ARTICLE

Check for updates

Spatial and temporal changes in population distribution and population projection at county level in China

Mei Sang[1], Jing Jiang[1✉], Xin Huang[1], Feifei Zhu[1] & Qian Wang[1]

Counties in China play a pivotal role in economic and social development, acting as essential leadership hubs for large and medium-sized cities, contributing to rural revitalization, and facilitating urban-rural integration. Using ArcGIS spatial analysis, this study examines the population distribution spatial and dispersion patterns in Chinese counties based on 40 years of data from the first to the seventh national population census. Results reveal noticeable growth trends and regional disparities in county populations, with an increase in large-population counties and a decrease in small-population ones. Recent population growth concentrates in urban agglomerations, metropolitan areas, and southeastern coastal regions, while reductions occur in the northeastern and Inner Mongolia border areas. Furthermore, the study identifies "high-high" agglomerations around provincial capitals and "low-low" agglomerations in economically underdeveloped western and northeastern border regions. China's population distribution spatial agglomeration has been increasing, with acceleration toward specific areas. The first through seventh census show rapid growth at low density, followed by growth at medium density, stable growth, and eventually negative growth. This suggests a likely slowdown and potential reversal in China's future population growth. Additionally, an Auto Regression Integrated Moving Average (ARIMA) model is employed to forecast China's total population, projecting a decline to 1343.68 million by 2035. The emergence of "population loss counties" in contemporary China underscores the need for a rational understanding of their development status and trends to optimize population development strategies and promote economic and social progress.

[1] College of Management, Gansu Agricultural University, 730000 Lanzhou, China. ✉email: jiangjing@gsau.edu.cn

Introduction

Changes in the world's human dynamics are crucial for understanding the current state of the world, which is faced with increasing challenges related to globalization. The process of globalization has made the world more dynamic than ever before. From an economic perspective, developing countries such as China and India have experienced unprecedented growth over the past 30 years. In addition to the realignment of the global economy, the process of economic globalization has also led to the redistribution of the global population, particularly in urban areas. Asian developing countries, especially China and India, are currently undergoing a significant transformation in the form of large-scale urbanization (Balsa-Barreiro et al., 2019). The demographic changes in China align with the overall global demographic trends. However, due to its immense population size, rapid demographic shifts, and aging population, China is expected to bear greater impacts from demographic changes. Looking back at the historical process of global population transitioning from high-speed growth to slowed growth, it is evident that successful socioeconomic development and policy responses have effectively alleviated unprecedented population growth pressures. Anticipating the future trajectory of demographic changes, especially gaining a more accurate understanding of China's position in global transformations, will aid us in promptly adapting to new population dynamics and actively addressing the challenges they bring.

The trend of urban economic development has been notably pronounced in recent years, resulting in a more concentrated population. Currently, approximately half of the global population resides in urban areas. It is projected that by 2050, this ratio will increase to two-thirds. Over time, the process of urbanization in developing countries is progressing at a faster pace compared to the historical experiences of developed nations. The urban population ratios in these countries are expected to surpass levels anticipated based solely on their economic development and wealth concentration. Consequently, the population movements witnessed by developed countries over the past two centuries are now unfolding in developing nations, albeit at an accelerated pace. Globally, the traces delineated by key indicators related to societal demographic trends underscore this prevailing pattern. In the works of Balsa-Barreiro et al. (2019), the global wealth shift toward the East is observable, while traces associated with population growth and urbanization rates are shifting toward the global South. Scholars have pointed out that the global population distribution occurs at two extremes: in large urban agglomerations and in rural deserts. However, Balsa-Barreiro et al. (2021) argue that this commentary is overly generalized and imprecise, emphasizing the necessity of analyzing these dynamics at other spatial scales. In order to conduct a more comprehensive analysis of the data, enabling us to observe significant changes in the spatial pattern of population distribution, Balsa-Barreiro et al. employ a local-scale approach to analyze rural and sparsely populated areas, revealing a sharp phenomenon of internal population redistribution. It demonstrates how urban dynamics are not scale-dependent and can even manifest in sparsely populated regions. Currently, our understanding of regions beyond urban areas is still lacking, necessitating the adoption of a multi-scale approach to examine the spatial behaviors of these dynamics at different spatial scales. Since the 21st century, with the acceleration of China's industrialization level and urbanization process, sustainable development has become the main theme of China's future development, ensuring the coordinated development of the regional populations, economies, and resource environments while achieving sustainable development has become an important issue at present (Yuan, 1993; Ye et al., 2001; Ellis and Ramankutty, 2008; Ramankutty and Foley, 1999).

As one of the most densely populated countries in the world, China's population problem has always been the focus of attention of the government and society. The population problem has always been intertwined with the rise and fall of the nation, the survival of the country, and the ups and downs of the society. It has always been a multifaceted and long-term problem involving various fields such as economy, society, politics, etc., which is related to China's comprehensive strength and international competitiveness. The population problem is not only manifested in the change in population number, but also its spatial distribution status is one of the focuses of many scientific studies (Guan et al., 2018; Clarke, 1978; Yue et al., 2005).

In 1935, Mr. Hu Huanyong, a Chinese population geographer, put forward the "Heihe-Tengchong" line for the first time on the basis of the national population density in his article "On the Distribution of China's Population", which vividly depicted the pattern of population distribution, which is "dense in the southeast and sparse in the northwest" (Hu, 1983). It vividly depicted the population distribution pattern of "dense in the southeast and sparse in the northwest". Subsequently, academics carried out a lot of related research work, ushering in the peak period of population geography research (Qi et al., 2016). From the 1950s to the present, China has carried out seven national population censuses, based on these censuses data, many population geography researchers have also carried out systematic research on the spatial distribution of the population by using advanced technical methods, the research scale mainly includes the global population distribution (Balk et al., 2006; Dobson et al., 2000; Ramankutty et al., 2002), national population distribution (Goldewijk, 2005; Liang et al., 2014), regional population distribution (Liu et al., 2014; Congdon et al., 1990; Briggs et al., 2007), and the research mainly focuses on population density, population concentration degree, population distribution imbalance, population potential and Lorentz curve and other indicators to conduct in-depth research on population distribution (Liu et al., 2008; Zhuo et al., 2009; Chi, 2012; Franklin, 2014; Liao et al., 2010). Optimizing the population development strategy should not only be based on the new pattern of China's population development but also focus on the country's medium and long-term development planning. In order to maintain the basic stability of social life and keep the population at an idealized state, it is necessary to make a reasonable prediction of the total population of the country and the region, so as to formulate the future population policy and promote the orderly development of the society. At present, population forecasting models mainly include two major categories, the first category is to establish forecasting models from the parameters of fertility and mortality population development (Booth et al., 2006), including the probabilistic population forecasting model (Alkema et al., 2011; Raftery et al., 2013; Gerland et al., 2014), Leslie matrix model (Anggreini, 2017; Li et al., 2018). The second category is to establish forecasting models for the time series of the total population, including regression model (Aalen, 1989; Poole and O'Farrell, 1971), gray forecasting GM model (Wang et al., 2018), artificial neural network model, combined forecasting model, etc., of which ARIMA (Auto Regression Integrated Moving Average) model, as a common the ARIMA model, as a common stochastic time-series model, can realize the optimal prediction in the sense of minimum variance, which is more effective and widely used in short-term population prediction (Piccolo, 1990; Shumway et al., 2006; Gilbert, 2005). Since the population forecasting method of the ARIMA model has strong applicability and high accuracy, this paper forecasts China's total population from 2023 to 2035 by building an ARIMA model.

Due to the limitation of data availability, existing studies on China's population distribution have mainly focused on recent years or on more developed regions, and there are almost no

large-scale studies of population distribution from the first to the seventh census at the county level across the country. Therefore, a nationwide study of population distribution and spatial and temporal changes at the county scale is important for scientifically portraying China's population distribution situation and pattern, deeply understanding population distribution patterns and future population evolution trends, and providing the scientific basis and decision support for strengthening population management, as well as coordinating regional sustainable development and ecological and environmental conservation.

