Chin. Geogra. Sci. 2012 Vol. 22 No. 3 pp. 367–380
doi: 10.1007/s11769-012-0534-0
www.springerlink.com/content/1002-0063

Origin Distribution Patterns and Floating Population Modeling: Yiwu City as a Destination

LI Hongsheng, WANG Yingjie, HAN Jiafu

(*State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China*)

Abstract: Existing quantitative migration studies are mainly at the inter-region or inter-province level for lacking of detailed geo-referenced migration data. Meanwhile, few of them integrate explorative spatial data analysis and spatial regression model into migration analysis. Based on aggregated registered floating population data from 2005 to 2008, the phenomena that population floating to Yiwu City in Zhejiang Province is analyzed at the provincial and county levels. The spatial layout of Yiwu′s pull forces is proved as a V-shaped pattern excluding Sichuan Province based on map visualization method. Using the migration ratio in 2007 as an explanatory variable, two models are compared using ordinary least square, spatial error model and spatial lag model methods for county-level data in Jiangxi and Anhui provinces. The model with migration stock provides an improved fitting over the model without migration stock according to the model fitting results. The floating population flocking into Yiwu City from Jiangxi is determined mostly by migration stock while the determinant factors are migration stock and distance to Yiwu City for Anhui. The distance-decay effect is true for migration flow from Anhui to Yiwu City while the distance rule is not confirmed in Jiangxi with the best fitting model. The correlation between per capita net income of rural labor forces and migration ratio is not significant in Jiangxi and significant but at the 0.1 level only in Anhui. Further analysis shows that the distance, income and man-land ratio are important factors to explain population floating at earlier stage. However, as the dynamic population floating process evolves, the determinant factor would be migration stock.

Keywords: floating population; origin distribution visualization; spatial regression model; Yiwu City; GIS

Citation: Li Hongsheng, Wang Yingjie, Han Jiafu, 2012. Origin distribution patterns and floating population modeling: Yiwu City as a destination. *Chinese Geographical Science*, 22(3): 367–380. doi: 10.1007/ s11769-012-0534-0

1 Introduction

Due to the rapid growth of market economy, China is witnessing great population movements which have never been seen in recent years. Currently, there is more than 2.0×10^8 floating population such as migrant workers all over the country according to National Population and Family Planning Commission of China at the end of 2008. In order to understand and evaluate the relevant challenges, it is necessary to analyze the origin and destination of population migration/floating in China and develop migration models for understanding migration determinants.

Ravenstain (1885) discussed the reasons and rules of migration for the first time, and was credited with the origination of distance decay theories of migration. Bogue (1969) proposed the ′push and pull theory′ which indicated that people move and migrate to other places only when their living conditions can be improved. Lee (1966) extended the theory and conceptualized migration as involving origins, destinations and the links between them. The migration of population is motivated by the push forces from the origin as well as the pull forces from the destination. The links imposes a cost on

Received date: 2011-10-27; accepted date: 2011-12-27

Foundation item: Under the auspices of National Natural Science Foundation of China (No. 41001314), Youth Science Funds of State Key Laboratory of Resources and Environmental Information System, Chinese Academy of Sciences (No. KA11040101), National Key Technology R & D Program of China (No. 2012BAI32B07)

Corresponding author: WANG Yingjie. E-mail: wangyj@igsnrr.ac.cn

the migration either directly, as in cost of searching for a job, or indirectly through affecting the amount of information available about a destination at an origin. Many literatures on migration modeling were published in recent several decades. Spatial interaction model based on gravity theory is the most common type of migration models and has been widely used in migration analysis (Haynes and Fotheringham, 1984; Fotheringham and O′Kelly, 1989). Multinomial logit model to migration based on discrete choice theory (Manski, 1977) has been applied in micro level (Liaw and Ledent, 1987; Liaw, 1990). In these models, the determinant factors of population migration have been categorized as follows: demographic factors, social and cultural factors, economic factors, housing factors, spatial and environmental factors (Champion *et al*., 1998). Todaro (1969) and Richard and James (1983) led to a better understanding of the factors which determine the individual′s decision in developing countries. Differences of income, social security, and economic development between urban and rural areas are recognized as the main driving forces of floating and migration of population based on the review of literatures on China in recent years (Cai, 1995; He, 1999; Li, 2003). They outlined the basic scenario of population migration/floating at the provincial scale. Few migration modeling and empirical determinants studies are conducted on smaller scale for non-availability of data. Based on registered individual floating population data from 2005 to 2008, population flocking to Yiwu City is analyzed at the provincial level within China as well as at the county level in Jiangxi and Anhui provinces.

In the last few decades, studies on a variety of social science disciplines have demonstrated the value gained by explicitly acknowledging spatial effects on issues concerning social processes and relevant explanatory statistical models. Spatial statistics and spatial econometrics (Anselin, 1988; 2000) have drawn demographers′ attention recently (Chi and Zhu, 2008; Paul *et al*., 2006). Applied demographers and population geographers are always interested in the spatial dimension of population, especially migration and floating such as the central place theory, gravity model, and origin-destination model. Few of them integrate spatial dimension into migration model for lacking of detailed data or unconsciousness of importance of spatial effect. Considering the spatial autocorrelation of variable data, this study tested the advantages of spatial error model and spatial lag model over ordinary linear regression model.

In this paper, explorative spatial data analysis and map visualization methods were used to extract the origin distribution patterns at the provincial and county levels. Choropleth maps overlaid with multi-ring buffers were novel visualization approaches for evaluation of distance law of migration. This paper describes the migration models with/without migration stock and demonstrates their potential by applying them to explain the floating population phenomena in Yiwu City. The dynamics of population floating to Yiwu City from Anhui and Jiangxi were analyzed by comparing the ordinary least square, spatial lag model and spatial error model. The first empirical analyses of floating population at the county level in this paper further improve our understanding on the floating population phenomena.

2 Data

2.1 Study area

The study area includes origins and destinations of floating population. In this paper, Yiwu City is chosen as the destination place. It is a representative city with rapid urbanization process in eastern coastal areas of China (Lou and Chen, 2007). Moreover, it is typical as an active net-immigration region (Liu *et al*., 2011). The whole China at the provincial level, Anhui and Jiangxi provinces at the country level are the origin places.

Yiwu City is located in the central area of Zhejiang Province, China, 132 km away from Hangzhou City, the capital of Zhejiang Province (Fig. 1). The area of Yiwu is 1100 km^2. The population accumulative effect caused by floating population is one of the main factors of rapid economic development of Yiwu City (Lou and Chen, 2007). A great number of rural labors from different areas are attracted to Yiwu City for expecting higher income. The whole city has a population of 7.239×10^5 registered as household and more than 1.114×10^6 as temporary residents at the end of 2007. The average growth rate of temporary residents is 18.7%, 18 times higher than that of the household register population. The economy of Yiwu has developed fast with a growth rate higher than 10% each year since 2000. In 2008, the whole city realized a domestic gross value of production of 4.933×10^{10} yuan (RMB). Yiwu is already the largest commodities trading center in the world.

