A Simulation and Application of Population and Employment Spatial Distribution at a Micro-scale: A case study of Beijing

Lianna He, Xiaochun Huang, Chengyin Cui and Zhenzhen Cui

Beijing Municipal Institute of City Planning & Design Beijing, China
email corresponding author: helianna@bmicpd.com.cn

Abstract

An important challenge in urban planning is to obtain detailed population and employment data, especially the spatial distribution of population and employment. This study uses data mining, GIS technology and advanced scheduling models to investigate the spatial distribution of an urban population at a micro-scale. Factors that influence urban spatial distribution were identified and their relationships with per capita construction area were analyzed. The normalized correlation coefficient was adopted as the weight coefficient for quantitative analysis, and a mathematical model of population distribution was developed. Beijing was used as a case study for testing the proposed technique. Data on land use, population and employment for Beijing were obtained from the national censuses, statistical yearbooks and other official sources. The proposed model was found to be suitable for application to spatial distribution problems involving urban populations. This study demonstrates how data mining can improve understanding of population and employment distribution.

1. Introduction

Population and employment are two major problems that confront humanity in the 21st century. China, as the most populous country in the world, faces a particular challenge. Solutions are critical to improve living standards and population quality, and to guarantee national strength. Therefore detailed demographic and employment data are urgently needed. Besides age structure, education level, employment type and capacity, it is essential to establish a refined spatial distribution, which can help devise scientific population policy, reasonably utilize human resources, and effectively distribute resources and urban layout. Population and employment information can be integrated with social information, such as resources, environment and economy, only if it is fused with spatial geographical information. Such integrated information can enable detailed research. Focusing on Beijing, this paper studies how to realize a scientifically reasonable population and employment distribution at a micro-scale based on large volumes of social information, 3S technologies, advanced scheduling models, and typical case analysis. This paper thus further supports the research on population and employment distribution.

2. Study area and data sources

The research in this paper covers the whole geographic area of Beijing at a micro-scale of land use units. The land use data include: (1) a map of land use in Beijing in 2010(based on 0.5 meter resolution remote sensing images, and categorizing property into living and employment types); (2) building distribution data (including the height, density, plot ratio and age of all constructions). The residential population data include: (1) Beijing data from the Fifth National Population Census, conducted in 2000,obtained from sub-district offices; (2) data from the Statistical Year book on the Residential Population of Beijing and its Sub-districts, for 2005, obtained from sub-district offices; (3) Beijing data from the Sixth National Population Census, conducted in 2010, obtained from sub-district offices; (4) population data from sample studies conducted on 2000 populated areas in Beijing in 2009, representing a 10% sampling rate of the occupied land area within the city (data included location of residential area, type, age, building volume, number of persons per household, residential population, household income etc.). The employed population data includes: (1) employed population data from the first economic census, conducted in 2005, obtained from sub-district offices; (2) employed population data from the second economic census, conducted in 2010, obtained from sub-district offices. All data are published by official organizations, cover the whole geographic area of Beijing, and are genuine and informative.

3. Research design and technique

3.1. Research objectives

Our access to multi-spatial data resources has improved with the development of computer technology, network technology and communication technology, and the establishment of mechanisms for data sharing among administrative organizations and platforms for data exchange. However, differences in data format, management and content limit the utilization of available data. This study creates spatial data on population and employment at a multi-time micro-scale by using information gathered by government organizations and dedicated researchers, and carefully combining large scale socioeconomic data on administrative boundaries and micro data on urban planning land use and construction distribution (see Fig.1).The data thus obtained are then applied to urban planning and research.

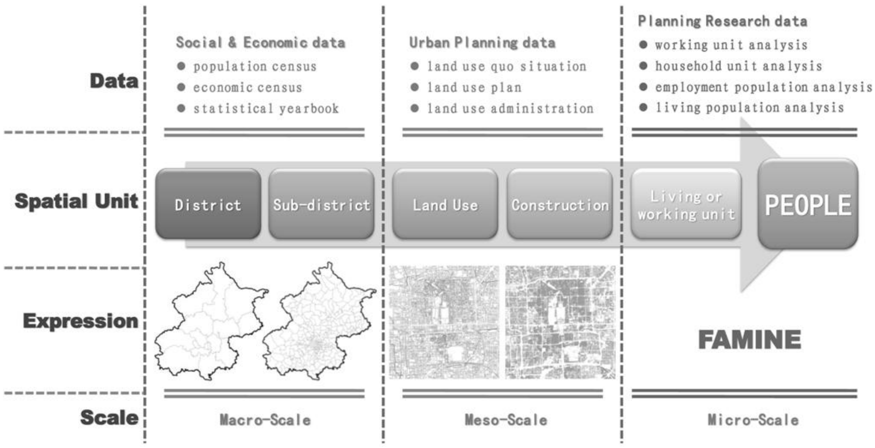

Fig. 1. Hierarchy analysis of multi-scale spatial data related to urban and rural planning

3.2. Research design

Density is a basic index of geographical distribution. The population density of inhabitancy and employment reflects the density degree and distribution of inhabitancy and employment in an area. The status of population and employment in any area can be calculated according to this density. The traditional extensive calculation focused on area density ignores the influences of land properties, architectural form and distribution, and cannot truly reflect the clustered status of inhabitancy and employment.