Research data and method

Research data. The data used in this article mainly comes from the national population census data, the main data bulletins of provincial-level administrative divisions (autonomous regions, municipalities directly under the central government) in the national population census, the China Statistical Yearbook, and the National Bureau of Statistics Website. County-level administrative division maps are sourced from the data center of the Resources and Environmental Sciences of the Chinese Academy of Sciences. The database of the resident population in county units from the first census to the seventh census was constructed with 2829 spatial units for the first, second, and third censuses, 2830 spatial units for the fourth, fifth, and sixth censuses, and 2849 spatial units for the seventh census, and the analysis in this paper does not include Hong Kong, Macao, and Taiwan for the time being.

Research methodology

Population density. Population density refers to the number of permanent residents on a certain unit of land area in a specific geographic area in a certain period of time, usually calculated by dividing the number of people by the area of the region, which is mainly used to measure the degree of intensity of the distribution of the population in a region, the formula is:

$$AD_i = \frac{P_i}{A_i} \quad (1)$$

Where AD_i is the population density in region i; P_i is the number of people in region i, A_i is the land area in region i.

Spatial autocorrelation. Spatial autocorrelation is often used to study changes in spatiotemporal patterns, and this paper quantitatively examines the agglomeration of the spatial distribution of population through GeoDa software and combined with ArcGIS spatial analysis and mapping functions. The method aims to explain the degree of agglomeration of specific attribute values in different spatial locations within a region and spatial autocorrelation includes global spatial autocorrelation and local spatial autocorrelation (Anselin, 1988; Miller, 2004; Tobler, 1970).

Global spatial autocorrelation is used to characterize the spatial correlation of geographic elements in the whole study area, and the global Moran's index (global Moran's I) is usually used to reflect the overall spatial correlation and variability, the value of Moran's I range from [-1,1], the Moran's I index is greater than 0 indicates spatial positive correlation, and the bigger the value the smaller the spatial variability is, the more obvious the agglomeration characteristics, the Moran's I index is less than 0 indicates a negative correlation, and the smaller the value the bigger the spatial variability is, and the more obvious the dispersed characteristics are, and the Moran's I index equals to 0 indicates that it is not correlated, and spatial

randomness is present.

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

Where I is the global Moran index; n is the number of samples in the study area; X_i, X_j is the spatial location i, j attribute value; W_{ij} is the spatial weight matrix.

Local spatial autocorrelation refers to the correlation between a specific location and its neighboring locations in geospatial space (Anselin, 1995). This method can reflect the agglomeration status of the population in the geospatial region and the degree of local correlation of each spatial unit, which is commonly expressed by local Moran's Index (local Moran's I) and is usually presented in five spatial relationships: high and high agglomeration, high and low disagglomeration, low and high disagglomeration, low and low agglomeration, and non-significant (no significant spatial relationship exists).

$$I_i = z_i \sum_{j \neq i}^{n} W_{ij} Z_j \quad (3)$$

Where I_i is the local Moran index; W_{ij} is the spatial autocorrelation weight matrix; Z_i, Z_j is the standard deviation normalized attribute value.

Based on the population-land Gini coefficient. The Gini coefficient is based on the principle of the Lorenz curve proposed as a measure of the degree of inequality of income or wealth within a certain range of residents in general, the Gini coefficient values the range of 0 to 1, the closer the value of 1 indicates that the higher the degree of inequality, the closer the value of 0 indicates that the distribution of the more tends to be equal. In this paper, the Gini coefficient is used to quantitatively analyze to reveal the degree of inequality in the distribution of population and land resources in Chinese counties, firstly, the land area is divided into n groups, and all the samples are arranged according to the population density m_i in descending order, then the formula of population-land Gini coefficient (G_R) is as follows:

$$G_R = 1 - \sum P_i(2Q_i - W_i) \quad (4)$$

Where G_R is the population-land Gini coefficient; P_i is the land ratio; Q_i is the cumulative population ratio; W_i is the population ratio.

ARIMA modeling. ARMA model is a common time-series analysis model, which combines the characteristics of autoregressive (AR) and moving average (MA) models, used to describe the stochastic variation in the time-series data (Newbold, 1983). ARMA model is also known as the Box-Jenkins method, initially proposed by the American statisticians Burks and Jenkins, the object of the model is a smooth stochastic process, which is now an important method to study the time-series model due to its simple and easy-to-use, easy to calculate, has become an important method to study the time-series model. Compared with it, the object of the ARIMA model is a non-smooth process, and its basic idea is to transform the non-smooth process into a smooth process by difference, which is essentially a combination of difference and the ARMA model. The ARIMA (p, d, q) model consists of three parameters, where p represents the autoregressive (AR) order of the model, indicating the relationship between the current observation and the past p observations; q represents the moving average (MA) order of the model, indicating the relationship between the current observation and the past q error terms; and d represents the differencing order of the sequence. This model can be categorized into three types: the AR (p) model, the MA (q) model, and the ARIMA (p, d, q) model.

Modeling process of ARIMA model:

(1) Load the data, and the obtained sequence of observations will be used to determine the smoothness of the data series through the graphical method and the unit root test;

(2) If the time series is non-stationary, the non-stationary series can be transformed into a stationary series by difference operation;

(3) Establish ARIMA (p, d, q) model for the differenced series;

(4) Test the established model;

(5) Use the tested model to forecast the series.

Results

A study of the spatial and temporal patterns of population change in the county

Distribution pattern of population size in the county. The population size of China's counties shows a clear trend of growth and regional differences. The average population size of Chinese counties grew from 228,000 to 498,000 during the period from the first to the seventh national census, with the largest average population growth from the second to the third national census, with a growth rate of 41.33%; the growth rate from the fourth national census showed a decreasing trend, and the average population growth from the sixth to the seventh national census was the smallest, with a growth rate of 5.64%. The Hu Huanyong line defines the main characteristics of China's population distribution in the east and west, and from the analysis of the population of the counties from the first to the seventh population census, the proportion of the southeast half of the wall to the total population on both sides of the Hu Huanyong line decreases from 94.94% during the period of the first popularization to 93.7% during the period of the seventh popularization, while the proportion of the northwest half of the wall to the total population increases from 5.06% during the period of the first popularization to 6.25% during the period of the seventh popularization, indicating that the population of the southeast half of the wall continues to decrease slightly, and the population of the northwest half of the wall continued to decrease slightly, while the population of the northwest half-wall gradually increased slightly. On the whole, the population ratio between the southeastern and northwestern parts of China remains at 94:6, while the distribution pattern of relatively dense and sparse areas remains stable. This once again confirms the regional differences in population distribution patterns under the influence of the natural geographical environment.

Based on the number of populations in counties from the first to the seventh census, using the Hu Huanyong line as a benchmark, it was concluded that counties and districts with larger populations (>500,000) were mainly located in the southeastern half of the country, and counties and districts with smaller populations (<100,000) were mainly located in the northwestern half of the country. The number of counties and districts with a population of less than 100,000 showed an overall decreasing trend during the period from the first to the seventh census, in which the number of counties and districts with a population of less than 100,000 accounted for about 28.21% of the total during the first census period, and the number of counties and districts with a population of less than 100,000 accounted for only 10.42% of the total during the seventh census period and were mainly distributed in the northwestern region of the Hu Huanyong line, including Tibet, Xinjiang, Qinghai, Gansu, Inner Mongolia, and Sichuan; the number of counties and districts with populations of more than 500,000 people. The number of counties with a population of more than 500,000 showed a general trend of growth, in which the number of counties with a population of more than 500,000 accounted for about 8.52% of the total in the

first national census period, and the number of counties with a population of more than 500,000 accounted for 36.85% of the total in the seventh national census period, which were mainly distributed in the southeast of the Hu Huanyong Line, including the Middle and Lower reaches of the Yangtze River plains, the North China plains, the Pearl River Delta, the Yangtze River delta, and the Sichuan Basin, and so on. All in all, the number of counties and districts with large-population sizes is increasing and the number of counties and districts with small-population sizes is decreasing during the period 1953–2020 (Figs. 1 and 2).