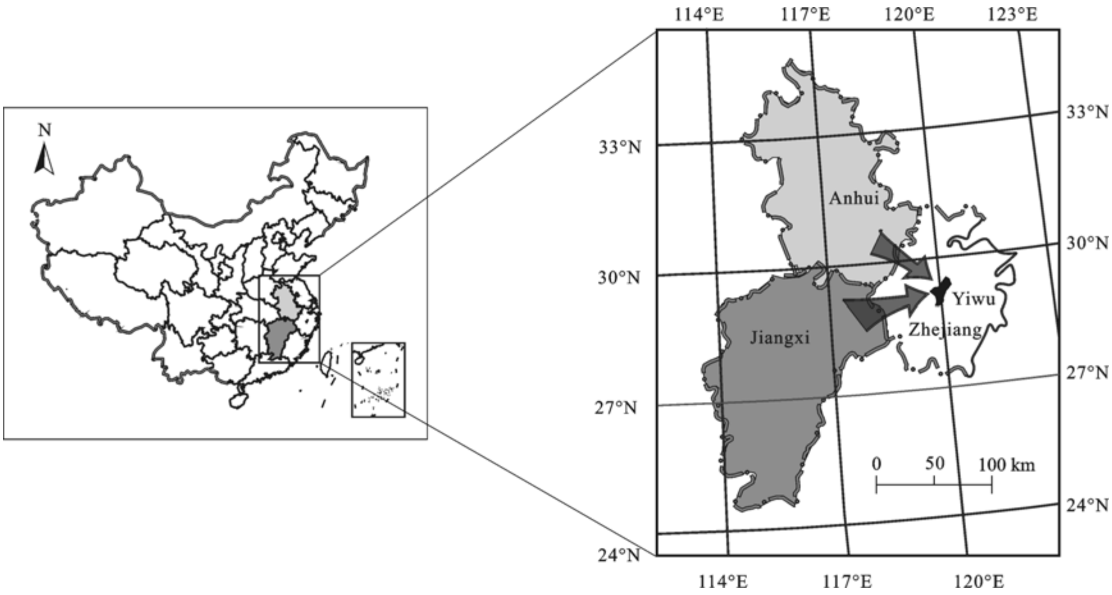


Fig. 1 Location of study area

Both Anhui and Jiangxi, located in the southeastern China, are chosen as the origins. Anhui is bordered by Zhejiang and Jiangsu on the east, Jiangxi on the south, Shandong on the north, Henan and Hubei on the west. The area of Anhui is over 1.3×10^5 km^2, with a total population of 6.6×10^7. Anhui Province is composed of 78 counties and municipal districts under its jurisdiction. Jiangxi lies between Zhejiang, Fujian, Guangdong, Hunan, Hubei and Anhui. Jiangxi Province covers an area of 1.7×10^5 km^2 and has a population of 4.3×10^7, consisting of 91 counties and municipal districts. The proportions of population engaged in agricultural production in Anhui and Jiangxi were 78% and 44% in 2007, respectively. The per capita net income of farmers was 4832 yuan and 4098 yuan respectively, only about half of the farmer income in Yiwu City.

2.2 Data

The registered floating population data at the individual level from 2005 to 2008 are acquired from Yiwu Municipal Public Security Bureau and used with permission. Private information of individuals has been removed from data records. A floating individual′s record contains birth date, age, gender, education, register date, reasons of floating, place name and administrative code of temporary residence, and six digits of identification number indicating their origin places. Data for Hongkong, Macao and Taiwan of China are not available.

The spatial data include the administration division polygons at the provincial level and the county level during the period of fifth census of China in 2000. The geographic code is used to link attribute data complies with the national standards. But part of identity card number indicating origin places is not in common use in organizations outside the public security and need further processing.

Other data like total population, gross domestic product (GDP), the annual per capita net income of rural households, cultivated land resources were obtained from the Anhui Statistical Yearbook (Anhui Statistical Bureau, 2008) and the Jiangxi Statistical Yearbook (Jiangxi Statistical Bureau, 2008).

3 Methods

3.1 Data preprocessing

The major challenge for data preprocessing is the differences of geographic code standards. Geographic codes are short alphabetic or numeric codes developed to represent provinces, prefectures, counties and subdivisions (e.g., townships or villages) for statistical and other purposes. There are many internal geographic code standards in different organizations like the statistics department and the public security department in China because of the historical reasons. The spatial polygon data used in this study come from the fifth census. The geographic codes are regulated by statistics department and used widely as the standard administrative division code. The floating population data come from the public security department. The geographic codes are inconsistent in many counties.

Another challenge is the frequent updates of administrative division and codes. Some administrative divisions are merged into a new one. For example, the Ninggang County and old Jinggangshan City are merged into a new Jinggangshan City in Jiangxi Province. Meanwhile, some divisions are divided into several subdivisions. One example is that the old Ji′an City is divided into Jizhou and Qingyuan districts.

Floating population data preprocessing involves four steps to set up the analysis datasets. To begin with, aggregate the individual data with geographic region code in identification card number at different levels. The aggregation need two, four and six digits of geographic region code at the provincial, city, and county levels, respectively, that at the county level is most difficult. Then, join the aggregated floating population data with the administrative division polygon data by county name. Furthermore, check the record which does not match and match them according to region name, geographic code and relative records on the adjustment of administrative divisions as much as possible. Much of unmatched cases are subject to add, delete or change administrative division. Finally, if there are still unmatched regions to any administrative division, distribute the floating population numbers to each county in average in each province.

ArcGIS software (ESRI) and Microsoft Excel have been used to build the datasets. Statistic data from the statistical yearbook, floating population data from different counties to Yiwu City are appended to the attribute tables of the administration division polygon data for further analysis.

3.2 Origin distribution analysis

After acquiring the spatially enabled floating population dataset, the explorative spatial data analysis and map visualization method have been used for distribution pattern analysis of floating population, so far two types of analysis have been made, the first is floating population from different provinces in China, the other is floating population from counties in Jiangxi and Anhui provinces into Yiwu City.

Spatial modeling often begins with explorative spatial data analysis (ESDA). Generally speaking, an ESDA is composed of visualizing spatial patterns exhibited in the data, identifying spatial clusters and spatial outliers, and diagnosing possible misspecification of spatial aspects of the statistical models, all of which can help better specify regression models (Anselin, 1996; Baller *et al.*, 2001). The core of ESDA techniques are measures of spatial autocorrelation, which can be defined as a similarity measure between two values of an attribute that are nearby spatially. The most well-known index of spatial autocorrelation is Moran′s I statistic (Moran, 1948).

The Moran′s I statistic of floating population at the provincial level has been calculated by using Geoda software (http://geodacenter.asu.edu/software/downloads). The Moran′s I statistic of migration ratio at the county level has been calculated and compared by using different spatial weight matrices for next spatial regression modeling in Jiangxi and Anhui provinces. Spatial clusters and spatial outliers have been explored by local indicators of spatial association (LISA) method to indicate the presence or absence of significant spatial clusters or outliers for each location.

Choropleth map is a type of map visualization method in which areas are colored or shaded to reflect the value of the mapped phenomenon or to display classes of values. Effective classification schema should be chosen to represent the distribution of data values and show the spatial pattern within the study area. Choropleth map of migration ratio and multi-ring buffer from Yiwu City in Anhui and Jiangxi have been overlaid to expose the floating patterns in these two provinces.