The key variables of this paper are “per capita building area of residential land of type X” and “per capita building area of employment land of type X”. This research incorporates the following assumptions and characteristics: (1) the superincumbent decomposition method is applied to translate official data recorded using larger units into data expressed in smaller units according to a fixed method, to prevent conflict between the total volume and authoritative data. (2) We assume full occupancy of land and construction for all the study objects, regardless of inhabitancy type and employment, and regardless of whether the occupants are owners or renters. (3) With regard to residential land, we only consider residential types oriented to family life (for example: residential quarters), not those oriented to collective life (such as: student dormitories, prisons and hotels). (4) With regard to employment land, because of the different systems used for land planning classification and national economic planning, and the difficulty of data comparison and transfer between these systems (for instance: the land use needs of the automobile industry span production, commercial service, and maintenance service), we consider all sub-categories simply as employment land (see Table.10). (5) We adapt the sample data-driven methods to the key parameter settings. (6) We emphasize the spatial relationships between basic infrastructure and public service facilities, such as the spatial distribution of the population, municipal traffic, and scientific, educational or other social services.

3.3. Technique

The technique is demonstrated by application to the spatial distribution of the residentialpopulationof Beijing.

3.3.1. Select method to model population spatial distribution

Analytical methods commonly applied to the spatial distribution of population include: (1) the population distribution equivalent region analysis method (Choropleth) (Dixon1972; Wright1936), which is suitable for grasping the macro trend of population distribution, but cannot reflect true residential distribution due to the large variation in population density within sectors; (2) the spatial distribution of population density attenuation model developed by Clark C (1951)and Sherratt (1960), which is applied to central cities and assumes that the spatial distribution of the urban population decreases progressively from central cities to their surrounding areas; (3) the digital population model(DPM), which defines a triaxial finite sequence $\{V_i = (A_i, B_i, C_i), i=1, 2, \ldots\ldots, n\}$, where $(A_i, B_i \in D)$ is a horizontal coordinate at the center of grid i, and C_i is the population in this grid. DPM reflects real population at the intension level by fully considering natural environment, such as elevation, terrain and green areas, congested and sparsely populated areas, and vocational partitions resulting from urban construction and socio-economic factors. Considering available data, this paper adopts the digital population model as being the best among the three methods.

3.3.2. To analyze similarity and difference among residential samples

We found high similarity in living area per capita within specific residential types and a big difference among different residential types based on the sample data (see Fig.2). We classify five types of residential areas as shown in the table below (see Table 1). The first type is the villa. This type of residential area has complete facilities, well arranged layout, good environment, and usually a low building density, plot ratio, and population density. Most villas are detached dwellings (three story or less) with areas of around 500-1000 square meters per household, and 200 square meters per capita. The second and third types are ordinary residential areas. These have a well arranged layout, with complete facilities and a good environment, and most are multi-story, medium-rise or high-rise commercial housing developments, constructed after 2000 – when the real estate development industry grew and the focus shifted to improving living conditions. Before 2000, most houses were built by companies for their workers. This type of housing usually comprised small apartments, with an area of 40 square meters for a one bedroom apartment, 60 square meters for two bedrooms, 75 square meters for three bedrooms, and a per capita building area of 18 square meters. Since 2000, most residential area developments have been commercial, and house sizes have greatly increased, with an area of 75 square meters for one bedroom, 110 square meters for two bedrooms, 140 square meters for three bedrooms, and a per capita building area of 35 square meters. The fourth type is the simple bungalow with general public service facilities, incomplete layout, bad environment, location on or among industrial land, and per capita residential area of less than 10 square meters. The fifth type is the residential area in the town or countryside. These residential areas are characterized by collective land ownership, low construction cost, and per capita residential area of almost 100 square meters.

R1 : villa	R2 : typical residential area built after 2000	R3 : typical residential area built in the 1990-2000
R5 : typical residential area in the countryside	R4 : simple cottages built before the 1970s	R3 : typical residential area built in the 1980-1990

Fig. 2. Pictures of typical residential areas from different periods

Table 1. Description of Residential Area Categories

Class code	Resident	Residence type	Per capita residential building area	Explanation
R1	Rich people in city Annual household income \$200 thousand	High-grade residential area	200m^2 per capita	Low building density, low plot ratio, low population density, extremely good afforestation in urban community, most of them are villa and semi-detached villa district
R2	Middle class in city Annual household income \$50-200 thousand	Ordinary residential area (constructed after the year of 2000)	35 m^2 per capita	Well overall arrangement with good environment, most of them are multi-storey, high-storey and middle-storey dwellings and constructed after the year of 2000
R3	Populace in city Annual household income \$10-50 thousand	Ordinary residential area (constructed before the year of 2000)	18 m^2 per capita	Well overall arrangement with good environment, most of them are multi-storey, high-storey and middle-storey dwellings and constructed before the year of 2000
R4	Poor population Annual household income less than \$10 thousand	Simple bungalow	10 m^2 per capita	Incomplete overall arrangement, bad environment crossover with industrial landing

R5	Farmer	Residential area in town and country	100 m^2 per capita	Located in the country-side, incomplete overall arrangement, bad environment, bad construction with a big area per capita

Per capita living space differs between geographical zones, even for a single type of residential area. We suppose Z_i (i=1, 2, ……307) is a unit of space in some street or town, which is recorded in the official demographic data; R_j (j=1, 2, ……5) is a residential land type. We can calculate per capita housing volume "$\square'_{Z_iR_j}$" for the areas characterized as different type R_j and different unit Z_i based on sample data. We can then calculate the proportional relationship to capacity of the resident population in different types of residential land based on construction volume in real R_j residential land, and then total population in unit Z_i relative to the total real population in R_j residential land according to the above proportional relationship. Table 2 describes different variables and their sources.