Divergence of population increase and decrease in counties. Between the first and the seventh censuses, at least more than 40% of the counties and districts with data from each of the two adjacent censuses showed an increasing trend in the rate of population change. The highest number of counties and districts, 96.43% of the total, experienced population growth between the second and third censuses; the number of counties and districts with population growth from the third census onwards showed a gradual downward trend; and the number of counties and districts with population growth between the sixth and seventh censuses was the lowest, accounting for 46.64% of the total. Among them, the number of counties and districts with population growth from the sixth to the seventh population census was 1301, accounting for 45.97% of the country, and the number of counties and districts with an increase of more than 20% was 483, accounting for 17.07% of the country's counties and districts, which were mainly located in the southeastern coastal areas and the economically developed areas of Beijing and Tianjin, and the populations of some areas in Xinjiang and Tibet also showed large growth; the number of counties and districts with a decrease in population was 1529, accounting for 54.03% of the country's population. There are 1529 counties and districts with population decreases, accounting for 54.03% of the country, of which 288 counties and districts with decreases of more than 20% belong to serious population loss, mainly distributed in Heilongjiang, Jilin, Liaoning, central and northeastern Inner Mongolia and the eastern part of Sichuan Province. About 17.81% of the county population decreased by 10% to 20%, which is an obvious loss of population, about 26.04% of the county population decreased by 10% or less, a mild loss. This reflects the fact that in recent years China's larger population increase in the region is mainly concentrated in the southeastern coastal areas and the more economically developed regions, they as China's science and technology development of the pilot area, the siphoning effect on the population is very obvious; the larger decrease in the region is mainly distributed in the northeast, Inner Mongolia's borders, along with the decline of the old industrial zones of the Northeast, the industrial structure into the offensive period, the population shows a large number of trends in the loss of population, so we have to pay further attention to the problem of population loss in the border areas, and strengthen the construction of border areas to ensure the security and stability of China's border areas (Fig. 3).

Spatial autocorrelation analysis of county population change. Exploratory spatial analysis of county population change based on GeoDa software was conducted so as to understand the spatial correlation of county population increase and decrease and its degree of correlation. From the perspective of global spatial autocorrelation, the negative correlation of the global Moran index in the spatial distribution of the rate of population change during the period from the first census to the seventh census to the period from the third census to the fourth census indicates that the spatial contraction of the population tends to be discrete,
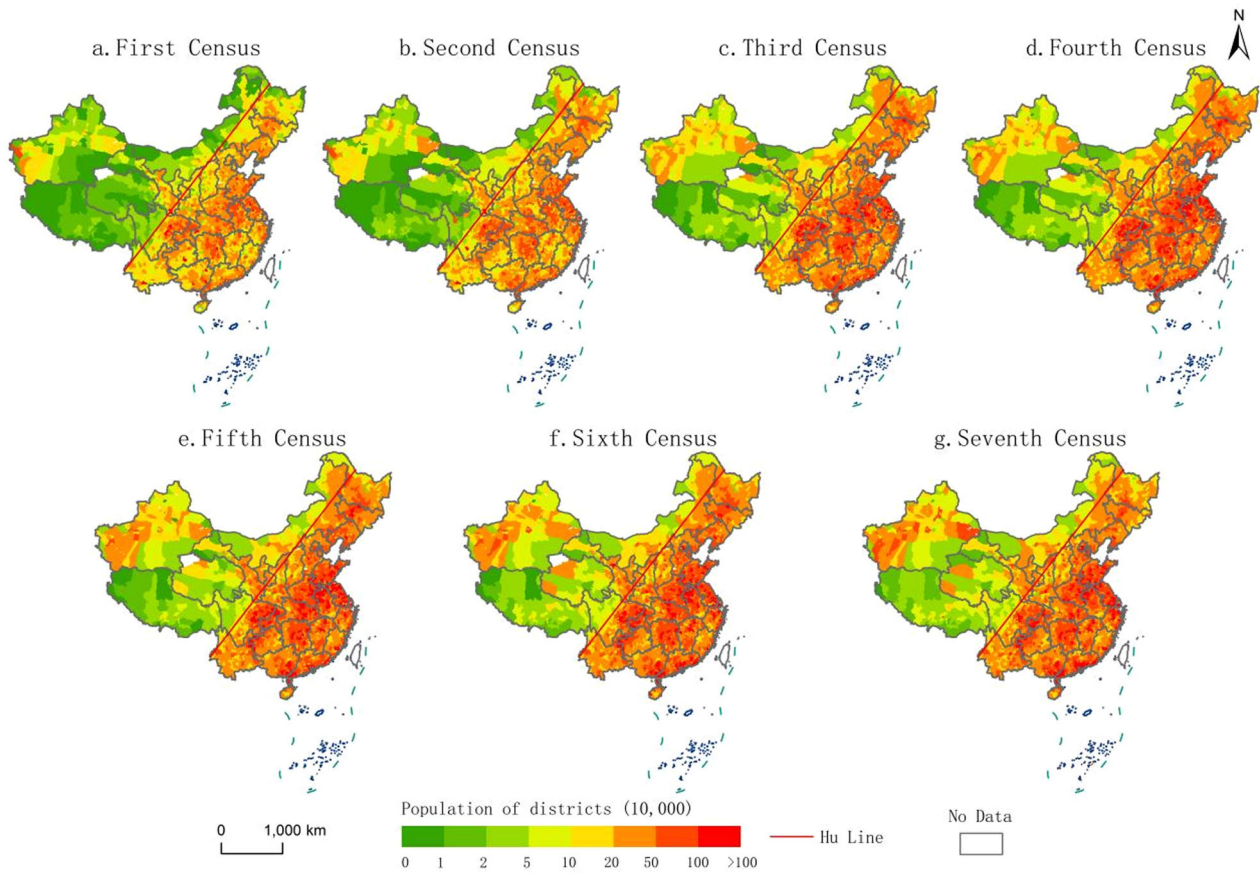
Fig. 1 Spatial distribution of county-level population in China. Note: Based on the Department of Natural Resources Standard Map Service website GS (2019)1815. Standard maps are produced with no modifications to the base map boundaries, same as below. **a** First census population distribution, **b** Second census population distribution, **c** Third census population distribution, **d** Fourth census population distribution, **e** Fifth census population distribution, **f** Sixth census population distribution, **g** Seventh census population distribution.
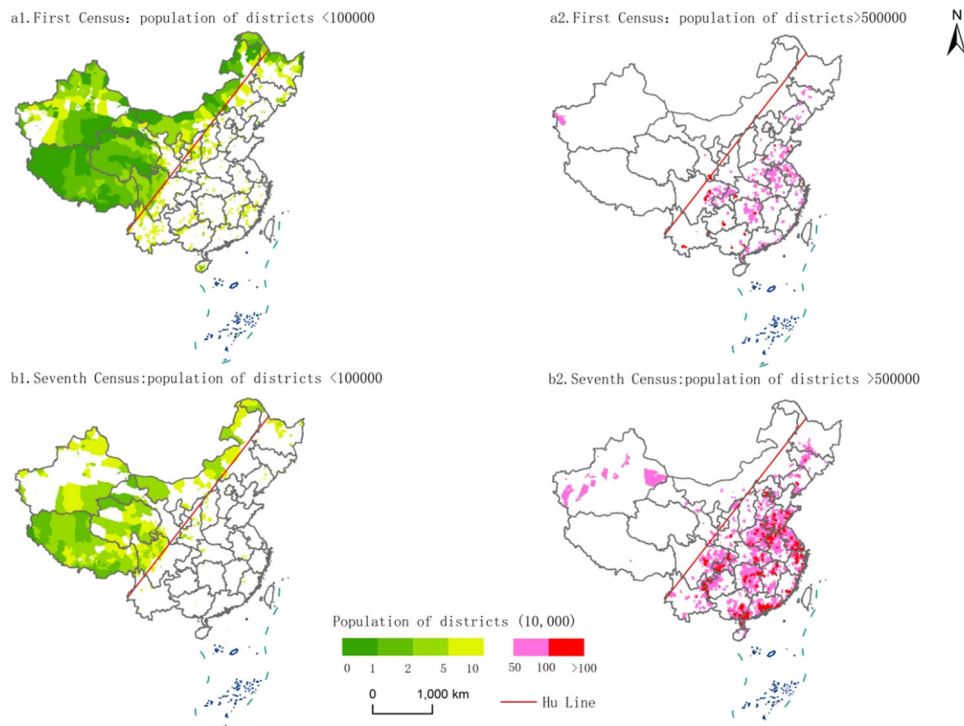
Fig. 2 Spatial distribution of specific population at county-level in China. a1 First census counted less than 100,000 people, **a2** First census counted more than 500,000 people, **b1** Seventh census counted less than 100,000 people, **b2** Seventh census counted more than 500,000 people.
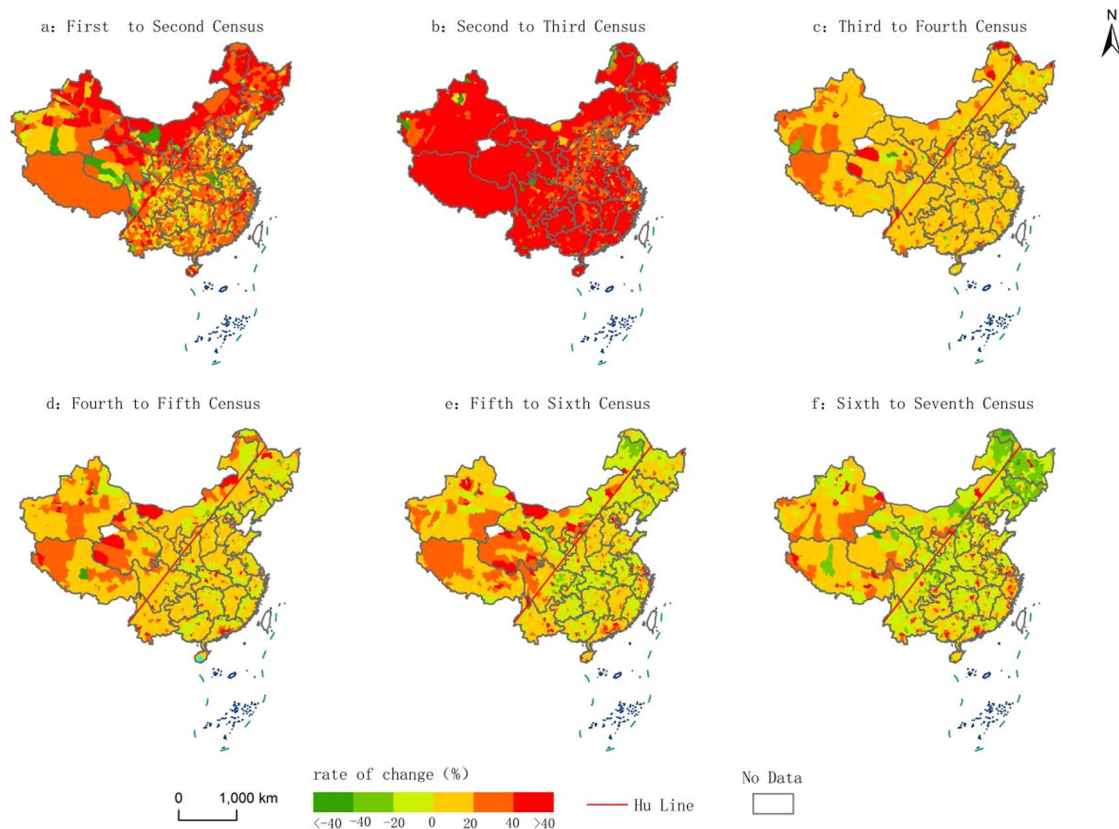
Fig. 3 Map of county-level population change rate in China. a Change rate of the first to second census, **b** Change rate of the second to third census, **c** Change rate of the third to fourth census, **d** Change rate of the fourth to fifth census, **e** Change rate of the fifth to sixth census, **f** Change rate of the sixth to seventh census.