3.3 Spatial regression modeling for determinant analysis of population floating

The determinant analysis tries to reveal what is behind the floating phenomena that people from different origins to Yiwu City. According to Todaro (1969), the migration model is specified as follows.

$$M_{ij} / P_i = f(D_{ij}, X_i, X_j) \qquad (1)$$

where M_{ij} refers to gross flow of migrants over a given period from one area i to another area j; P_i refers to the population of origin places at the beginning of the period; M_{ij}/P_i refers to migration ratio which most of the time is used as the dependent variable in the model. There are three types of variables in the process of population floating and migration, i.e., 1) social-economic factors of origin places X_i like the population density, amount of cultivated land per person, and education; 2) comparative advantage factors of destination over origin X_j like average wage ratio, employment rate ratio and employment chance ratio; 3) obstacle factors D_{ij} from origin place i to destination place such as Euclidean distance, degree of transportation convenience, clan, geographical relationships, *etc*.

3.3.1 Two migration models

In order to understand the dynamics of floating population to Yiwu City, key variables need to be specified in the model. Many empirical results implied that the selecting appropriate variables are crucial for migration modeling, because variable selection must be based on the specific county or areal contexts on the one hand, and also dependent upon the data availability on the other hand (He, 1999). What′s more, many variables may be highly correlated, which is known as multicollinearity. For example, the per capita gross domestic product is usually correlated significantly with the annual per capita net income. Generally speaking, population floating is a complex social-economic phenomenon and involves population variables like population density, rural labor force, education, *etc*., social variables like floating experience, clan, fellow-townsman relationship, and physical variables like distance, and other geographic variables like altitude, slope, land quality, soil fertility, cultivation index, *etc*.

According to previous study and current conditions, several variables like man-land ratio, per capita net income of rural labor forces, distance to Yiwu City, and migration stock have been chosen for modeling. The main reasons are given as follows. To start with, as there are only floating data of one destination from different origins in this study, the social-economic ′pull′ factors of Yiwu City is the same in the model. They may be reflected in the constant term of the function. Next, Substantial achievements and empirical researches in this field of migration of population and labor forces have illustrated that levels of economic development is the key variable that explains the floating direction and volume of labor forces (He, 1999; Yang *et al*., 1999; Wang and Fan, 2009). In the mainstream model on the floating of labor forces, the variable of income often is adopted. Because most of floating people are rural labor forces, the annual per capita net income of rural households (yuan) is used to reflect the income of different origins. In addition, the opportunity cost of labor withdrawn from agriculture in the origin place depends most directly on the quantity and quality of land available per agricultural worker. In this study, man-land ratio is defined as cultivated land resources (ha) divided by the total population (10 000 persons), which should be replaced by rural labor forces but data are missing at the year-end. Fourthly, distance to Yiwu City should be adequately reflected in the model. It may be also proxy for non-economic costs of floating. Fifthly, the clan contacts variables may play an additional role in helping migrants directly to find work in destination places. But this variable is very hard to be quantified. Migrant stock (defined as number of migrants who moved previously to a destination) may be used as alternative of clan contacts or other cultural factors. The volume of would-be migrants is an increasing function of the stock of previous origin-to-destination migrants (He, 1999). Other previous studies indicate that total annual investment or foreign direct investment is important factor of population floating (He, 1999; Wang and Fan, 2009), which are not included here.

The above considerations lead to the following two migration models with and without migration stock in double-log form.

$$\ln MR = \alpha_1 \ln MLR + \beta_1 \ln PCNIR + \lambda_1 \ln DYW + \theta_1 \quad (2)$$

$$\ln MR = \alpha_2 \ln MLR + \beta_2 \ln PCNIR + \lambda_2 \ln DYW + \varepsilon_2 \ln MS + \theta_2 \quad (3)$$

where ln*MLR* is man-land ratio; ln*PCNIR* refers to per capita net income of rural labor forces; ln*DYW* represents distance to Yiwu City in Euclidean space; ln*MS* refers to migration stock in previous year; ln*MR* refers to migration ratio from different places to Yiwu City and is used as dependent variable in model.

3.3.2 Regression analysis

Standard linear regression and two popular spatial regression models including the spatial error model and the spatial lag model (Anselin, 1988; 2000) have been used to analyze the determining factors behind the floating phenomena that people from different origins to Yiwu City. A standard linear regression assumes that the error terms are independently, identically, and normally distributed. But the independence assumption of the errors is often violated due to spatial autocorrelation. Spatial linear regression model is a generalization of standard linear regression models in which spatial autocorrelation is allowed and accounted for explicitly by spatial models. The two migration models and three regression methods have been compared in two provinces (Jiangxi and Anhui) at the county level.

Spatial weight matrix is the foundation of all the spatial analysis, e.g., Moran′s I test and spatial regression analysis. A spatial weight matrix should be specified corresponding to the neighborhood structure so that the resulting variance-covariance matrix can be expressed as a function of a small number of estimate parameters relative to the sample size (Anselin, 2002). There are many spatial weight matrices like rook's and queen′s contiguity weight matrix of order one or higher, the k-nearest neighbor weight matrix, the general distance weight matrix, *etc*. There are no practical guidelines for the selection of neighborhood structure except for the theoretical analysis in Arthur (2010). Spatial weight matrix with high coefficient of spatial autocorrelation along with a high level statistical significance of specified variable has been used after calculation and comparison as in Chi and Zhu (2008).

Data-driven approach to model assessment and comparison has been used in this study to determine which model is the better one accounting for the spatial autocorrelation. Akaike′s Information Criterion (AIC) and Schwartz′s Bayesian Information Criterion (BIC) are often used, which measure the fit of the model to the data but penalize models that are overly complex. Models with a smaller AIC or BIC are considered as the better ones in the sense of model fitting balanced with model parsimony.

4 Results and Analyses

The aggregated individual data have been matched and joined with administrative division polygon data by using the data preprocessing methods described above. Floating population size is acquired from Anhui and Jiangxi provinces to Yiwu City at the county level.

4.1 Origin distribution patterns at provincial level

As a strong economic city, the ′pull′ forces of Yiwu City result in different effect on different people in different places. Jiangxi Province ranks first in all of 31 provincial-level administrative divisions in average from 2005 to 2008. Followings are Anhui Province, Henan Province and Guizhou Province which are smaller than Jiangxi Province. Although Yiwu is located in Zhejiang Province, the floating population from Zhejiang to Yiwu City ranks fifth. Hunan Province is located near to Yiwu City, but the floating population is smaller than that from Guizhou Province. In Sichuan Province, over 8.0×10^6 workers are employed outside the province and rank the first each year. But when it goes to the floating population to Yiwu City, Sichuan Province is only the seventh largest province. They are abstracted towards other economic center.