$$\theta'_{Z_iR_j} = \frac{S'_{Z_iR_j}}{P'_{Z_iR_j}} = \frac{\sum_{k=1}^{n} S'^{k}_{Z_iR_j}}{\sum_{k=1}^{n} P'^{k}_{Z_iR_j}} \quad \textbf{(Eq. 1.1)}$$

$$P_{Z_iR_j} = P_{Z_i} \cdot \frac{S_{Z_iR_j}/\theta'_{Z_iR_j}}{\sum_{j=1}^{n} S_{Z_iR_j}/\theta'_{Z_iR_j}} \quad \textbf{(Eq. 1.2)}$$

$$S_{Z_iR_j} = \sum_{l=1}^{n} S^{l}_{Z_iR_j} \quad \textbf{(Eq. 1.3)}$$

$$P_{Z_iR_j} = \sum_{l=1}^{n} P^{l}_{Z_iR_j} \quad \textbf{(Eq. 1.4)}$$

Table 2. Variable Sources and Descriptions

Variable	Implication	Source
$P'_{Z_iR_j}$	The population in residential landing type J in street or town of NO. i	See Eq. 1.1
$P'^k_{Z_iR_j}$ (k=1, 2, ……n)	The population of NO. k residential district in residential landing type J in street or town of NO. i	Sample survey for residential quarter in 2009
$S'_{Z_iR_j}$	The construction volume in residential landing type J in street or town of NO. i	See Eq. 1.1
$S'^k_{Z_iR_j}$ (k=1, 2, ……n)	The construction volume of NO.K residential district in residential landing type J in street or town of NO. i	sample survey for residential quarter in 2009
$\theta'_{Z_iR_j}$	The construction volume per capita in residential landing type J in street or town of NO.i	See Eq. 1.1
P_{Z_i}	The population in street or town of NO. i	population census data
$P_{Z_iR_j}$	The population in residential landing type J in street or town of NO.i	See Eq. 1.2
$P^l_{Z_iR_j}$ (l=1, 2, ……n)	The population in first residential landing in residential landing type J in street or town of NO. i	?
$S_{Z_iR_j}$	The construction volume in residential landing type J in street or town of NO. i	See Eq. 1.3
$S^l_{Z_iR_j}$ (l=1, 2, ……n)	The construction volume in first residential landing in residential landing type J in street or town of NO. i	Land use status chart

Whether per capita residential area is constant across different blocks, given the same residential land type R_j, and the same unit Z_i, in a street or town is a key question. Analysis of sample data showed that per capita residential area differs even within the same type of residential land, because coverage of populated area Z_i differs among urban regions. For instance, if Z_i is a less than 5 square kilometer area in the central city, the per capita construction volume will be comparable among different residential areas.

However, if Z_i denotes the urban-rural fringe, and has an area of more than 10 square kilometers, differences in infrastructure and public service facilities will cause large variance in the concentration of residential areas.

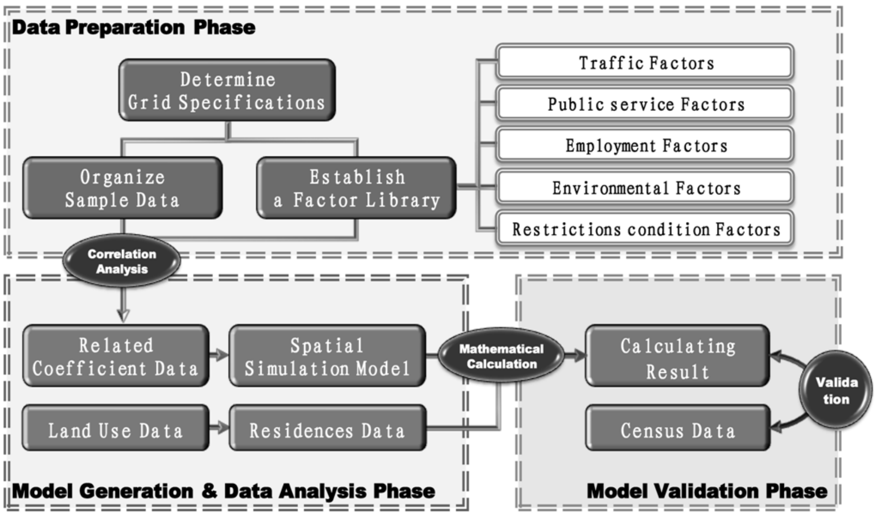

Fig. 3. Data processing and analysis techniques

Therefore, the second step is to establish a spatial population distribution model by analyzing the spatial connection between samples in different types of residential area R_j and factors influencing urban land use arrangement, and then to calculate the distribution of the current residential population (see Fig.3). This method requires relative analysis of the relationship between samples of the per capita residential area and factors influencing spatial distribution in a city. Correlation analysis is a statistical method commonly used to measure the intimacy degree of various variables, and also the degree of linear relation and direction between two variables. The numerical value of a related coefficient ranges between -1 and +1: the bigger the absolute value, the higher the interdependence. Three related coefficients usually exist between two assigned variables, namely the Pearson correlation coefficient, Spearman's rank correlation coefficient, and Kendall rank correlation coefficient. The latter two are suitable for rank correlation analysis, while, the Pearson coefficient is suitable for analysis of constant variables or measurement of equal intervals. We take the Pearson coefficient as the correlation value of factors, and calculate the normalized weight coefficient as the scientific measure of population distribution.

