APPLICATION OF REMOTE SENSING TECHNOLOGY TO POPULATION ESTIMATION

ZHANG Bao-guang

(*College of Urban and Environmental Science, Tianjin Normal University, Tianjin 300074, P. R. China*)

ABSTRACT: This paper attempts to explore a new avenue of urban small-regional population estimation by remote sensing technology, creatively and comprehensively for the first time using a residence count method, area (density) method and model method, incorporating the application experience of American scholars in the light of the state of our country. Firstly, the author proposes theoretical basis for population estimation by remote sensing, on the basis of analysing and evaluating the history and state quo of application of methods of population estimation by remote sensing. Secondly, two original types of mathematical models of population estimation are developed on the basis of remote sensing data, taking Tianjin City as an example. By both of the mathematical models the regional population may be estimated from remote sensing variable values with high accuracy. The number of the independent variables in the latter model is somewhat smaller and the collection of remote sensing data is somewhat easier, but the deviation is a little larger. Finally, some viewpoints on the principled problems about the practical application of remote sensing to population estimation are put forward.

KEY WORDS: remote sensing technology population estimation mathematical model

Remote sensing is a new, comprehensive detection technology. In recent years remote sensing technology has been introduced not only into the domain of natural sciences as an effective means for resources investigation, environmental monitoring and regional development, but also into the domain of the social sciences including the realm of demographic research.

The demographic factor plays an increasingly important role in the development of modern society therefore demographic research is becoming one of the main research directions of social science. In particular, the collection and analysis of demographic information, and monitoring and understanding demographic dynamics are vital tasks with which demographic circles and the geographer, urban planner and manager are confronted.

At present the collection of population data depends mainly on the census. However, the census creates a great drain on manpower, financial resources and time. Moreover, between the two sets of census data there is no way of monitoring the urban changes.

After several years the census data will be out of date and their value will be impaired. One of the most complicated problems which faces the urbanologists is the lack of timely and reliable quantitative information on population.

Remote sensing technology has opened a new avenue of urban population estimation and provided a new means of researching the dynamic changes of population.

1 PRINCIPLES AND METHODS

So far the collection of remote sensing data depends mainly on the photographic sensor systems(e.g. black-and-white, color and false-color photographs) and non-photographic sensor systems (e.g. radar and thermal IR images). There are generally two kinds of information on the remote sensing photographs and images. One is directly visible, i.e. the phenomena that are visible on photographs and images in reality; the other is indirectly visible, i.e. the phenomena that are

not visible on the photographs and images, but inferable from other visible phenomena. Being classified with indirectly visible remote sensing information, the population is invisible on the remote sensing images at all, but appreciable by the aid of measuring visible variables, e.g. the number of residential building and the area of built-up zone.

There are three methods of population estimation by remote sensing: residence count method, area (density) method and model method.

The essence of the residence count method lies in identifying and classifying the residential units by using remote sensing images, and calculating the number of each type of residential units. The number of people for each type of residential units can be acquired by sampling different types of residential units or statistical investigation. The estimated population can be obtained by summing up the products of number of residential units for each type and its population coefficient or calculated by the formula (1):

$$P = \sum_{i=1}^{n} (H_i D_i) \tag{1}$$

where P is the estimated population; H_i and D_i are the number of residential units and corresponding number of people in a family for different types respectively, and n is the number of types of residential units.

The essence of area (density) method lies in identifying and classifying the residential land, and calculating the area of residential land for each type, population density which may be obtained in accordance with vital statistics or on-the-spot investigation. The estimated population can be obtained by summing up the products of area of residential land for each type and its population coefficient, or calculated by the formula (2):

$$P = \sum_{i=1}^{n} (A_i D_i) \tag{2}$$

where P is the estimated population; A_i and D_i are the area of residential land for each type and corresponding population density respectively, and n is the number of types of residential land.

The essence of the model method lies in measuring the variations on remote sensing images in the area of built-up zone, inter-town transport line and distance to nearest neighboring large city etc. And a model of population estimation is developed by using regression analysis. The model in its general form can be represented by the formula (3):

$$y = (x_1, x_2, \ldots, x_n) \tag{3}$$

where y is the amount of population and $x_1, x_2, \ldots, x_n$ are the variables of remote sensing indicators taken from remote sensing images.

2 HISTORY AND STATUS QUO

The researches on population estimation by remote sensing technology dated from the early 1970s and were mostly concentrated on the western urban environment. The residential buildings in western cities are variform and characterized by their strip distribution.

A research into population estimation on the basis of remote sensing data has made by an American scholar comprehensively using a residence count method and model method and taking Washington D. C. as an example (HORTON, 1974). In that research on the basis of 1:50 000 scale aerial photographs and vital statistics the regression models of population estimation for Washington proper and its suburbs are developed respectively as following:

For the city proper

$$y_1 = 1522.18 + 6.78x_1 + 83.86x_3 + 18.48x_4 - 790.49x_5 + 19.20x_2 \tag{4}$$

For the suburbs

$$y_2 = 1654.5 + 192.46x_2 + 18.12x_3 + 2.33x_1 + 23.62x_4 + 14.71x_5 \tag{5}$$

where y_1 and y_2 are the numbers of people in the city proper and its suburbs respectively and x_1, x_2, x_3, x_4 are respectively the numbers of entrances to one-, over 15-, 6–14-, 2–5- storey buildings and the distance to the central commercial district of the city.