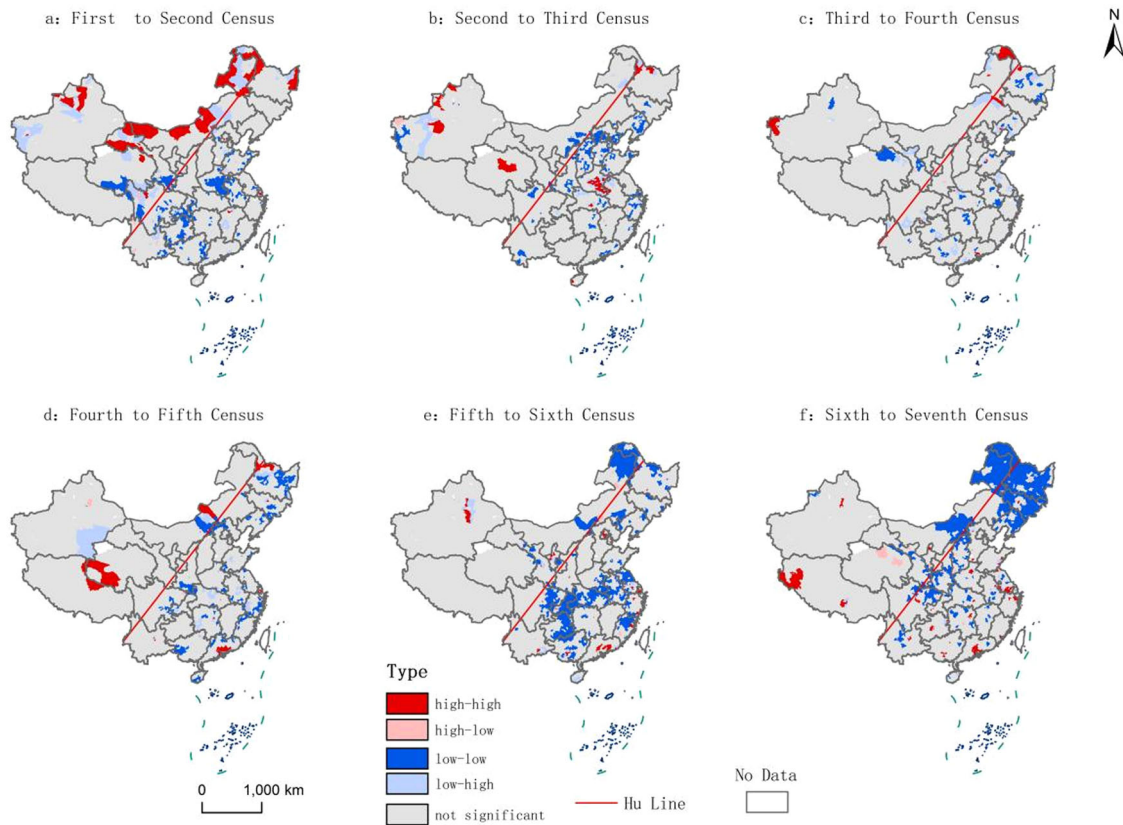| Table 1 Global Moran's *I* index of population change rate in China.					
Time	**First to second census**	**Second to third census**	**Third to fourth census**	**Fourth to fifth census**	**Fifth to sixth census**	**Sixth to seventh census**
Moran's *I*	0.189	0.011	−0.002	0.017	0.074	0.189

and the significant positive autocorrelation exists in the spatial distribution of the rest of the several censuses indicates that there is significant spatial agglomeration in the spatial contraction of the population, and from the fourth census onwards, the global Moran index continues to rise, and the distribution of the population shows a significant enhancement trend. Based on GeoDa, the univariate local Moran index analysis was conducted, and LISA maps were generated to characterize the distribution of spatial units in terms of agglomeration and deagglomeration, in which "high-high" agglomeration zones refer to the areas of continuous and substantial population growth, and "low-low" agglomeration zones refer to the areas of concentrated and contiguous population decline. The "high-high" agglomerations refer to areas of large-population growth and the "low" agglomerations refer to areas of population decline. Taking the recent sixth to seventh population census as an example, there are a total of 138 high and high agglomerations, mainly located in Changchun City, Jilin Province; Qingdao City, Shandong Province; Zhengzhou City, Henan Province; Wuhan City, Hubei Province; Changsha City, Hunan Province; Chengdu City, Sichuan Province; Nanjing City, Suzhou City, and Wuxi City, Jiangsu Province; Hangzhou City, Hangzhou City, and Ningbo City, Zhejiang Province; Guiyang City, Guizhou Province; Nanning City, Guangxi Province; Guangzhou City, Guangzhou and the

East Aster area, Guangdong Province; Xiamen City, Fujian Province; and Xizang City, Tibet. The total number of low-low agglomerations is 361, mainly in Heijiliao and Liaojiao Provinces, the southeast of Gansu Province, the central part of Inner Mongolia, the south of Shaanxi Province, the east of Sichuan Province, and the central part of Yunnan Province. It can be seen that the "high-high" agglomeration areas are mainly distributed around the provincial capital cities, while the "low-low" agglomeration areas are mainly distributed in the northeast border area and the economically underdeveloped western region, indicating that the agglomeration capacity of China's provincial capital cities has been comprehensively improved and the population in northeast China shows an obvious contraction trend (Table 1 and Fig. 4).