3.3.3. Determine grid specifications and establish a factor library of spatial distribution in the city

Common grid square analysis tools include square grid, Thiessen polygon mesh, and center circle analysis. After consideration of the advantages and disadvantages of the above three methods of population analysis, as well as the current data and the calculation complexity, we use a square grid with side lengths of 1km, 2km and 3km. 1km represents a convenient walking distance for daily life activities, while 3 km is the upper limit for easy shopping. The advantages of square grid are that the studied units are constant, do not overlap, and have standardized shape and area. The factor library of urban spatial distribution is based on years of accumulated data gathered or provided by various official and academic organizations. This factor library mainly includes five factors: traffic condition, public service facilities condition, employment condition, environment condition and restriction condition.

3.3.4. Analysis of the degree of association and normalized weight

Single factor analysis

Traffic is important to urban spatial arrangement, and travel convenience can reflect the skeleton or texture of a city. The traffic factor is divided into three sub-factors: urban traffic, urban rail traffic, and outbound traffic. The analysis of these sub-factors is shown in Table 3.

Table 3. Analysis of the traffic factor

Association Class		Associated content		Degree of association (%)					
Class	Sub-class	Field name	Field interpretation	1K.A	1K.B	2K.A	2K.B	3K.A	3K.B
Traffic	road traffic	D-GJ	The shortest distance to bus station	18.4	13.8	18.4	12.0	18.4	11.8
		D_GJX	The shortest distance to bus line	21.6	16.2	21.6	14.1	21.6	13.9
		T_GJ	The amount of bus stations in grid	12.2	9.2	21.4	13.9	19.4	12.5

		T_GJ_L	The total road length in grid	15.1	11.3	14.1	9.2	13.7	8.8
	track traf-fic	D_DT	The shortest distance to subway	18.1	13.6	18.1	11.8	18.1	11.6
		D_DTX	The shortest distance to subway line	17.0	12.8	17.0	11.1	17.0	10.9
		T_DT	The amount of subway station in grid	4.4	3.3	10.9	7.1	13.7	8.8
		T_DT_L	The total subway length in grid	6.3	4.7	11.8	7.7	13.4	8.6
	Out-bound traffic	D_WT	The shortest distance to outbound traffic sta-tion	20.2	15.2	20.2	13.2	20.2	13.0

note: 1k: square grid with side lengths of 1 km; 2k: square grid with side lengths of 2 km; 3k: square grid with side lengths of 3 km
A: association degree, namely the weight coefficient; B: normalized weight coefficient(as listed in the subsequent table).

The distribution and intensity of public service facilities clearly reflect the level of convenience, prosperity and development potential. This paper applies proximity analysis and density analysis to public service facilities. According to urban planning land use classification criteria, service facilities can be categorized as: administrative (C1), commercial(C2), cultural and entertainment (C2), sporting(C4), health care services (C5), education and scientific research institutions (C6), facilities housing cultural relics (C7), and municipal services (U). Table 4 lists the analysis results.

Table 4. Analysis of the public service factor

Associa-tion Class	Associated con-tent	Degree of association (%)

Class	Sub-class	Field name	Field inter-pretation	1K.A	1K.B	2K.A	2K.B	3K.A	3K.B
Pub-lic ser-vice	Jobs	D-XZ	The short-est distance to adminis-trative cen-ter	8.9	5.6	8.9	5.4	8.9	4.9
		D_SY	The short-est distance to commer-cial center	19.4	12.2	19.4	11.8	19.4	10.8
	Ser-vice facility	SO	The amount of public ser-vice fa-cility in grid	14.4	9.1	19.2	11.7	20.9	11.6
		S_C1	The amount of public ser-vice fa-cility of class C1 in grid	30.9	19.5	22.1	13.5	21.1	11.7
		S_C2	The amount of public ser-vice fa-cility of class C2 in grid	14.4	9.1	17.6	10.7	20.4	11.3
		S_C3	The amount of public ser-vice fa-cility of class C3 in grid	8.9	5.6	16.7	10.2	15.4	8.5
		S_C4	The amount of public ser-vice fa-cility of class C4 in grid	10.0	6.3	14.4	8.0	8.3	4.6

		S_C5	The amount of public ser-vice fa-cility of class C5 in grid	10.0	6.3	14.1	8.6	12.6	7.0
		S_C6	The amount of public ser-vice fa-cility of class C6 in grid	9.4	5.9	11.4	6.9	16.1	8.9
		S_C7	The amount of public ser-vice fa-cility of class C7 in grid	16.7	10.5	5.4	3.3	15.8	8.8
		S_U	The amount of public ser-vice fa-cility of class U in grid	15.8	9.9	15.1	9.2	21.4	11.9

To shorten commute times, most residents of metropolises in China favor residential areas located close to their work, even if this means sacrificing quality of external environment. Therefore, employment condition and employment distribution are increasingly closely related. The factors of employment condition in this paper fall into two classes: job analysis and employment land analysis.

Employment land is calculated using the “Technical standards for Controlling Detailed Planning in the Central City of Beijing (2008.2 version)” and the standard for employment estimate in the central city given in Table 5.

