CHINESE GEOGRAPHICAL SCIENCE
Volume 13, Number 1, pp. 34 – 38, 2003
Science Press, Beijing, China

APPLICATION OF GEOGRAPHICAL PARAMETER DATABASE TO ESTABLISHMENT OF UNIT POPULATION DATABASE

DONG Chun[1], ZHAO Rong[1], LIU Ji-ping[1], WANG Gui-xin[2], WU Xi-zhi[3]

*(1. Government GIS Research Center, Chinese Academy of Surveying and Mapping, Beijing 100039, P. R. China;
2. Institute of Population Research, Fudan University, Shanghai 200443, P. R. China;
3. Statistics Department, Renmin University of China, Beijing 100872, P. R. China)*

ABSTRACT: Now GIS is turning into a good tool in handling geographical, economical, and population data, so we can obtain more and more information from these data. On the other hand, in some cases, for a calamity, such as hurricane, earthquake, flood, drought etc., or a decision-making, such as setting up a broadcasting transmitter, building a chemical plant etc., we have to evaluate the total population in the region influenced by a calamity or a project. In this paper, a method is put forward to evaluate the population in such special region. Through exploring the correlation of geographical parameters and the distribution of people in the same region by means of quantitative analysis and qualitative analysis, unit population database (1km × 1km) is established. In this way, estimating the number of people in a special region is capable by adding up the population in every grid involved in this region boundary. The geographical parameters are obtained from topographic database and DEM database on the scale of 1∶250 000. The fundamental geographical parameter database covering county administrative boundaries and 1km × 1km grid is set up and the population database at county level is set up as well. Both geographical parameter database and unit population database are able to offer sufficient conditions for quantitative analysis. They will have important role in the research fields of data mining (DM), Decision-making Support Systems (DSS), and regional sustainable development.

KEY WORDS: geographical parameter database; unit population database; quantitative analysis; weight coefficient; standardized weight coefficient

CLC number: P208 Document code: A Article ID: 1002-0063(2003)01-0034-05

1 INTRODUCTION

Now GIS is turning into a good tool in handling geographical, economical, and population data, so we can obtain more and more information from these data (CHOU, 1997). On the other hand, more than 85% of the information on government affairs are related to the information on spatial position in the world (ZHANG, 2001). So we conclude that much information on national economy and social development has certain correlation with different geographical factors in a region. In order to explore the determinate correlation of geographical factors and economical factors by means of quantitative analysis, DONG Chun *et al.* put forward the concept of geographical parameters database (DONG, 2000; DONG *et al.*, 2001), and established the fundamental geographical factors databases covering county administrative boundaries and 1km × 1km grid within the territory of Fujian Province (DONG *et al.*, 2000a).

In some cases, for a calamity, such as hurricane, earthquake, flood, drought etc., or a decision-making, such as setting up a broadcasting transmitter, building a chemical plant etc., we have to evaluate the total

Received date: 2002-09-29
Foundation item: Under the auspices of the High-tech Research and Development (863) Program (No. 2001AA135080); Technology Base Project Foundation of the Ministry of Science and Technology of China in 2000; the National Social Science Foundation of China (No. 00BRK006).
Biography: DONG Chun (1971 –), female, a native of Shiping County of Yunnan Province, assistant professor, specialized in the fields of GIS, application of spatial data, data mining, economics of population, resources and environment etc.

population in the region influenced by a calamity or a project.

In this paper, a method is put forward in evaluating the population in such special region. By exploring the correlation of geographical parameters and the distribution of people in the same region by means of quantitative analysis and qualitative analysis, unit population database (1km × 1km) is established. In this way, estimating the number of people in a special region is capable by adding the population in every grid involved in this region.

The geographical parameters are obtained from topographic database and DEM database on the scale of 1 : 250 000. The fundamental geographical parameter databases covering county administrative boundaries and 1km × 1km grid is set up, and the population database at county level is set up as well.

2 CONTENTS OF GEOGRAPHICAL PARAMETERS DATABASE

Inhabitant sites, water system, highway line, and railway line etc. are most important landform elements. By them the particular landscape in a geographical region is described, they are important parameters because they have tight correlation with population distribution. Thus, the above-mentioned factors are important contents in geographical parameter database.

Geographical factors are the essential components in substance and energy of geographical environment. They are self-existent, having different properties, but obeying certain rules (WANG, 1997). In this paper, geographical parameter database is defined as database that stores the quantitative values of multifarious geographical factors in each geographical cell. The conception and definition can be found in DONG's reference (DONG, 2000).