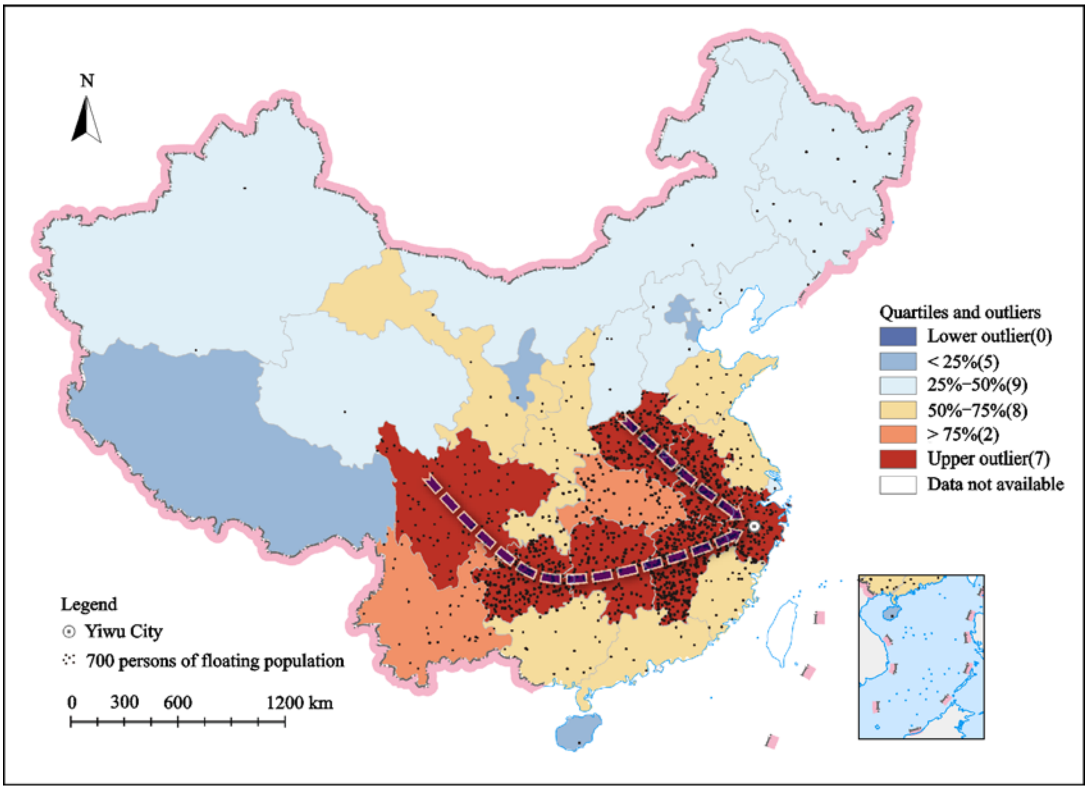
Explorative spatial data analysis and map visualization methods have been used to get the distribution pattern map of floating population to Yiwu City (Fig. 2). Thus the figure is overlaid by two representative methods using aggregated temporary registered residence data at the provincial level in 2007. One of them is the dot density map created by ESRI ArcMap software. The other is the box map created by Geoda software.

Explorative spatial data analysis can be of great help in understanding the origin distribution pattern. Box map at a hinge level of 3.0 indicates the outliers are Jiangxi, Henan, Anhui, Guizhou, Zhejiang provinces. Box map at a hinge level of 1.5 indicates more outliers like Hunan and Sichuan provinces. The global Moran′s I statistic of floating population to Yiwu City from different provinces in 2007 is 0.2379 with *p*-value smaller than 0.01 (spatial weight matrix used here is rook′s contiguity with order 1). This proved the existence of positive spatial autocorrelation. LISA analysis allows us to know which high values of floating population areas are surrounded by high values in the neighboring areas, i.e. which is called the high-high clusters. There are clearly spatial clusters in the floating population distribution pattern at the provincial level according to LISA analysis results (Table 1). Provinces classified as high-high clusters are Jiangxi, Anhui, Zhejiang, Hunan, and Hubei provinces. Conversely, Inner Mongolia and Xinjiang

Table 1 LISA cluster and local Moran′s statistic of floating population to Yiwu City

LISA cluster	Province	Local Moran′s I	Population size	*p*-value
High-high	Jiangxi	2.064045	232957	0.020
	Anhui	1.964500	136549	0.002
	Zhejiang	1.012428	96500	0.032
	Hunan	0.702096	81178	0.030
	Hubei	0.312792	45248	0.002
Low-high	Fujian	–0.432964	14994	0.022

Notes: Province whose floating population size smaller than 10 000 is not listed. Rook′s contiguity with order 1 is used



Number in bracket indicates number of provinces in quartile

Fig. 2 Distribution pattern map of floating population to Yiwu City

regions are classified as low-low clusters. This demonstrates that population floating, as a complex social-economic phenomenon, is not a stochastic process. Two provinces in low-high class are Fujian and Hainan. There is no any province in high-low class. Others are not significant statistically. The box plot at a hinge level of 3.0 of local Moran's I statistic indicates the Jiangxi and Anhui provinces are outliers in high-high spatial clusters.

The spatial variation of origin distribution of floating population is illustrated clearly on the dot density map in Fig. 2. Dot density maps show floating population density graphically rather than showing density value. Each dot represents 700 floating people to Yiwu City. Compared to the other provinces, the dots of the above outliers are much closer. The most densely-populated areas are Anhui and Zhejiang provinces.

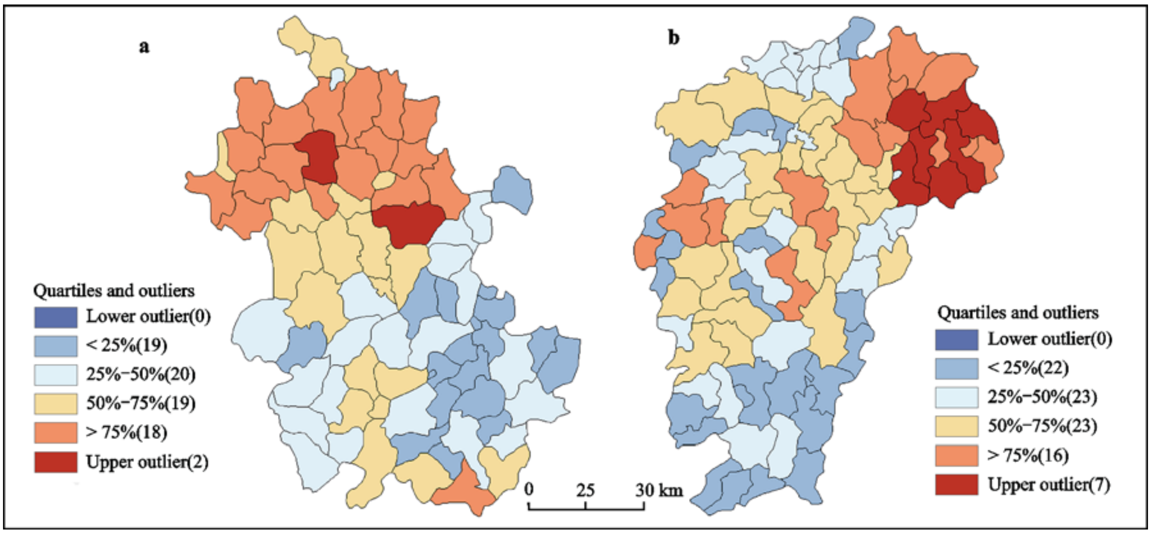
The most significant characteristic is the two axes or pathways on Fig. 2. One axis extends from Henan to Yiwu City via Anhui and Zhejiang. Another axis winds its way from Sichuan to Yiwu City via Guizhou, Hunan, Jiangxi, and Zhejiang. These two axes connect the above detected seven outliers and contribute more than 80% of floating populations to Yiwu City. The sum of seven upper outliers in all the 31 provinces (autonomous regions, municipalities) achieves 83%. The distribution follows the 80/20 principle or the Pareto Principle: about 80% floating population to Yiwu City come from the 20% of the provinces, while about 20% (17.05%) floating population come from other 80% provinces. The overall spatial layout seems a sideling U-shape. Taking into account the specificity of the Sichuan Province with the biggest population and the biggest floating population in China, the spatial layout of Yiwu's pull is a V-shaped pattern excluding Sichuan Province.