Table 5. Standards used for employment estimate

Land classification		Construction area per capita (square meter per capita)	
C1	administration	building area	50
C2	commercial finance	building area	25
C3	cultural entertainment	building area	50
C4	sports facility	building area	75
C5	health care	building area	40
C61	universities and colleges	building area	200
C65	scientific research	building area	50
R5	elementary, secondary school and kindergarten	building area	60
U	Municipal transportation infrastructure	building area	20
M1	industrial land	building area	50
W	warehouse land	building area	50
D1	military land	building area	80

Employment center is identified using color gradation analysis charts based on the above standards. The colors range from dark to light to show job quantity within a unit grid, and the areas with the densest job allocation are clearly Zhongguancun, Financial Street, Jian'guomen, the Central Business District, and Sanyuan Bridge. All these areas provide more than 20000 jobs per square kilometer. (see Fig.4).

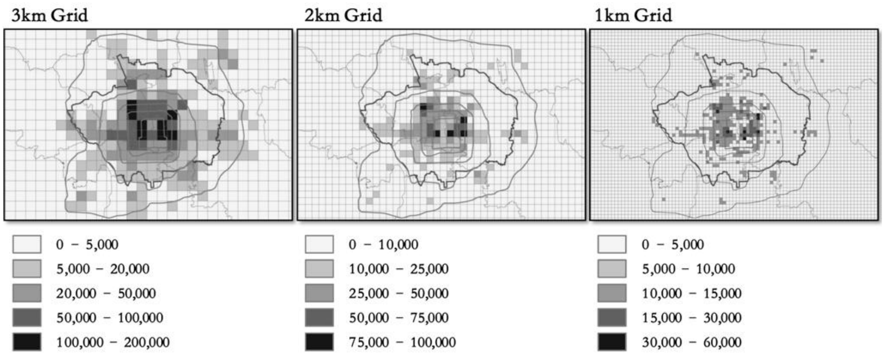

Fig. 4. Job quantity analysis chart for different grid scales

Table 6. Analysis of factors that influence employment

Association Class		Associated content		Degree of association (%)					
Class	Sub-class	Field name	Field interpretation	1K.A	1K.B	2K.A	2K.B	3K.A	3K.B
Employment	Jobs	D_JOB	The shortest distance to job center	16.7	15.2	16.7	15.1	16.7	13.0
		E_JOB	Post amount in grid	22.3	20.3	19.2	17.1	23.4	18.3
		E_100	The amount of top 100 enterprises in grid	9.4	8.6	10.0	9.1	18.7	14.6
	Employment land	YD_C	Total land area for type C in grid	17.0	15.5	14.4	13.0	17.3	13.5
		JZ_C	Construction volume of type C in grid	15.8	14.4	13.4	12.1	13.4	10.5
		YD_M	Total land area for type M in grid	12.6	11.5	17.3	15.7	19.4	15.1
		JZ_M	Construction volume of type M in grid	15.8	14.4	19.4	17.6	19.2	15.0

The results are obtained through comprehensive analysis of the relationship among all factors influencing employment condition and per capita residential area in Table 6.

Environment condition refers to the eco-environment status of a residential area located on a grid. It is determined by public green area, river area, scenic spot area and park quantity. The analysis is shown in Table 7.

Table 7. Analysis of environmental factors

Association Class		Associated content		Degree of association (%)					
Class	Sub-class	Field name	Field interpretation	1K.A	1K.B	2K.A	2K.B	3K.A	3K.B
Environment	Environmental greening	S_G	Total number of public service facilities of type G in grid	11.8	55.7	10.0	44.2	20.2	76.2
		YD_GE	Total land area of G and E1 in grid	9.4	44.3	12.6	55.8	6.3	23.8

Restrictions condition refers to space distribution factors that are incompatible with high population concentration, such as cemeteries or conservation districts associated with historic sites, as listed in Table 8.

Table 8. Analysis of restrictions condition factors

Association Class		Associated content		Degree of association (%)					
Class	Subclass	Field name	Field interpretation	1K.A	1K.B	2K.A	2K.B	3K.A	3K.B
Re-striction	Restricted condition	D-GM	The shortest distance to cemetery	17.6	59.1	17.6	59.1	17.6	59.1
		D_LS	The shortest distance to conservation districts of historic sites	12.2	40.9	12.2	40.9	12.2	40.9

Dual factors analysis

This paper discusses the factors that strongly influence the keyvariable of per capita construction area, which include traffic condition, public service condition, employment condition, environment condition and restrictions condition. Table 9 categorizes the weight coefficients of factor analysis for each factor to rank the influence of the factors.

Table 9. Normalized weight coefficient for factors

Factor category	Normalized weight coefficient (%)		
	1K	2K	3K
Traffic conditions	28.7	31.2	29.2
Public service facility conditions	36.8	24.6	36.1
Employment conditions	23.6	22.4	24.1
Environment conditions	4.6	4.6	5.0
restricted conditions	6.4	6.1	5.6

Table 9 shows the influence of five categories of city layout factors on per capita residential area under different grids, and reveals public service facilities as the most important influence, followed by traffic condition and employment condition, while the influences of the environment and restrictions conditions are smaller.