Analysis of the type and equilibrium of population evolution in counties

Types of population change in the county. The population density of counties and counties from the first to the seventh national population density and population growth rate of counties and counties from the first to the seventh national population growth rate of two types, coupled to get the population of China's counties and the magnitude of change in the combination of the type of

Fig. 4 LISA map of county-level population change rate in China. a First to second census LISA map, **b** Second to third census LISA map, **c** Third to fourth census LISA map, **d** Fourth to fifth census LISA map, **e** Fifth to sixth census LISA map, **f** Sixth to seventh census LISA map.

analysis from the table below can be concluded, the first to the second national census of low-density rapid growth as the main trend of change accounted for the total number of counties and counties of 26.72%, the second to the third national census of the magnitude of its growth increased in the medium-density rapid growth in the counties and counties accounted for the total number of counties and counties of 42.28%, the thrid to the fourth census, the fourth to the fifth census, the fifth to the sixth census was mainly medium-density stable growth trend and the growth rate was gradually downward trend from 42.59% to 22.16%, the sixth to the seventh census there are 777 counties and districts for the medium-density negative growth, accounting for 27.46% of the total number of counties and districts, from the first to the seventh population changes can be seen that China is about to usher in an inflection point of the population, slowing down or slowing down the growth of the total number of counties and districts, the population is about to change. From the type of population change from the first to the seventh census, it can be seen that China is approaching the inflection point of its total population, and that slower or negative growth will become a general rule that should be given great attention in the future (Table 2).

Taking the sixth to the seventh general population survey as an example, the low-density areas are mainly concentrated in the three northeastern provinces and the southern part of Yunnan Province, the Qian-Gui junction area and the northwestern half of the Hu Huanyong Line; among them, there are 408 counties and districts with negative low-density growth, accounting for 14.42% of the total counties and districts, mainly located in the northeast of the large and small Xing'anling and the Changbai mountain range, the Junggar Basin of Xinjiang, the Yinshan Mountain Range of Inner Mongolia, the Qilian Mountains of the Gan-Qing region and the mountainous areas of Yunnan and

Sichuan provinces. Yunnan Province and Sichuan Province in the mountainous areas; low-density stable growth of 220 counties, accounting for 7.78% of the total number of counties, mainly in Xinjiang, Tibet, Inner Mongolia, Qinghai, Ningxia, Gansu, the north and the western part of Sichuan Province; low-density fast-growing counties and districts there are only 39 of its share of a relatively small, mainly in Xinjiang, Hami, Changji, Hotan, Aksu, Lhasa, Tibet, Ali, Linzhi, Chamdo In the west of Xinjiang Province, there are only 39 counties with low-density rapid growth, mainly in Hami, Changji, Hotan and Aksu in Xinjiang, Lhasa, Ali, Linzhi and Chamdo in Tibet, Ganzi and Golog in Sichuan and Qingdao, and Ordos and Yulin in Neimenggu and Shaanxi.

High-density areas are primarily concentrated in the North China Plain, the plains of the middle and lower reaches of the Yangtze River, the southeastern coast, and the border areas of Sichuan Province and Chongqing Municipality. There are 344 counties and districts experiencing negative high-density growth, constituting 12.16% of the total. These areas are mainly situated in the eastern part of Tianjin Municipality, the central part of Hebei Province, the northern and central part of Henan Province, the central part of Shaanxi Province, the central part of Jiangsu Province, the central and southern parts of Shandong Province, the eastern parts of Sichuan Province, Tianmen City, Xiaogan City in Hubei Province, Loudi City, Hengyang City in Hunan Province, and the eastern coastal area of Guangdong Province. Furthermore, 11.81% of the total number of counties and districts exhibit high-density and stable growth, mainly concentrated in Shandong, Henan, Hebei, and the southeastern coastal area. Another 278 counties and districts, accounting for 9.83% of the total, show high-density and rapid growth. These areas are primarily distributed in Beijing, Tianjin, Chongqing, Xi'an City in

Table 2 Coupling characteristics of population density and growth types in counties in China in the first and seventh counties.

Density		First to second census	Second to third census	Third to fourth census	Fourth to fifth census	Fifth to sixth census	Sixth to seventh census
Low density (<100 persons per km²)	Negative growth (<0) (number of districts, percentage)	(107, 3.78%)	(18, 0.64%)	(33, 1.17%)	(136, 4.81%)	(235, 8.30%)	(408, 14.42%)
	Stable growth (0–25%)(number of districts, percentage)	(341, 12.05%)	(33, 1.17%)	(650, 22.98%)	(483, 17.07%)	(348, 12.30%)	(220, 7.78%)
	Rapid growth (>25%)(number of districts, percentage)	(756, 26.72%)	(1005, 35.52%)	(56, 1.98%)	(60, 2.12%)	(70, 2.47%)	(39, 1.38%)
Medium density (100–500 persons per km²)	Negative growth (<0) (number of districts, percentage)	(175, 6.19%)	(29, 1.03%)	(84, 2.97%)	(324, 11.45%)	(552, 19.51%)	(777, 27.46%)
	Stable growth (0–25%) (number of districts, percentage)	(436, 15.41%)	(89, 3.15%)	(1205, 42.59%)	(899, 31.77%)	(627, 22.16%)	(339, 11.98%)
	Rapid growth (>25%) (number of districts, percentage)	(641, 22.66%)	(1196, 42.28%)	(51, 1.80%)	(76, 2.69%)	(91, 3.22%)	(91, 3.22%)
High density (>500 persons per km²)	Negative growth (<0) (number of districts, percentage)	(49, 1.73%)	(54, 1.91%)	(107, 3.78%)	(125, 4.42%)	(243, 8.59%)	(344, 12.16%)
	Stable growth (0–25%) (number of districts, percentage)	(45, 1.59%)	(36, 1.27%)	(589, 20.82%)	(620, 21.91%)	(438, 15.48%)	(334, 11.81%)
	Rapid growth (>25%) (number of districts, percentage)	(279, 9.86%)	(369, 13.04%)	(54, 1.91%)	(107, 3.78%)	(226, 7.99%)	(278, 9.83%)

Shaanxi Province, Zhengzhou in Henan Province, Zhengzhou City in Henan Province, Changsha City in Hunan Province, Chengdu City in Sichuan Province, Kunming City in Yunnan Province, Jinan City, Linyi City in Shandong Province, the northern part of Zhejiang Province, the southern part of Jiangsu Province, the eastern part of Fujian Province, and the eastern part of Guangdong Province (Fig. 5).

Changes in the Gini coefficient of the population. The population-land distribution in China's counties is extremely unbalanced, and the trend of spatial concentration in the distribution of the population has increased significantly. The Gini coefficient of China's spatial distribution of the population is higher than 0.7 during the period from the first to the seventh population census, showing the obviously unbalanced distribution of population and land, of which the Gini coefficient of population-land distribution is the smallest in the fourth population census, and the smallest in the seventh population census, which is 0.7504 and the Gini coefficient of the seventh population census is 0.7895. The Gini coefficient shows a gradually increasing trend from the fourth to the seventh census, with a rapid increase of 0.0388 from the third to the seventh census, indicating that the trend of population centralization is very significant from the third census. The Gini coefficients of the population on both sides of the Hu Huanyong line show that the Gini coefficient of the southeastern half of the wall is about 0.5, the Gini coefficient of the northwestern half of the wall is about 0.8, and the Gini coefficient of the northwestern half of the wall is about 1.6 times that of the southeastern half of the wall, which further illustrates the imbalance in the distribution of the population on both sides of the Hu Huanyong line.