Because of the special geographic conditions and landform character, population distribution in vertical direction in China has its particularity and diversity (ZHANG *et al.*, 1996). All of them are very interesting, and any other country is difficult to compare with it (HU, 1990; JENNIFER *et al.*, 1997). Therefore, altitude, slope grade, slope aspect etc. (here, they are classified by DEM factor) are the important contents of geographical parameter database. At present, they can only be obtained from DEM database on the scale of 1: 250 000.

With the development of resource satellite data at high precision, remote sensing technology, and GIS technology, it is possible to obtain high-precision land cover data, and some researches have proved that different types of land-covering can support different weights of population. So, land-covering factor should be put into geographical parameter database.

Economical development factors, such as Gross Domestic Product (GDP), yearly output of crop, amount of pupil, aggregate investment of fixed assets and so on, are correlative factors of population distribution. The relationship between geographical factors and economical development factors can be examined from the next two illustrations (Fig. 1). In Fig. 1a, the vertical axis represents amount of total population in each county, and the horizontal axis represents amount of pupil in the corresponding county in Guizhou Province in 2000. In Fig. 1b, the vertical axis represents amount of total population, and the horizontal axis represents yearly output of crop in the corresponding county in Guizhou Province in 2000. Every dot plotted in the diagrams represents a county according to its corresponding values. The strong relations are found between these two variables in the diagrams. The computed PEARSON's correlation coefficient (DONG, 2000; DONG *et al.*, 2000b) equals 0.925 and 0.883 respectively ($n = 83$), which indicates a significant positive correlation between the two groups of variables. But, because it is difficult to acquire the quantitative value of an economical development factor in each cell at size of 1km × 1km, in this paper, they are omitted very reluctantly.

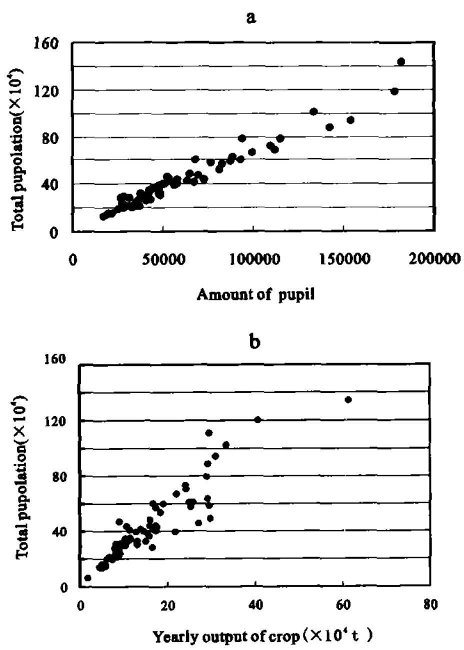

Fig. 1 The relationship between geographical factors and economical development factors in Guizhou Province in 2000

3 ESTABLISHMENT OF CORRELATION FACTORS MODEL OF POPULATION DISTRIBUTION IN GUIZHOU PROVINCE

3.1 Factor Analysis in the Same Sort

In this paper, geographical parameters database and population database covering county administrative boundaries in Guizhou Province have been established. There are such factors as landform factor, DEM factor, land cover factor and so on in the former database, and total population in the latter one. After multi-factor analysis, the correlation between geographical factors and the factor of total population is shown by Pearson's correlation coefficient. Then weight coefficients and standardized weight coefficients of every parameter and every group of factors are acquired.

3.1.1 Landform factor

Though, the quantitative value in any geographical cell is the accumulative total of length (arc), number of points (point), or area (polygon) (DONG *et al.*, 2001). In consideration of the error of the spatial database on the scale of 1∶250 000, the factors which are polygons in map, such as residential parcels and water system represented as polygons, are omitted, and only the parameters which are represented as lines and points are applied in the analysis. Table 1 shows the landform factor's weight coefficient and standardized weight coefficient to total population in the same unit. Here, there are four factors only, and we obtained them by adding up all the factors in the same sort.

Weight coefficients indicate the degree and direction of the relationship between the variable of landform factors and the variable of total population. The value ranges from 0 to 1. A higher value of one variable implies a lower value of the other. If the value is not significantly different from 0, there is no correlation between the variables, or they are independent each other. Standardized weight coefficients come from weight coefficients, the sum of all standardized weight coefficients in Table 1 is exactly 1.

Table 1 The weight coefficient and standardized weight coefficient of each landform parameter to total population

Landform factor	Weight coefficient	Standardized weight coefficient
Number of habitat sites	0.789	0.482
Length of railway line(km)	0.059	0.036
Length of Highway line(km)	0.506	0.310
Length of water system(km)	0.282	0.172

From Table 1 we know that habitat sites and highway line have stronger correlation to population distribution than railway line and water system in Guizhou Province.