3.3.5. Mathematical calculation of population distribution for residential land within a single area and with same type but different location

Based on data analysis, we propose calculating current population distribution by measuring the factor quantification in each unit of each grid using GIS technology, as follows:

X_{11}, X_{12}, ……, X_{1q}, (q=9) ——factors related to traffic condition

X_{21}, X_{22}, ……, X_{2q}, (q=11)——factors related to public service facility condition

X_{31}, X_{32}, ……, X_{3q}, (q=7) ——factors related to employment condition

X_{41}, X_{42}, ……, X_{4q}, (q=2) ——factors related to environment condition

X_{51}, X_{52}, ……, X_{5q}, (q=2) ——factors related to restrictions condition

X is a regularization value designed to handle the big gap in the wild function range of each factor (see Eq.2.1). Among these regularization values, X_{max} denotes the biggest, and X_{min} the smallest, which enables convenient weighting through integrating these values between 0 and 1.

$$X_{\text{rule value}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (\text{Eq.2.1})$$

Suppose the normalized weight coefficient for single factor is b_{uv}, and the weight coefficients for other similar factors are a_1,a_2, a_3, a_4, a_5, then the comprehensive weight coefficient for each relative factor is $a_u b_{uv}$, and the weight coefficient C_l for population distribution in block l of residential area in grid R_j can be calculated using Eq.2.2. Among the calculated values, l is a serial number. This method supposes relative factors as weight coefficients; the total sum for normalized weight coefficients in the same type of factor is 1.

$$C_l = \frac{a_u b_{uv} X_{uv}}{\sum_{u=1}^{5} \sum_{v=1}^{q} a_u b_{uv} X_{uv}} \quad (\text{Eq. 2.2})$$

This method considers numerous factors related to city layout to calculate per capita residential area in different parcels of residential land of type R_j. However, ignoring the relevance of independentvariablefactors impacts the final result. Besides, this calculation method has to be processed under a known population and construction volume, which naturally creates limitations.

3.3.6. Calculation of employment distribution

The method used to calculate residential population is suitable for calculating the spatial distribution of employment population, but must establish the one-to-many relationship between city planning land category and national economic category. This paper classified 395 industries to analyze their use of urban-rural planning land, as shown in Table 10. The calculation method is characterized by the employment population average principle under the same zone. Per capita employment construction area of type X is the core key variable.

Table 10. Comparison table for national economic category and urban-rural planning land (excerpt)

Category	National economic category		Category for urban and rural planning land	Land function classification explanation
Class	P	Education		
Large Class	84	Education		
Sub-class	841	Preprimary education	R5	Education land equipped with residential land
Sub-class	842	Elementary education	R5	Education land equipped with residential land
Sub-class	843	Secondary education	R5	Education land equipped with residential land
Sub-class	844	Higher education	C61	Higher education land
Sub-class	849	Other education	C63	Land for adult education and spare-time school
	……			

Class	Q	Health, social security and social welfare		
Large Class	85	Health	C5	land for medical purpose
Sub-class	851	Hospital	C51	land for medical purpose
Sub-class	852	Health-center and community medical service	C51	land for medical purpose
Sub-class	853	Clinic medical	C51	land for medical purpose
Sub-class	854	Motherhood and child protection	C51	land for medical purpose
Sub-class	855	Technical service of family planning	C51	land for medical purpose
	……			

4. Researchresults and application example

4.1. Research results

The results realize refined spatial distribution simulation between population and employment at a micro-scale from the space dimension, and establish spatial data that link employment and multi-layer population (such as district and sub district), as well as land use, construction, and human behavior. To take the residential population and employment population in the Financial Street Area as an example, as Fig.5 and Fig.6 show, each point in the figures represents 50 people, which eliminates the distortion in previous extensive average space distribution, and provides solid data for refined planning formation, research and management.

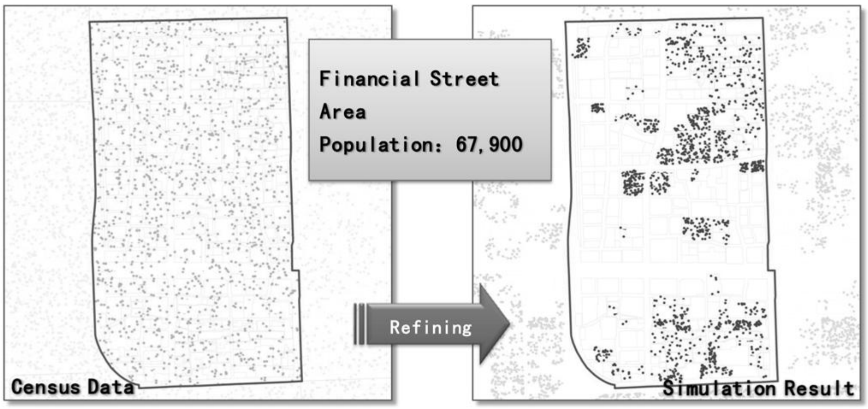

Fig. 5. Spatial distribution of residential population in financial street offices

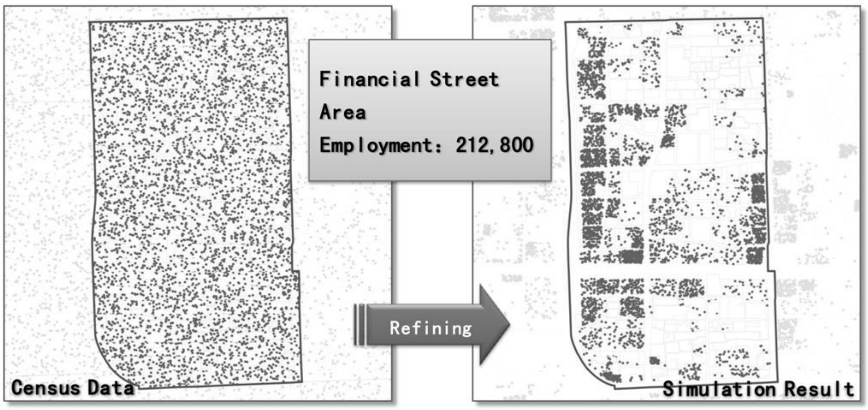

Fig. 6. Spatial distribution of employment in financial street offices

For the error analysis, we took Xicheng District as an example. Under the premise of the known total population and residential building area, we did the simulation calculation, and got the following result (table 11).