4.2 Origin distribution patterns at county level

In order to better understand the phenomenon of population floating to Yiwu City from different origins at the county level, Anhui and Jiangxi provinces are further analyzed for several reasons. To begin with, they are the top two provinces of floating population to Yiwu City. The ratio of floating population from these two provinces is up to 35.7% in average from 2005 to 2008. Furthermore, they are outliers by using local Moran's I statistics in high-high spatial clusters. Finally and most intuitively, they are adjacent provinces to Zhejiang which Yiwu City belongs to.

4.2.1 Explorative spatial data analysis results of floating population

Box maps of floating population of two provinces in 2007 are plotted to show the spatial distribution of outliers (Fig. 3). Box map at a hinge level of 3.0 indicates the outliers in Anhui are Dingyuan and Mengcheng counties. Box map at a hinge level of 1.5 indicates more outliers like Funan, Bozhou, Linquan, and Guzhen counties. Box map at a hinge level of 3.0 indicates the outliers in Jiangxi are Geyang, Guixi, Shangrao, Leping, Qianshan, Dexing and Yushan counties. Box map at a hinge level of 1.5 indicates more five outliers. There are five counties in Jiangxi Province while there are only one county with floating population to Yiwu City larger than 10 000. In the detected outliers at a hinge level of 1.5, there are two impoverished counties in Anhui while three in Jiangxi. The floating population size of outliers is larger in Jiangxi than that in Anhui. Moreover, the outliers are more spatially-concentrated in Jiangxi than that in Anhui.



Number in brackets indicates number of counties in quartile

Fig. 3 Box map of floating population to Yiwu City in Anhui (a) and Jiangxi (b) provinces

The Moran scatters plot as well as the LISA analyses have been done in Anhui and Jiangxi (Fig. 4). The global Moran's I statistic of floating population to Yiwu City from different counties in Anhui is smaller than that in Jiangxi in 2007. The spatial weight matrixes used here are chosen according to Table 2. This means that there exists positive spatial autocorrelation and the spatial autocorrelation is stronger in Jiangxi than in Anhui. There are clear spatial clusters in the floating population distribution pattern at the county level according to LISA analysis. Polarization phenomenon is obvious both in Anhui and Jiangxi. There are many counties in high-high cluster and low-low cluster while there are few counties in low-high cluster and high-low cluster except counties are not significant. Woyang County with 5080 floating people in Anhui is categorized as high-high cluster and significant with *p*-value smaller than 0.01. Notably, the local Moran's I statistics of three counties Hengfeng, Qianshan, and Geyang with more than 6000 floating people to Yiwu City in Jiangxi are 3.79, 6.16, 13.2, respectively. Geyang County can be seen as a kernel with largest floating population size in the southeast part of Jiangxi.

4.2.2 Migration ratio analysis using Choropleth map with multi-ring buffer

Map visualization methods give another perspective on floating population to Yiwu City. Migration ratio used here is defined as size of floating population per 10 000 persons. The migration ratio of Jiangxi and Anhui provinces is 45.1 and 20.5 respectively in 2007. The highest migration ratio (803) in Geyang is more than one thousand times the lowest one (0.77) in Xunwu in Jiangxi Province. The range of migration ratio in Jiangxi is two times larger than the range in Anhui. The rate of maximum and minimum migration ratio in Jiangxi is about three times the rate in Anhui. This shows the degree of unevenness of migration ratio in Jiangxi is larger. There are 11 counties where migration ratio is greater than 100 in Jiangxi Province while there are only two counties in Anhui Province. The Dingyuan County, rated number one in Anhui, is the eighth highest one if it belongs to Jiangxi Province.

Choropleth maps have been created to show the differences of migration ratio in the two provinces (Fig. 5). A multi-ring buffer around Yiwu City has been created using the multi-ring buffer tool in ArcToolbox and overlaid on the choropleth maps. Each ring is 50 km width. As the distance to Yiwu City increase, the migration ratio is supposed to decline according to the distance rule of migration. There are two counties whose migration ratio is larger than 100 in Anhui. One is Xiuning County, close to Jiangxi and Yiwu. The other is Dingyuan County, located far away from Yiwu City. Both of these two counties in Anhui start the declining pattern of migration ratio as the distance increases in the southwest and northwest parts, away from Yiwu City. The distance effect is not true in other areas. However, the result in Jiangxi is the same to above expectation. The migration ratio in Jiangxi Province declines as the distance to Yiwu City increases roughly.

