19(17): 10650.
- [13] Mu H, Li X, Wen Y, et al. A global record of annual terrestrial Human Footprint dataset from 2000 to 2018. *Scientific Data* 2022; 9(1): 176.
- [14] Zhao N, Liu Y, Cao G, Samson EL, Zhang J. Forecasting China’s GDP at the pixel level using nighttime lights time series and population images. *GIScience & Remote Sensing* 2017; 54(3): 407–425.
- [15] Li X, Gong P, Zhou Y, et al. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environmental Research Letters* 2020; 15(9): 094044.
- [16] Gong P, Li X, Wang J, et al. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment* 2020; 236: 111510.
- [17] Forey E, Lortie CJ, Michalet R. Spatial patterns of association at local and regional scales in coastal sand dune communities. *Journal of Vegetation Science* 2009; 20(5): 916–925.
- [18] Ke G, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 2017.