Table 11. Error Analysis of population distribution simulation results

Sub-Zone Name Of Xicheng District	Census Data	Statistics of Calculating Result	Error Rate
Changqiao	92158	59587	-35.3%
Yuetan	149121	161818	8.5%
Xinjiekou	80873	49968	-38.2%
Xichanganjie	81282	56567	-30.4%
Fengsheng	46322	32213	-30.5%
Fuwai	71596	84078	17.4%
Erlonglu	64411	36027	-44.1%
Fusuijing	84514	65515	-22.5%
Zhanlanlu	77306	89681	16.0%
Dewai	119523	204645	71.2%
Total Population of Xicheng District	869113	840099	-3.3%

By comparison between the census data and statistics of calculating result, we recognize there is being a gap, but the population proportions of each sub-zone is roughly consistent (Fig. 7).

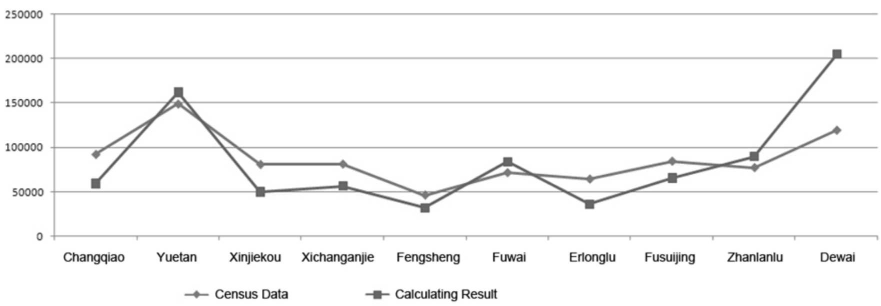

Fig.7. Error Analysis of population distribution simulation results(Xicheng District)

4.2. Planning application

4.2.1. Case 1: Concentration degree of Beijing population analysis and dynamic simulation

A city belongs to its people, so city planning must center on human demands. Therefore, understanding of the dynamic feature of the quantity and spatial distribution of residents is the basis for effective arrangement of city planning land and public service facilities. According to the research method, we realize the spatializationthroughapplying human rules to the data from the Fifth National Population Census in 2000, the Statistical Yearbook on the Residential Population of Beijing and its Sub-districts in 2005, and the Sixth National Population Census in 2010.The results show that population growth in Beijing remains centered on the central city and new towns from 2000 to 2010. Notably, population growth from 2000 to 2005 was concentrated between the second and fourth ring roads in the north, while, from 2005 to 2010, with the implementation of the south-east oriented development direction, population growth was focused in the south-east, and the population concentration in the main function zone also changed considerably, especially in Zhongguancun. Meanwhile, three new cities (Yizhuang, Shunyi and Tongzhou) alleviated population pressures in the central city. (Fig. 8)

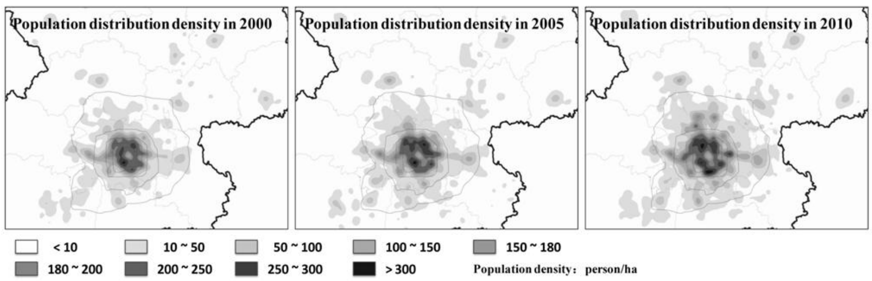

Fig. 8. Population concentration in Beijing from 2000 to 2010

4.2.2. Case 2: Analysis of employment and residence in the central city

Employment and residence are two main elements of land use in the city, and their spatial relationship largely decides features and traffic efficiency. Therefore, how to effectively allocate land use between employment and residency is a permanent topic of interest. The spatial relationship between employment and residency during China's central planning economy was characterized by work unit housing compounds, which were a balanced solution for the conditions of the time. With the marketization of employment and residency that followed the economic reform, spatial isolation between employment and residency increased and caused various problems. Based on the result of the refined spatial distribution between the residential and employment populations, quantitative analysis of the employment and residency situation was conducted in various planning areas in the central city (see Fig. 9). The proportions of employment and residency in the west and north-west areas within the second ring road are well balanced according to the international reference standard of 0.5-0.7. Meanwhile, residency slightly exceeds employment in the south-east, employment considerably exceeds residency within the second ring road in the south-west and east, and residency exceeds employment in the periphery.