The trend of rapid centralization of population distribution is also evident in population density. During the period from the first to the seventh census, the number of counties and districts with population densities of 500 persons per km² gradually increased from 373 to 957. The ratio of the population residing in these districts to the total population of the country grew from 21.59% to 53.88%. However, the land area of these districts expanded from 1.41% to 6.74% of the land area of the whole country. At the time of the first census, the 1314 counties and districts with a population density of more than 150 persons per km² accounted for a total population ratio of 68.06% and a land area ratio of 13.15%. By the time of the seventh census, the number of counties and districts with a population density of more than 150 persons per km² had increased to 1845. These areas had a total population ratio of 85.5% and a land area ratio of 23.55%, indicating that approximately 24% of China's land supports 86% of the population.

This further indicates that the spatial agglomeration of China's population distribution continues to increase and that China's population is accelerating its concentration in a few regions (Table 3 and Fig. 6).

China's population sequence forecasting based on ARIMA modeling

Smoothness analysis of time-series data. Population, as an important factor in the study of urban of the urban system, plays an important role in grasping the characteristics of population distribution in each region, predicting the future population and the optimal allocation of resources in each region, and can strongly support the optimal layout of the urban system. Scientific and reasonable prediction of the population is of great significance for the formulation of the countermeasures and measures for the coordinated development of the population, resources, and the environment, and for the compilation of territorial spatial planning. In this paper, we choose to use the total
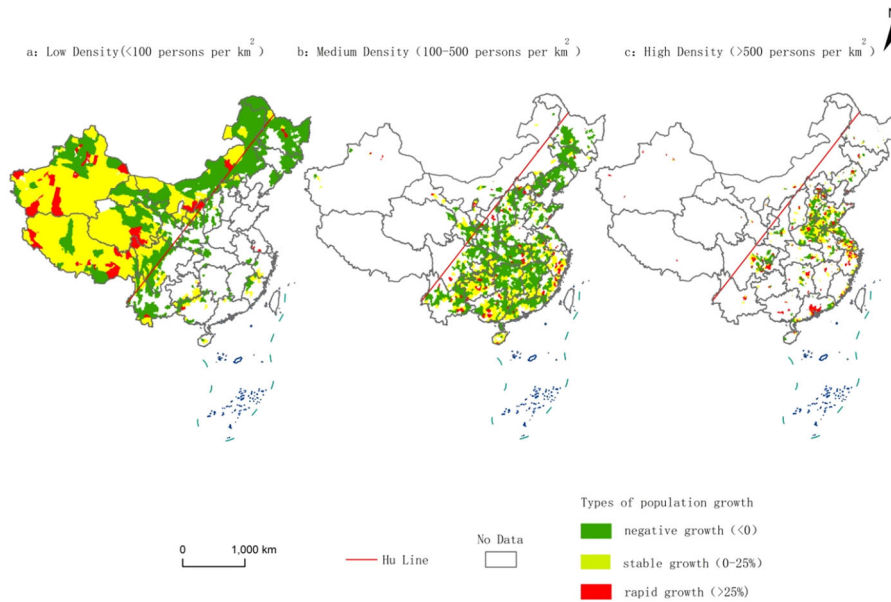
Fig. 5 Type of county-level population change in China from the sixth to seventh census. a Low population density (<100 persons/km^2), **b** Medium population density (100-500 persons/km^2), **c** High population density (>500 persons/km^2).

		First census	Second census	Third census	Fourth census	Fifth census	Sixth census	Seventh census
Table 3 Changes in the Gini coefficient of population distribution of counties in China from the first to the seventh general population.								
	China	0.7769	0.7638	0.7508	0.7504	0.7543	0.7671	0.7896
Hu huan yong line	Half of the south-east wall	0.5547	0.5362	0.5049	0.5119	0.5234	0.5526	0.5978
	Half of the north-western wall	0.8436	0.8274	0.8092	0.8004	0.8008	0.8024	0.8146

population from 1982 to 2022 as the research object, with a sample capacity of $n = 41$ (data from China Statistical Yearbook), by modeling China's population growth trend and forecasting China's future population development trend. Since the population number is a typical time-series indicator, this paper uses the ARIMA model to fit the population series, and this paper utilizes Eviews 12.0 to do the time-series plot of the original series data to get Fig. 7.

From the time-series graph in Fig. 7, we can find that the population of China is generally increasing year by year, and does not fluctuate up and down around a certain mean value, so it can be known that the series is a non-stationary series. At the same time, the unit root test is performed on the original population data series, and Table 4 is obtained.

As shown in Table 4, the value of the t-statistic of the ADF test is −2.41, which is obviously larger than the critical value of −2.60 with a significance level of 10%, so it is determined that there is a unit root in the series, and it is a non-stationary series. Since the population size is a typical stock indicator and has a continuous growth trend, the series needs to be transformed into a smooth series by differential treatment.

From the test results in Table 5, it can be seen that the p-value is 0.0001 after the second-order differencing and the t-statistic value is −5.49, which is obviously smaller than the critical value of −2.6 at the significant level of 10%, so it can be judged that the sequence no longer exists as a unit root after the second-order differencing, and therefore, the ARIMA (p, d, q) model is introduced to determine that the order of the differencing is the second order. From Fig. 8, it can be seen that the autocorrelation

and partial autocorrelation coefficients of the second-order differenced series fall within the dashed line of the random interval, indicating that the series is smooth and can be used.
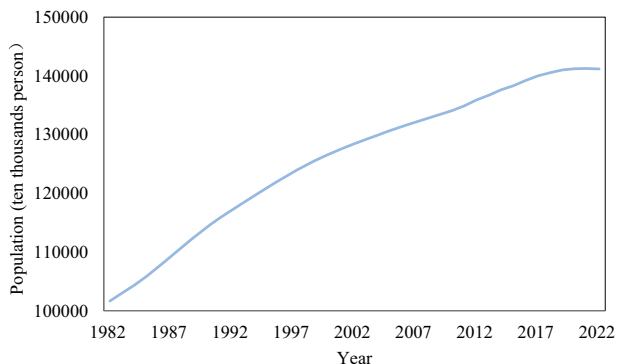
Model identification, building. Since the trend of the series is eliminated after the second-order differencing, resulting in $d = 2$ in the ARIMA (p, d, q) model, it can be seen from Fig. 8 that the autocorrelation coefficient and partial autocorrelation coefficient of the series both begin to decay after the second-order lag. Thus, it can be assumed that the values of p and q for the model are both 2. After comparing and optimizing the results of the model through the establishment of multiple models, it is determined that the prediction model of the present paper is the ARIMA (2, 2, 2) model, and the results of the constructed model are as follows.

From Table 6, it can be seen that the parameter results of the ARIMA model: p-value of AR (1), AR (2), MA (1), MA (2), and the value of the constant term C are all less than 0.05, which passes the significance test. All autocorrelation coefficients and partial autocorrelation coefficients in Fig. 9 are within the dashed line, and the p-values are all greater than 0.05 progressively approaching 1, indicating that the residuals are random sequences that satisfy the white noise test and that the ARIMA (2, 2, 2) model can be used for prediction.