3.1.2 Altitude factor

The characteristics of population distribution in vertical direction show that population distribution has intensive tendency to the low and flat topography. That is, with the ascent of altitude, population drops promptly (ZHANG *et al.*, 1996).

But because a majority of land in Guizhou Province is located in alps and coulees, population distribution in vertical direction in this province has its special phenomenon. That is, population inhabit mainly in the medium altitude of 800 – 1000m and 1200 – 2000m, and less people live in the region with low altitude. This is contrary to the common law of population distribution in vertical direction. Table 2 shows the weight coefficient and standardized weight coefficient of each altitude belt to total population.

Table 2 Weight coefficient and standardized weight coefficient of each altitude belt to total population

Altitude (m)	Weight coefficent	Standardized weight coefficient
<100	0.019	0.006
100 – 300	0.179	0.055
300 – 500	0.442	0.136
500 – 800	0.050	0.015
800 – 1000	0.724	0.223
1000 – 1200	0.004	0.001
1200 – 1500	0.683	0.212
1500 – 2000	0.669	0.206
2000 – 3000	0.356	0.110
3000 – 5000	0.116	0.036
> 5000	0.000	0.000

3.1.3 Slope grade factor

From Table 3 we can conclude that the smaller degree of inclination of a slope represents the more population distribution in Guizhou Province. This is accordant to the common principle of population distribution in vertical direction (ZHANG *et al.*, 1996).

Table 3 The weight coefficient and standardized weight coefficient of each slope grade to total population

Slope grade (degree)	Weight coefficient	Standardized weight coefficient
0 – 13	0.738	0.243
13 – 17	0.546	0.180
17 – 19	0.442	0.145
19 – 21	0.379	0.125
21 – 23	0.329	0.108
23 – 25	0.304	0.100
25 – 28	0.301	0.099

3. 1. 4 *Slope aspect factor*

From Table 4 we can conclude that in these slope aspect belts exposed to the sun have the most population distribution in Guizhou Province. This is accordant to the common law of population distribution in vertical direction (ZHANG *et al.*, 1996).

Table 4 The weight coefficient and standardized weight coefficient of each slope aspect to total population

Slope aspect belt	Weight coefficient	Standardized weight coefficient
Slope in the shade	0. 170	0. 148
Slope half exposed to the sun	0. 186	0. 163
Slope exposed to the sun	0. 788	0. 689

3. 1. 5 *Land area factor*

The relationship between land area and total population can be examined from Fig. 2. In Fig. 2, the vertical axis represents total population, and the horizontal axis represents land area in the same county in 2000. In the example, the positive correlation between the two factors in each county is obvious. The computed r equals 0. 691, which indicates a significant positive correlation between the two variables. So the land area factor should be put into the correlation factors.

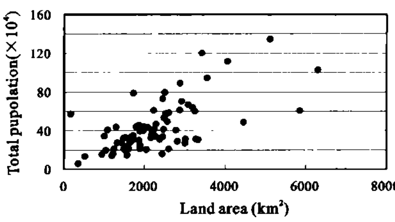

Fig. 2 The relationship between land area and total population

3. 2 Factor Analysis in Sort

Here, in order to evaluate the correlation of each sort factor to total population reasonably, we create a new variable corresponding to each sort factor by the following equation:

$$Y_i = \sum_{i=1}^{5} \sum_{j=1}^{83} a_{ij} X_{ij} \qquad (1)$$

Hereinto, X_{ij} denotes a factor value from a certain sort, and a_{ij} denotes the standardized weight coefficient of the corresponding factor. Then, we can get the weight coefficient of each sort factor by calculating correlative coefficient of Y_i and total population.

In this paper, the population distribution factors acquired from Guizhou Province are sorted as landform factor, altitute belt factor, slope grade factor, slope aspect factor, and land area factor. Table 5 shows the weight coefficients and standardized weight coefficients of the all five sorts of factors.

Table 5 The weight coefficient and standardized weight coefficient of five sort correlation factors to total population

Factors name	Weight coefficient	Standardized weight coefficient
Landform factor	0. 685	0. 194
Altitude belt factor	0. 771	0. 218
Slope grade belt factor	0. 753	0. 213
Slope aspect belt factor	0. 632	0. 179
Land area factor	0. 691	0. 196

We can see from Table 5 that all of the five sorts of factors have certain correlation with population distribution.