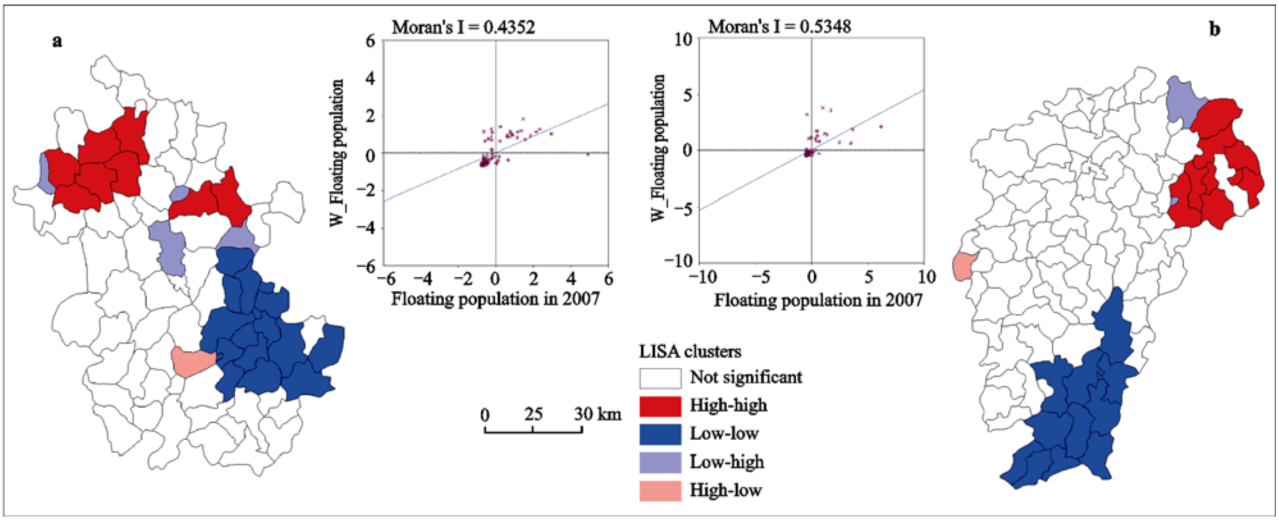
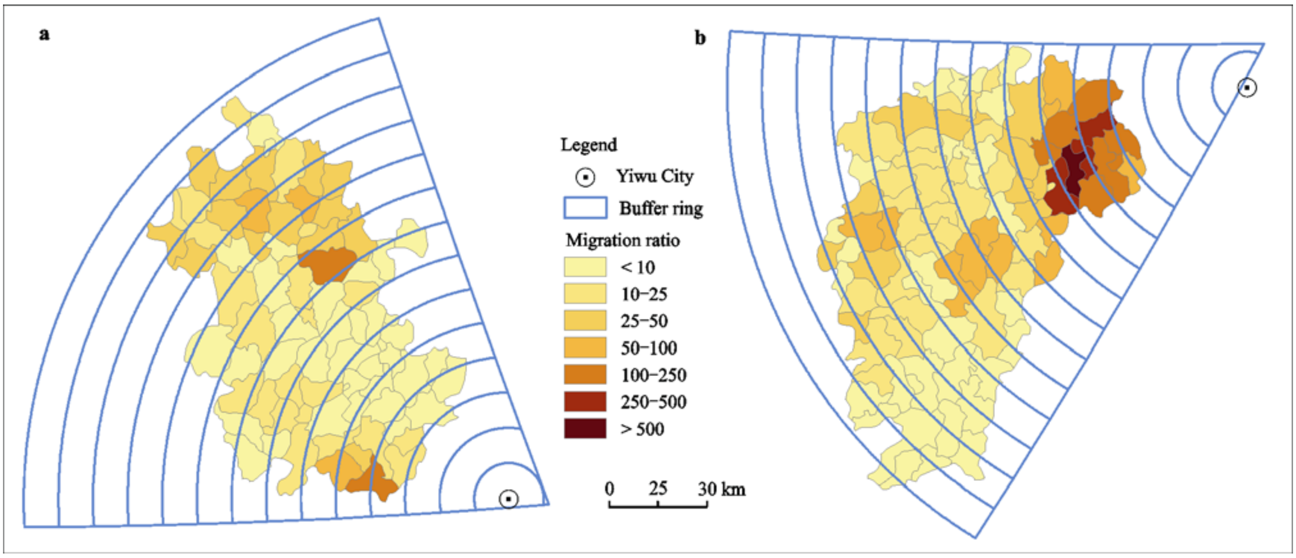


Fig. 4 LISA analysis on floating population to Yiwu City in Anhui (a) and Jiangxi (b) provinces



Interval of multi-ring buffer is 50 km

Fig. 5 Choropleth map of migration ratio in Anhui (a) and Jiangxi (b) provinces

4.3 Determinant analysis

The choice of spatial weight matrix is very important and the one with largest Moran′s I statistic is selected. Spatial autocorrelation of migration ratio (logarithm version) of floating population from different counties of these two provinces can be expressed by global Moran′s I statistic. After repeatedly calculating using queen, rook, general distance, and nearest neighbor methods and different parameters, the 3-nearest neighbor weight matrix is selected for high coefficient (Moran′s I = 0.67; p-value < 0.01) for analysis in Jiangxi Province, while the general distance weight matrix (Moran′s I = 0.57; p-value < 0.01) is selected for analysis in Anhui Province.

The gathering time of variables should be clearly stated in modeling analysis. Distance to Yiwu City is almost constant if the administrative division is not adjusted. There are two variables (man-land ratio and per capita net income) whose gathering time is 2007. Migration stock data are a lagging indicator on population floating and refers to that in 2006.

4.3.1 Model fitting results of floating population from Jiangxi

The results from multiple regression analysis of migration ratio in Jiangxi Province are shown in Table 2. The Moran′s I test on the residuals after fitting the ordinary least square (OLS) suggests that there is strong evidence of spatial autocorrelation among the residuals. Thus the independence assumption of the error term appears to be violated. In order to explain the spatial autocorrelation, the spatial linear regression models were applied. The 3-nearest neighbor weight matrix is selected because of high coefficient (Moran′s I = 0.67; p-value < 0.01) for spatial regression analysis.

The migration ratio has correlation with the distance to Yiwu City in all the three estimates of model 2. Compared with the standard linear regression residuals, the residuals from the spatial linear regression were much reduced in size. The dramatic reduction of the Moran′s I statistic from 0.352 to 0.040 shows that spatial autocorrelation among these residuals was eliminated. The standard linear regression indicates the correlations among migration ratio and man-land ratio and rural labor income are not significant, while the correlation between migration ratio and man-land ratio is significant at the 0.5 level according to spatial version of model 2. The three estimates results of model 2 shows strong positive correlations between migration ratio and the distance variable at the 0.0001 level. The migration ratio in near counties is larger than that in counties far away from Yiwu City in Jiangxi Province which is shown in Fig. 5b. The spatial error model (SEM) has slightly higher goodness of fitting than spatial lag model (SLM) based on AIC value even though the Moran′s I value of residuals is larger in SEM.

The correlation between migration ratio and migration stock is notable with high significance level (p-value < 0.0001) in OLS, SLM and SEM of model 3.

Table 2 Regression results of models for floating population to Yiwu City from counties in Jiangxi Province

Independent variables	Model 2			Model 3		
	OLS	SLM	SEM	OLS	SLM	SEM
lnMLR07	–0.078	–0.136*	–0.148*	–0.096*	–0.134*	–0.084*
lnPCNIR07	–0.072	0.025	0.058	0.026	0.084	0.019
lnDYW_E	–3.003*****	–1.772*****	–2.860*****	–0.485**	0.147	–0.628
lnMS06	–	–	–	0.805*****	0.737*****	0.751*****
Constants	22.226*****	13.158 **	20.740*****	0.849	–3.575*	2.052*
Spatial lag effects	–	0.404*****	–	–	0.277*****	–
Spatial error effects	–	–	0.403*****	–	–	0.311***
Log likelihood	–132.300	–124.167	–124.398	–76.475	–69.248	–72.651
AIC	272.600	258.334	256.797	162.950	150.497	155.302
Moran′s I for residuals	0.352***	0.040*	0.043*	0.226***	–0.002	0.039*

Notes: *, **, ***, ****, ***** represent significance at 0.5, 0.1, 0.01, 0.001 and 0.0001 respectively; InMLR07 is man-land ratio in 2007; lnPCNIR07 represents per capita net income of rural labor forces in 2007; lnDYW_E represents distance to Yiwu City in Euclidean space; lnMS06 refers to migration stock in 2006; OLS is ordinary least square regression method; SLM is spatial regression using spatial lag model; SEM is spatial regression using spatial error model; AIC means Akaike′s Information Criterion

The OLS results of model 3 also confirm the distance-decay effect while results from the spatial version of model 3 failed to support the distance law in population floating. There is negative correlation between migration ratio and man-land ratio from the three fitting results at the 0.5 level. Results from model 3 show that there was no correlation between migration ratio and the per capita net income of rural labor forces. The results from model 2 also show the same results. The OLS fitting results shows the negative correlation between migration ratio and distance while the SLM and SEM fitting results failed to support this. The SLM estimates have higher goodness of fitting than SEM estimates in model 3 no matter from the AIC values or from the Moran′s I value of residuals.