Population projections based on the ARIMA (2, 2, 2) model. By using Eviews software, the population size of China is analyzed. The original population sequence was first tested for stabilization and unit root to determine the difference order $d = 2$. Then, the ARIMA model was identified on the model, and p and q values
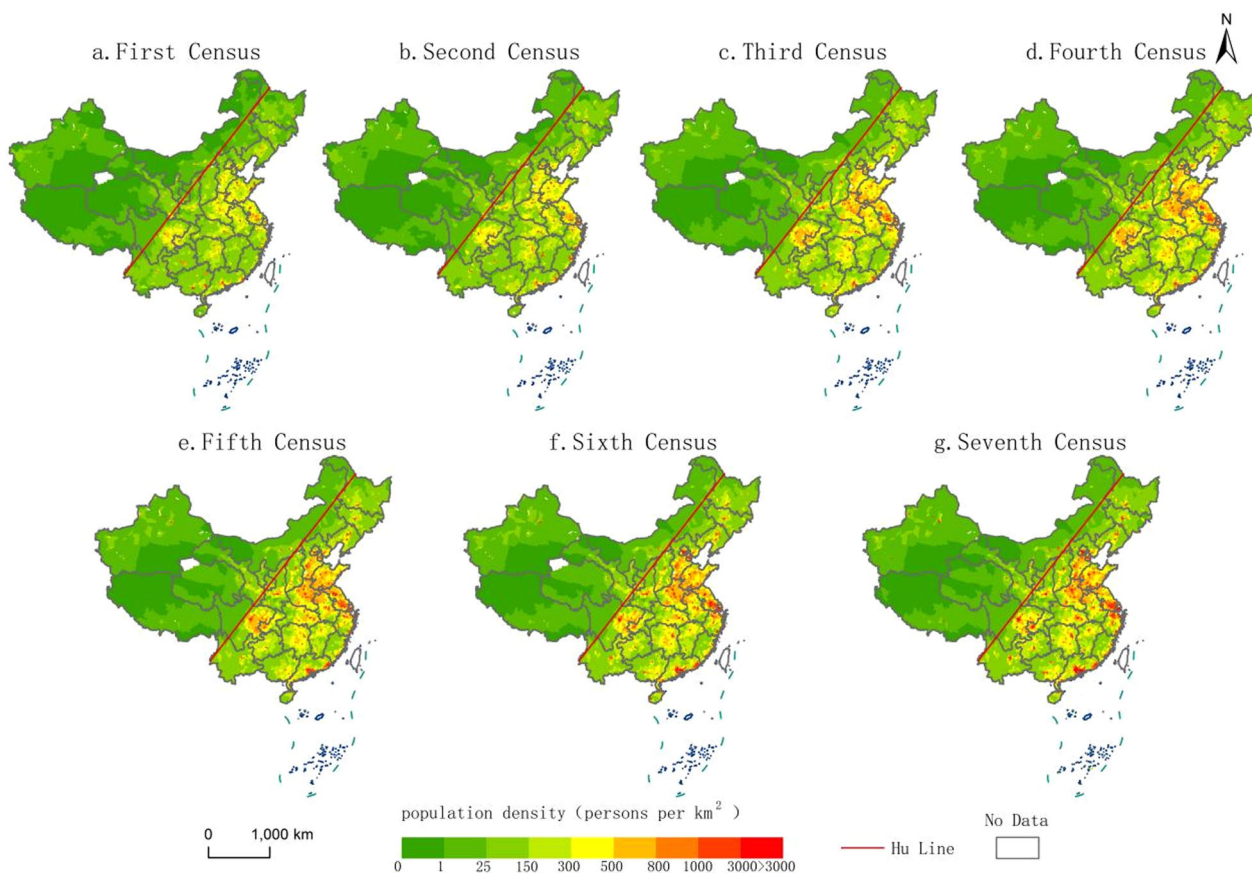
Fig. 6 Spatial distribution of county-level population density in China. a First census population density distribution, **b** Second census population density distribution, **c** Third census population density distribution, **d** Fourth census population density distribution, **e** Fifth census population density distribution, **f** Sixth census population density distribution, **g** Seventh census population density distribution.

Fig. 7 Time-series distribution of China's population.

Table 4 Unit root test results of the original population series.

		t-statistic	Prob.*
Augmented Dickey-Fuller test statistic		−2.415566	0.1442
Test critical values:	1% level	−3.610453	
	5% level	−2.938987	
	10% level	−2.607932	

*MacKinnon (1996) one-sided *p*-values.

were determined by the model parameter estimation using autocorrelation and partial autocorrelation tests. Finally, the optimal model was identified as ARIMA (2, 2, 2), and this model was used to predict the population of China from 2023 to 2035. Figure 10 shows that China's population gradually decreased from 1410.05 million in 2023 to 1343.68 million in 2035.

Conclusion and discussion

Conclusion. Based on the data from the first to the seventh census, this paper constructs the county population spatial database of the first to seventh census administrative divisions, establishes the ARIMA model to predict the population in China

from 2023 to 2035, and finds that the county population in China shows an obvious growth trend and regional differences; the population growth in China or negative will become a general rule; the county population and land distribution in China is extremely unbalanced and the spatial concentration trend of population distribution is obviously strengthened; the future population development in China will face the challenge of negative population growth.

As one of the two major factors of production, the change of labor is bound to have an impact on China's economic development. Since 2022, China's total population has entered negative growth, and by 2035, China's population will be less than 1.35 billion. The low number of births will inevitably aggravate the aging of the population, which puts forward a higher demand for China to promote the long-term balanced development of the population and promote the transformation of emergency aging.

For the future coming labor shortage era, the Chinese government based on equal basic public services of county contraction, complies with the trend of population, changes the traditional concept of "growth", "contraction" as the opportunity for future transformation. Simultaneously, it is enhancing infrastructure, guiding the transformation of population loss development through convenient and improved measures, and actively formulating effective development strategies. The government is also working on the rational optimization of population spatial layout to improve labor productivity in response to population growth.

From the above data results, we can see that the application of the ARIMA model in population prediction in China is somewhat scientific, and its data also has certain statistical significance. It can help China to find out the key to the problem in time and to put forward timely solutions to this problem, so as to promote the growth of people's living standards and economic level. The ARIMA model is one of the most widely used methods in time-series analysis technology. Its theory is mature, taking into account factors such as the trend of the sequence, cyclical changes, and random interference. This consideration makes the fitting results theoretically convincing and applicable in practice, promoting the widespread use of the ARIMA model in various fields. It can provide a valuable reference for future directions.

Fig. 9 AERMA (2, 2, 2) residual series autocorrelation and partial autocorrelation plots.

Table 5 Unit-root test results for second-order difference population series.

		t-statistic	Prob.*
Augmented Dickey-Fuller test statistic		−5.490382	0.0001
Test critical values:	1% level	−3.615588	
	5% level	−2.941145	
	10% level	−2.609066	

*MacKinnon (1996) one-sided p-values.

Fig. 8 Autocorrelation and partial autocorrelation plots of second-order difference series.
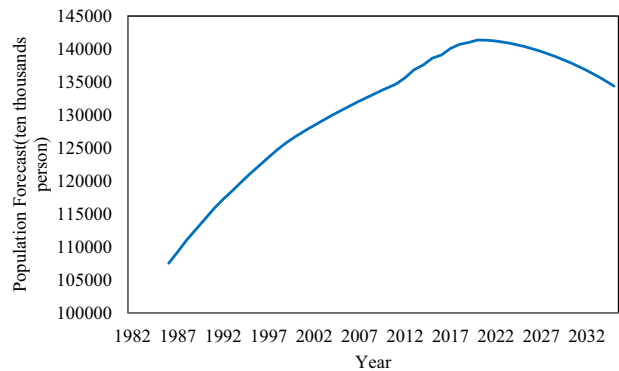
Fig. 10 China population model projection results.

Table 6 Table of correlation parameters for AERMA (2, 2, 2) fitting.

Variable	Coefficient	Std. error	t-statistic	Prob.
C	−53.83592	22.63544	−2.378391	0.0235
AR(1)	1.272941	0.13383	9.511631	0.0000
AR(2)	−0.646288	0.105829	−6.106898	0.0000
MA(1)	−1.455524	0.085314	−17.06074	0.0000
MA(2)	0.947317	0.061452	15.41562	0.0000
R-squared	0.30397	Mean dependent var		−42.67568
Adjusted R-squared	0.216966	S.D. dependent var		117.7943
S.E. of regression	104.2353	Akaike info criterion		12.25627
Sum squared resid	347,679.7	Schwarz criterion		12.47396
Log-likelihood	−221.7409	Hannan-Quinn criter.		12.33301
F-statistic	3.493753	Durbin-Watson stat		2.091987
Prob (F-statistic)	0.017791			
Inverted AR roots	0.64−0.49i	0.64 + 0.49i		
Inverted MA roots	0.73−0.65i	0.73 + 0.65i		

Discussion. Population is a prominent global issue. By conducting spatial and temporal analysis of the population in Chinese counties from the first to the seventh national census, we can gain a detailed understanding of the spatial and temporal distribution of China's population. This enables us to make predictions about China's population and comprehend the patterns of spatial and temporal distribution of county population, providing valuable insights for future trends in population change. The time-series analysis reveals that China's population exhibited a trend of low-density rapid growth, followed by medium-density rapid growth, stable growth, and ultimately negative growth. This long-term negative growth trend poses potential risks to the sustainable development of regional socio-economics (Yuan et al., 2023). Furthermore, from a spatial perspective, the distribution of China's population still follows the approximate ratio of 96:4 between the southeastern half of the Hu Huanyong line and the northwestern half of the country, as previously concluded by Qi et al. (2015). This distribution pattern is primarily influenced by economic and administrative factors. He (2018) points out that the quantity, structure, and distribution of the population are fundamental conditions for national development and also serve as a metric for measuring economic and social development. These factors directly influence the value orientation and path choices for China's future development. From the perspective of population projections, it is evident that China's total population has entered a negative growth phase after 2022, with the population projected to be less than 1.35 billion by 2035. Furthermore, there is an anticipated phased acceleration trend in the future, aligning with the research findings of Chen (2022). Additionally, a phased acceleration trend is expected in the future. However, there are some limitations in this study. Firstly, it does not analyze the influencing factors of population change using specific indicators. Secondly, considering the regional variations in population distribution and dispersion, an in-depth examination of population dispersion characteristics across subregions (east, middle, west, north, south, southeast, northwest, and southwest) was not conducted. Finally, while this study focuses on the county administrative unit scale, townships, as the smallest administrative unit of China's census data publicly released, can be a more detailed and specific description of the spatial pattern of China's population distribution. Further research should explore the population distribution of townships in China to enhance the scale and dimension of population geography enhance.