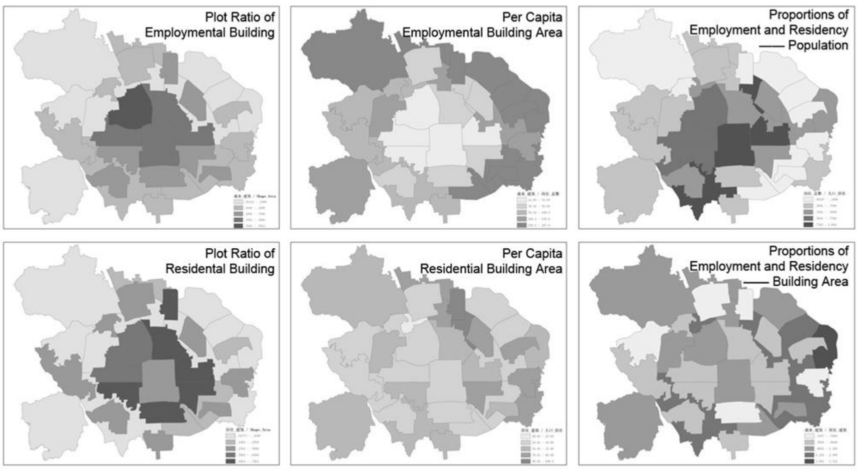

Fig. 9. Analysis of employment and residency in central Beijing in 2010

4.2.3. Case 3: Analysis of residency capacity and employment land use

Employment and residency land, and construction volume are important factors for measuring the reasonability of urban land use arrangement, and also are one of the indexes used to indicate urban quality of life and analyze planning direction. This research can calculate the per capita construction index of construction land in different cities and regions (see Table 12), and then determine population volume in different areas. The differences between this population volume and planning targets provide important evidence for regional development analysis, development status control, and construction process evaluation.

Table 12. Comparative analyses of construction targets and planning of residency and employment in the central city and new developing zone (unit: square meters per capita)

<table>
<tr><td colspan="2">Class</td><td>R1</td><td>C1</td><td>C2</td><td>C3</td><td>C6</td><td>……</td><td>M</td><td>U1</td></tr>
<tr><td colspan="2">Programming guidelines</td><td>30</td><td>50</td><td>25</td><td>50</td><td>50</td><td></td><td>50</td><td>20</td></tr>
<tr><td rowspan="5">Main Central city</td><td>Old city</td><td>21</td><td>51</td><td>20</td><td>40</td><td>20</td><td></td><td>32</td><td>2</td></tr>
<tr><td>Central area(outside the old city)</td><td>28</td><td>82</td><td>25</td><td>32</td><td>32</td><td></td><td>55</td><td>15</td></tr>
<tr><td>Bianyuan Group</td><td>32</td><td>60</td><td>46</td><td>95</td><td>65</td><td></td><td>120</td><td>23</td></tr>
<tr><td>Green Space</td><td>29</td><td>45</td><td>40</td><td>163</td><td>25</td><td></td><td>219</td><td>15</td></tr>
<tr><td>Total sum to central city</td><td>29</td><td>69</td><td>28</td><td>45</td><td>36</td><td></td><td>122</td><td>19</td></tr>
<tr><td rowspan="2">New Developing Zone</td><td>North-west part</td><td>32</td><td>84</td><td>10</td><td>125</td><td>479</td><td></td><td>217</td><td>180</td></tr>
<tr><td>South-west part</td><td>33</td><td>69</td><td>25</td><td>81</td><td>87</td><td></td><td>189</td><td>101</td></tr>
</table>

5. Conclusions

This paper attempts to solve the spatial distribution of urban population using data mining technology. Using current GIS technology and spatial data, we analyze the relationship between per capita construction area and various factors by identifying factors that influence urban spatial distribution, which include construction condition, traffic condition, public service facilities condition, employment condition, environment condition and restrictions condition. We then establish a mathematical model of population distribution by taking the normalized correlation coefficient as the weight coefficient of quantitative analysis, and finally realize a scientifically reasonable quantitative distribution of residency. This residency distribution provides a solid base for better management and utilization of population information, and for comprehending the relationship among population, resources and environment.

References

Chen Nan, etc. GIS and quantitative analysismethod in populationeconomics [M]. Beijing: sciencepress, 2007.
Li Chenming, etc. Population geographical information system [M]. Beijing: sciencepress, 2005.
Ye Jia'an, etc. Geographic information and planning support system [M]. Beijing: sciencepress, 2006.
Li Heping, etc. Social investigation methods for urban planning [M]. Beijing: China Building Industry Press, 2004.
Long Ying. Principle and application of planning support system [M]. Beijing: Chemical Industry Press, 2007.
Tian Li, etc. comparison of planning construction and development of the world famous metropolis [M] . Shanghai: chinabuildingindustrypress, 2010.
Zhang Hongyang, Yang Hu. Chongqing populationdistribution model and featureanalysis [J]. Population & Economics, 2003, (2): 53-58.
Huang Rongqing. The change of population distribution in the process of urbanization in Beijing since 1980 [J]. Demographic research, 2005, 29(5): 19-26.
Meng Xiaochen, etc. research review to the balance between employment and residency [J]. Urban development, 2009, 16(6): 23-26.
Liu Zhilin, etc. Segregationphenomenon between employment and residency and their features in metropolis of China [J] . Urban development, 2009, 16(9): 110-117.
Li Xin, etc. Data Assimilation——new thought to multiple geographical space data [J]. Science & Technology Review, 2004, 12: 13-16