4.3.2 Model fitting results of floating population from Anhui

The results from multiple regression analysis of migration ratio in Anhui Province are shown in Table 3. The assumption of independence of the residuals is violated because of strong evidence of spatial autocorrelation among the residuals in model 2 and 3. In order to account for the spatial autocorrelation, spatial linear regression models like SLM and SEM are fitted. The general distance (50 km) weight matrix is selected for weight matrix calculation.

Compared with the regression residuals by ordinary least square method, the residuals from the spatial lag model and spatial error model in model 2 were much reduced in size, and spatial autocorrelation among these residuals was essentially eliminated. Fitting results of model 2 show that the migration ratio has notable positive correlation with the man-land ratio in all the three estimates with high significance level. The correlation between migration ratio and per capita net income of rural labor forces is also significant at different levels among OLS, SLM and SEM fitting. The distance-decay effect is only confirmed by the best fitting results using SLM in model 2. According to the likelihood statistics, the AIC value and the Moran′s I statistics of residuals, the fitting results in model 2 show a tight race between spatial lag model and spatial error model.

The improvement of spatial autocorrelation among these residuals was obvious by spatial regression method in model 3 and the Moran′s I statistics were reduced from 0.255 to 0.015 and –0.019, respectively. The migration ratio was positively significantly correlated with the man-land ratio at the 0.1 level. There is negatively correlation between migration ratio and the per capita net income of rural labor forces at the 0.1 significance level. Meanwhile, the migration ratio has notable negatively correlation with distance variables. Strong positive correlations exist between migration ratio and migration stock (p-value < 0.0001). According to the AIC value, the spatial lag model provides better fitting than spatial error model while spatial error model is better than OLS.

5 Discussion

Existing statistical system of floating population is not enough to support analysis work of origin, destination and flow direction of population. The registered floating population data used in the study are the best available data. Still there is no guarantee for ′registration when coming′ and ′deregistration when leaving′ on one hand.

Table 3 Regression results of models for floating population to Yiwu City from counties in Anhui Province

Independent variables	Model 2			Model 3		
	OLS	SLM	SEM	OLS	SLM	SEM
lnMLR07	0.571***	0.480****	0.469****	0.185**	0.205**	0.210**
lnPCNIR07	–1.465***	–0.803**	–0.685*	–0.527**	–0.349**	–0.463**
lnDYW_E	0.070	–0.257*	0.014	–1.179*****	–1.131*****	–0.936****
lnMS06	–	–	–	0.807*****	0.668*****	0.718*****
Constants	10.501*	5.914*	5.056*	7.240***	5.454**	5.723**
Spatial lag effects	–	0.667*****	–	–	0.341*****	–
Spatial error effects	–	–	0.706*****	–	–	0.484*****
Log likelihood	–100.930	–82.792	–84.212	–46.048	–36.459	–40.665
AIC	209.861	175.584	176.425	102.097	84.918	91.329
Moran′s I for residuals	0.457***	–0.017*	–0.019	0.255***	0.015*	–0.019

On the other hand, the difference of geocoding schema between different department and the frequent updates of administrative divisions are main challenges for spatially-enabled data preprocessing.

Map visualization method as well as explorative spatial data analysis can do great help in detecting the spatial pattern of floating population to/from a certain place. Two provinces, i.e. Jiangxi and Anhui as origin places, are found to be outliers and typical for further analysis on determinant factors. Explorative spatial data analysis using box map at the county level illustrates the concentration or decentralization trends of floating population. Migration ratio map with multiple ring buffers from origin to destination is proved as an excellent media for analyzing the relationship between migration ratio and distance.

Some recommendations for floating population model selection can be acquired based on fitting results. The model fitting results confirm that the theoretical deduction of the spatial linear regression can reduce the residuals and eliminate the spatial autocorrelation among the residuals. The estimate results in the two spatial models in Jiangxi and Anhui provinces further suggest a preference for the spatial lag model rather than the spatial error model. Thus the spatial lag model can be viewed as an appropriate alternative to standard regression approaches in floating population models. Based on the likelihood statistics, the AIC and the Moran′I statistics of residuals, the model 3 provide an improved fitting over the model 2. Model 2 is used to explain the decision to migrate when the migration stock is low at the initial stage or there is little knowledge for floating population about their friends, relatives or hometown fellows in Yiwu City. Model 3 introduces migration stock variables and will explain much more than model 2. But one risk of using model 3 is that migration ratio maybe closely relate to the migration stock. The migration stock variables may appear to exert a strongly positive effect on floating and migration even if in fact there is no such relationship.

Floating population from Jiangxi and Anhui ranks the top two in all of 31 provinces. The Yiwu City is more attractive to people in Jiangxi Province than those in Anhui Province. However, the determinant factors seem different according to the best fitting results using SLM in model 3. In the case of Jiangxi Province, the SLM fitting results show that the migration ratio is significantly positively correlated with migration stock at the 0.0001 level. The migration ratio is negatively significantly correlated with man-land ratio at the 0.5 level. In the case of Anhui Province, the SLM fitting results show that the migration ratio is correlated with migration stock positively (p-value < 0.0001), distance to Yiwu City negatively (p-value < 0.0001), the per capita net income of rural labor forces negatively (p-value < 0.1) and the man-land ratio positively (p-value < 0.1). The migration ratio of floating population in Jiangxi flocking into Yiwu City is determined mostly by the migration stock. The distance law of migration failed to explain the migration ratio because of the impact of large migration stock. However the migration ratio map with multiple buffer rings in Fig. 5 shows the obvious distance-decay effect in Jiangxi Province. It appears that the dynamic population floating process from counties in Jiangxi to Yiwu City is in its developed stage. Subsequent would-be floating population is influenced by migrants who moved previously to Yiwu City greatly. However, the population floating process from counties in Anhui to Yiwu City is in its developing stage. Thus the model fitting results show that the determinant factors are composed by several social-economic and obstacle factors. Meanwhile, the pull forces from other industry centers located in Jiangsu, Shanghai and other provinces (municipalities) restrict the development of population floating process from Anhui to Yiwu City. The correlation is negative in Jiangxi Province between migration ratio and man-land ratio. This is in line with the theoretical assumption and people′s understanding. However, the model fitting results in model 2 and 3 show the positive correlation of these two variables in Anhui Province. This apparent paradox needs further analysis in future research.

Theoretically speaking, social-economic factors of origins and destinations and obstacle factors from origin places to destination places are related to population floating and migration. Ravenstein's (1885) distance-decay function in migration is confirmed according to the best estimates of model 2 in Jiangxi/Anhui Province and the model 3 in Anhui Province. However, the best fitting results by SLM version of model 3 in Jiangxi Province do not support the distance law of migration. However, the fitting results of model 2 and spatial pattern by map visualization illustrate obvious distance-decay effect. The introduction of migration stock which has strong ability to explain the floating population phenomena submerges the reality of the distance law. The correlation between per capita net income of rural labor forces and migration ratio is not significant according to the three versions of the two models in Jiangxi. These results indicate obvious difference with previous studies at the provincial level (Cai, 1995; He, 1999; Li, 2003; Wang and Fan, 2009). The range of the income in log-form within the counties in Jiangxi is as small as 1.13. The main reason is the small disparity of per capita net income of rural labor forces. Multivariate spatial interaction approach (He, 1999) suggest that migration stock is an important variable to reflect both migration policy change and economic development strategy. This study also demonstrates that migration stock has strong ability to explain floating population phenomena at the county level.