Data availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

References

Aalen OO (1989) A linear regression model for the analysis of life times. Stat Med 8(8):907–925

Alkema L, Raftery AE, Gerland P, Clark SJ, Pelletier F, Buettner T, Heilig GK (2011) Probabilistic projections of the total fertility rate for all countries. Demography 48(3):815–839

Anggreini D (2017) The Female Population Growth Projection Year 2021 in Trenggalek Regency by Leslie Matrix Model on the Birth Rate and Life Expectancy. Biol Med Nat Prod Chem 6(2):37–45

Anselin L (1988) Spatial econometrics: methods and models. Springer, New York

Anselin L (1995) Local indicators of spatial association—LISA. Geogr Anal 27(2):93–115

Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A (2006) Determining global population distribution: methods, applications and data. Adv Parasit 62:119–156

Balsa-Barreiro J, Li Y, Morales A (2019) Globalization and the shifting centers of gravity of world's human dynamics: Implications for sustainability. J Clean Prod 239:117923

Balsa-Barreiro J, Morales AJ, Lois-González RC (2021) Mapping population dynamics at local scales using spatial networks. Complexity 2021:1–14

Booth H, Hyndman RJ, Tickle L, De Jong P (2006) Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. Demogr Res 15:289–310

Briggs DJ, Gulliver J, Fecht D, Vienneau DM (2007) Dasymetric modelling of small-area population distribution using land cover and light emissions data. Remote Sens Environ 108(4):451–466

Chen W (2022) Forecasting negative population growth and population ageing in China. Soc Sci J (05):133–144 (In Chinese)

Chi G (2012) The impacts of transport accessibility on population change across rural, suburban and urban areas: a case study of Wisconsin at sub-county levels. Urban Stud 49(12):2711–2731

Clarke JI (1978) Population geography. Prog Hum Geogr 2(1):163–169

Congdon P, Batey P, Regional Science Association (1990) Advances in regional demography: information, forecasts, models. J R Stat Soc 39(3):382–384

Dobson JE, Bright EA, Coleman PR et al. (2000) LandScan: a global population database for estimating populations at risk. Photogramm Eng Remote Sens 66(7):849–857

Ellis EC, Ramankutty N (2008) Putting people in the map: anthropogenic biomes of the world. Front Ecol Environ 6(8):439–447

Franklin RS (2014) An examination of the geography of population composition and change in the United States, 2000–2010: Insights from geographical indices and a shift-share analysis. Popul Space Place 20(1):18–36

Gerland P, Raftery AE et al. (2014) World population stabilization unlikely this century. Science 346(6206):234–237

Gilbert K (2005) An ARIMA supply chain model. Manag Sci 51(2):305–310

Goldewijk KK (2005) Three centuries of global population growth: a spatial referenced population (density) database for 1700-2000. Popul Environ 26:343–367

Guan X, Wei H, Lu S, Su H (2018) Mismatch distribution of population and industry in China: pattern, problems and driving factors. Appl Geogr 97:61–74

He D (2018) Strengthening research to address the population challenges in the new era. Popul Res 42(02):3–6. (In Chinese)

Hu HY (1983) Research on population distribution in China. East China Normal University Press, Shanghai, (In Chinese)

Liang Z, Li Z, Ma Z (2014) Changing patterns of the floating population in China, 2000–2010. Popul Dev Rev 40(4):695–716

Liao Y, Wang J, Meng B, Li X (2010) Integration of GP and GA for mapping population distribution. Int J Geogr Inf Sci 24(1):47–67

Li S, Yang Z, Li H, Shu G (2018) Projection of population structure in China using least squares support vector machine in conjunction with a Leslie matrix model. J Forecast 37(2):225–234

Liu Y, Stillwell J, Shen J, Daras K (2014) Interprovincial migration, regional development and state policy in China, 1985–2010. Appl Spat Anal Policy 7:47–70

Liu XH, Kyriakidis PC, Goodchild MF (2008) Population-density estimation using regression and area-to-point residual kriging. Int J Geogr Inf Sci 22(4):431–447

Miller HJ (2004) Tobler's first law and spatial analysis. Ann Assoc Am Geogr 94(2):284–289

Newbold P (1983) ARIMA model building and the time series analysis approach to forecasting. J Forecast 2(1):23–35

Poole MA, O'Farrell PN (1971) The assumptions of the linear regression model. Trans Inst Br Geogr 52:145–158

Piccolo D (1990) A distance measure for classifying ARIMA models. J Time Ser Anal 11(2):153–164

Qi W, Liu S, Zhao M, Liu Z (2016) Chinas different spatial patterns of population growth based on the "Hu Line". J Geogr Sci 26:1611–1625

Qi W, Liu S, Zhao M (2015) Study on the stability of Hu Line and different spatial patterns of population growth on its both sides. Acta Geogr Sin 70(4):551–566. (In Chinese)

Raftery AE, Chunn JL, Gerland P et al. (2013) Bayesian probabilistic projections of life expectancy for all countries. Demography 50(3):777–801

Ramankutty N, Foley JA (1999) Estimating historical changes in global land cover: Croplands from 1700 to 1992. Glob Biogeochem Cy 13(4):997–1027

Ramankutty N, Foley JA, Olejniczak NJ (2002) People on the land: changes in global population and croplands during the 20th century. AMBIO 31(3):251–257

Shumway RH, Stoffer DS, Shumway RH, Stoffer DS (2006) ARIMA models. Time series analysis and its applications: with R examples. Springer, New York

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46(sup1):234–240

Wang ZX, Li Q, Pei LL (2018) A seasonal GM (1, 1) model for forecasting the electricity consumption of the primary economic sectors. Energy 154:522–534

Ye D, Xu W, He W, Li Z (2001) Symmetry distribution of cities in China. Sci China Earth Sci 44:716–725

Yuan H (1993) Study on the rationality of Chinese population distribution. Geogr Res 12(3):64–69. (In Chinese)

Yuan X, Du P, Tong YF, Lu JH, Yin DT (2023) Negative population growth and high-quality population development in China. J Beijing Adm Coll (03):11–25 (In Chinese)

Yue TX, Wang YA, Liu JY, Chen SP et al. (2005) Surface modelling of human population distribution in China. Ecol Model 181(4):461–478

Zhuo L, Ichinose T, Zheng J, Chen J, Shi PJ, Li X (2009) Modelling the population density of China at the pixel level based on DMSP/OLS non-radiance-calibrated night-time light images. Int J Remote Sens 30(4):1003–1018

Acknowledgements

Author contributions

Conceptualization: Mei Sang, Jing Jiang, and Xin Huang; Methodology: Mei Sang and Jing Jiang; Formal analysis and investigation: Mei Sang, Jing Jiang, Xin Huang, Feifei Zhu, and Qian Wang; Writing—original draft preparation: Mei Sang and Jing Jiang; Writing—review and editing, Mei Sang and Jing Jiang; Funding acquisition: Jing Jiang; Supervision: Jing Jiang. All authors have read and agreed to the submitted version of the manuscript.

Competing interests

The authors declare no competing interests.

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-024-02784-1.

Correspondence and requests for materials should be addressed to Jing Jiang.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.