3. 3 Establishment of Unit Population Database

In this paper, all parameters in every grid (1km × 1km) covering the whole Guizhou Province are acquired with GIS technology. They are expressed as the following:

$X_{11}, X_{12}, \ldots, X_{1P}$ $(P=4)$ Landform factor
$X_{21}, X_{22}, \ldots, X_{2P}$ $(P=11)$ Altitude factor
$X_{31}, X_{32}, \ldots, X_{3P}$ $(P=7)$ Slope grade factor
$X_{41}, X_{42}, \ldots, X_{4P}$ $(P=3)$ Slope aspect factor
$X_{51}, X_{52}, \ldots, X_{5P}$ $(P=1)$ Land area factor

If standardized weight coefficient for each parameter to population size expressed as b_{ij} (the range of i and j equal to X), and standardized weight coefficient of each sort factor expressed as a_1, a_2, a_3, a_4, a_5, then the final weight coefficient for each correlative parameter is $a_i b_{ij}$. The weight coefficient for each grid to population distribution is c_{mn}, and

$$C_{mn} = \frac{a_i b_{ij} X_{ij}}{\sum_{i=1}^{5} \sum_{j=1}^{p} a_i b_{ij} X_{ij}} \qquad (2)$$

where m is the sequence number for each county in Guizhou Province; n, the sequence number for each grid in a county.

4 DISCUSSION AND CONCLUSION

In order to analyze the correlation between geographical parameters and population size in a region, in this paper the content of geographical parameter database is extended from landform factor to landform factor, altitude factor, slope grade factor and land area factor.

Based on our analysis on geographical parameter database and population size in the year of 2000 in Guizhou Province of China, we can conclude that there is significant linear correlation between population size and some geographical parameters. In fact, almost all of the geographical parameters in our database are correlative to population distribution in Guizhou Province. Of course, each parameter has different correlation coefficient from other parameters, so it has its own weight coefficient to population distribution. And based on this principle, unit population database is established in Guizhou Province.

We have to say that the weight coefficient of each parameter to population distribution in this paper is not the same with other regions.

Both geographical parameter database and unit population database are able to offer sufficient conditions for quantitative analysis. They will have important role in the research fields of data mining (DM) and Decision-making Support Systems (DSS) (TAYLOR *et al.*, 1999).

REFERENCES

CHOU Y H, 1997. *Exploring Spatial Analysis in Geographic Information Systems*[M]. Santa Fe: Onword Press. 1 - 210.

DONG Chun, 2000. *A Research for the Establishment and Application of Geographical Factors Databases*[D]. Beijing: Chinese Academy of Surveying and Mapping. (in Chinese)

DONG Chun *et al.*, 2001. The quantitative study of geographical elements—establishment of geographical factors database[A]. In: *Proceeding of the 20th International Cartographic Conference*[C]. Beijing: Publishing House of Surveying and Mapping. 1197 - 1207.

DONG Chun, ZHANG Qing-pu *et al.*, 2000a. A discussion on the establishment and application of geographical factors databases [J]. *Remote Sensing Information*, 57(1): 12 - 16. (in Chinese)

DONG Chun, WU Xi-zhi, CHEN Bo, 2000b. Study on application of PLS in correlation analysis between geographical parameters and economical development parameters[J]. *Science of Surveying and Mapping*, 25(4): 48 - 51. (in Chinese)

HU Huan-yong, 1990. The distribution, regionalization and prospect of population in China[J]. *Acta Geographica Sinica.* 45(2): 139 - 145. (in Chinese)

JENNIFER B Hughes, GRETCHEN Daily, PAUL R Ehrlich, 1997. Population diversity: its extent and extinction[J]. *Science*, 278(5338): 689 - 692.

TAYLOR K, WALKER G *et al.*, 1999. A framework for model integration spatial decision support system[J]. *International Journal of Geographical Information System*, 13(6): 533 - 555.

WANG Gui-xin, 1997. *Population Distribution and Regional Economic Development in China —A New Approach of Population Distribution Economics*[M]. Shanghai: East China Normal University Publishing House, 1 - 104. (in Chinese)

ZHANG Qing-pu, 2001. Government GIS and its application for decision support[A]. In: *Proceeding of the 20th International Cartographic Conference*[C]. Beijing: Publishing House of Surveying and Mapping, 1386 - 1391.

ZHANG Shan-yu *et al.*, 1996. *The Rule of Population Distribution in Uprightness and Research on Reasonable Population Redistribution in Mountainous Area in China*[M]. Shanghai: East China Normal University Publishing House, 8 - 26, 162 - 167. (in Chinese)