6 Conclusions

In this paper, the origin distribution pattern of floating population in Yiwu City is analyzed using explorative spatial data analysis and map visualization methods. Using the migration ratio in 2007 as an explanatory variable, two models are fitted using OLS, SEM and SLM approaches with county-level data in Jiangxi and Anhui provinces. The determinant factors are analyzed based on model comparison in the two provinces and with previous studies.

To better understand population floating phenomenon, data on floating population size, structure (in terms of sex, age, ethnic, occupation, *etc*.) and flow direction are essential and should be collected, processed, released periodically. ESDA and map visualization methods have great potential in spatial analysis of floating population data. The V-shaped origin distribution pattern at the provincial level and the distance-decay effect in Jiangxi are detected. In choosing modeling methods among OLS, SLM and SEM, empirical study in this paper suggests replacing OLS with spatial linear regression to reduce the residuals and eliminate the spatial autocorrelation among the residuals. And it further suggests a preference for the spatial lag model rather than the spatial error model. The model 3 with migration stock variable provides an improved fitting over the model 2. The floating population flocking into Yiwu City from Jiangxi is determined mostly by the migration stock while the determinant factors are migration stock and the distance to Yiwu City in Anhui. The correlation between per capita net income of rural labor forces and migration ratio is not significant in Jiangxi and significant only at the 0.1 level in Anhui. The distance-decay effect is true for migration flow from Anhui to Yiwu City while the distance rule is not confirmed in Jiangxi with the best fitting model. The dynamic population floating process from counties in Jiangxi to Yiwu City is in its developed stage. Meanwhile, the population floating process from counties in Anhui to Yiwu City is in its developing stage. According to the analysis in this paper, the distance, income and man-land ratio are important factors to explain population floating at earlier stage. However, as the dynamic population floating process evolves, the determinant factor will be the migration stock. The accumulative effect of migration stock depends mostly on social network which is built based on clan, hometown fellow, schoolfellow, *etc*. These findings support the theoretical and practical importance of further social network research of rural labor forces.

References

Anhui Statistical Bureau, 2008. *Anhui Statistical Yearbook*. Beijing: China Statistics Press.

Anselin L, 1988. *Spatial Econometrics*: *Methods and Models*. Boston: Kluwer Academic.

Anselin L, 2000. Part 2 The Link between GIS and spatial analysis: GIS, spatial econometrics and social science research. *Journal of Geographical Systems*, 2: 11–15. doi: 10.1007/s101090050023

Anselin L, 2002. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3): 247–267. doi: 10.1016/S0169-5150(02)00077-4

Anselin L, 1996. The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fischer M (ed.). *Spatial Analytical Perspectives on GIS*. London: Taylor & Frances, 111–125.

Arthur G, 2010. Spatial autocorrelation. In: Manfred M F and Arthur Getis (eds.). *Handbook of Applied Spatial Analysis*. Berlin Heidelberg: Springer-Verlag, 255–278.

Bogue D J, 1969. *Principles of Demography*. New York: Johnson Wiley and Sons, 753–754.

Baller R D, Anselin L, Messner S F *et al*., 2001. Structural covariates of US county homicide rates: Incorporating spatial effects. *Criminology*, 39(3): 561–590. doi: 10.1111/j.1745-9125.2001.tb00933.x

Cai Fang, 1995. Reasons, trends and policies of population migration and floating. *Population Science of China*, 51: 8–16. (in

Chinese)

Champion A, Fotheringham S, Rees P *et al*., 1998. *Determinants of Migration Flows in England*: *A Review of Existing Data and Evidence*. Newcastle: Department of Geography, University of Newcastle.

Chi G Q, Zhu J, 2008. Spatial regression models for demographic analysis. *Population Research Policy Review*, 27(1): 17–42. doi: 10.1007/s11113-007-9051-8

Fotheringham A S, O′Kelly M E, 1989. *Spatial Interaction Models*: *Formulations and Applications*. Dordrecht: Kluwer.

Haynes K E, Fotheringham A S, 1984. *Gravity and Spatial Interaction Models*. Beverly Hills, California: Sage.

He J S, 1999. *Multivariate Spatial Interaction Models as Applied to China's Inter-provincial Migration 1982–1990*. Saskatoon: University of Saskatchewan.

Jiangxi Statistical Bureau, 2008. *Jiangxi Statistical Yearbook*. Beijing: China Statistics Press. (in Chinese)

Lee S E, 1966. A theory of migration. *Demography*, 3(1): 47–57. doi: 10.2307/2060063

Li Qiang, 2003. An analysis of push and pull factors in the migration of rural workers in China. *Social Sciences in China*, 1: 125–136. (in Chinese)

Liaw K L, 1990. Joint effects of personal factors and ecological variables on the interprovincial migration pattern of young adults in Canada. *Geographical Analysis*, 22: 189–208. doi: 10.1111/j.1538-4632.1990.tb00205.x

Liaw K L, Ledent L, 1987. Nested logit models and maximum quasi-likelihood method. *Regional Science and Urban Economics*, 17(1): 67–88. doi: 10.1016/0166-0462(87)90069-X

Liu S H, Hu Z, Deng Y, Wang Y J, 2011. The regional types of China′s floating population: Identification methods and spatial patterns. *Journal of Geographical Sciences*, 21(1): 35–48. doi: 10.1007/s11442-011-0827-8

Lou Honghao, Chen Xiuying, 2007. Study on urbanization of Yiwu City. *Economy Geography*, 27(5): 790–795. (in Chinese)

Manski C, 1977. The structure of random utility models. *Theory and Decision*, 8(3): 229–254. doi: 10.1007/BF00133443

Moran P, 1948. The interpolation of statistical maps. *Journal of the Royal Statistical Society B*, 10: 243–251.

Paul R voss, Katherine J C W, Roger B H, 2006. Explorations in spatial demongraphy. In: William A K (eds.). *Population Change and Rural Society*. Netherlands: Springer, 407–429.

Ravenstein E G, 1885. The laws of migration. *Journal of Statistical Society of London*, 48(2): 167–235. doi: 10.2307/2979181

Richard A, James C K, 1983. *Population Growth*, *Employment and Economic-Demographic Interactions in Kenya*: *Bachue-Kenya*. New York: Martin′s Press.

Todaro M P, 1969. A model of labour migration and urban unemployment in less developed countries. *The American Economic Review*, 59(1): 138–148.

Wang Huabo and Fan C Cindy, 2009. The influence factor analysis on population migration among the Chinese provinces. *Population Journal*, 5: 50–53. (in Chinese)

Yang Yunyan, Chen Jinyong, Liu Ta, 1999. Population migration in China: Multi-region model and empirical analysis. *Population Science of China*, 73: 20–26. (in Chinese)