Some attempts to do population estimation on the basis of remote sensing data have been made by some other American scholars using a residence count method and area (density) method for Birmingham, Chicago, Boston, nearby regions of Atlanta in Georgia and four cities in California respectively (BARRETT and CURTIS, 1986). The results of these researches indicate that the color IR photography is an effective means of estimating the amount of residence units and area of residence lands. Another interesting research into urban population estimation for China on the basis of Landsat information was conducted by an American scholar through combining area (density) method with model method in the light of Chinese cities which are variform and characterized by multicentric development. In this research on the principle of population growth decided by growth of residence area holding the population and in accordance with the built-up zone's areas of thirteen cities in China taken from Landsat imagery and vital statistics, a regression model of urban population estimation for China have been

created as following:

$$\lg P = 5.3304 + 0.4137 \lg A \qquad (5)$$

where P is the number of people in a city and A is the area of built-up zone.

Besides, an attempt to do rural population estimation has been made by an American scholar using residence count method on the basis of 1:25 000 scale aerial photographs and vital statistics for Xinjie rural district of Hong Kong. This research is of certain significance for solving rural population estimation under different cropping systems and economic conditions.

Oversea researches indicate that large and middle scale aerial photographs are usually used for population estimation on the basis of remote sensing data. The error of population estimation by remote sensing technology is 3.0%−4.5%. The scale of aerial photograph is closely related with the accuracy of population estimation by remote sensing technology. The estimation results being obviously on the low side are of frequent occurrence.

3 NEW METHOD

Today the application of remote sensing to population estimation models is still in the exploratory stage. The author of this paper has conducted some researches on this problem and gained some knowledge. What follows is an attempt to explore a new method of urban small-regional population estimation by remote sensing technology, comprehensively using a residence count method, area (density) method and model method, taking Tianjin City as an example.

In its general form the mathematical model, which intended to estimate the small regional population parameters with variables taken from remote sensing images, can be represented by the following equation:

$$y = f(x_1, x_2, \ldots, x_n) \qquad (6)$$

where y is the amount of population ; $x_1, x_2, \ldots, x_n$ are variables taken from remote sensing images, respectively, amounts of 6-, 5-, 4-, 3-, and 2-storey residential buildings, the distribution area of one-storey residential buildings and the distance to the central commercial district of the city.

IR color photographs on the scale of 1:25000 and acquired in 1989 are the basic remote sensing information. IR color film is more sensitive to the electromagnetic radiation in the longer wavelength bands as compared with normal color film and possesses higher haze-penetration capability. On it there is a higher contrast ratio of the cultural to natural characteristics. All of this is especially valuable for the estimation of urban residential buildings.

In order to determine the number of entrances to the residential buildings from remote sensing images, first of all, it is necessary to ascertain different types of buildings by their sizes and to identify the residential buildings of each subdistrict. It must be pointed out that, on IR color aerial photographs, the color signatures of subject are determined by the spectral reflectance properties and color composition and contrast with their original visual color signatures.

The process of image interpretation and data collection is as following. First the boundary lines of 72 subdistricts are drawn on IR color photographs in accordance with relevant material. Within the limits of each subdistrict the types of different residential buildings are recognized by computer automatic classification program in combination with visual interpretation and on-the-spot inspection, and the numbers of entrances to residential buildings with different storeys, the distribution areas of one-storey houses and the distance from the location of concentrated residential areas to the central commercial district of the city are then calculated by computer automatic polygon program. The remote sensing data acquired from color IR photographs are all given in Table 1.

In order to ascertain the relationship between the variables of remote sensing indicators and the regional population, the subdistricts of the city are taken as observation units; the independent variables are interpreted as variables of indicators of population taken from remote sensing. The dependent variable are get from data of population census of China in 1990 (Tianjin Statistical Bureau, 1991). The multiple regression analysis is used for studying the relationship of them.

The analysis of population estimation includes two regressions. In the former case, the dependent variable is interpreted as the number of people in each subdistrict of Tianjin City in 1990, the independent variables include following remote sensing data: the number of the entrances to 6-storey buildings (x_1), the number of entrances to 5-storey buildings (x_2), the number of entrances to 4-storey buildings (x_3), the number of entrances to 3-storey buildings (x_4), the number of entrances to 2-storey (x_5), the area of one-storey houses (x_6), and the distance to the central commercial district of the city (x_7). In latter case, the dependent variable is still interpreted as the number of people in each subdistrict of Tianjin in 1990, but the independent variables as following remote sensing data: the number of entrances to 5- and 6-storey buildings (x_1), the number of entrances to 4-storey buildings (x_2), the number of

Table 1 Population and its variables of remote sensing indicators in subdistricts of Tianjin City

Items				Subdistricts									
				(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	...	(72)
				For the former case of regression									
Variables of remote sensing indicators	Number of entrances to buildings	6-storey	x_1	48	11	–	–	68	–	28	58		4
		5-storey	x_2	132	9	8	–	141	11	31	194		24
		4-storey	x_3	202	64	2	–	78	6	135	68		12
		3-storey	x_4	144	43	7	6	97	13	33	88		117
		2-storey	x_5	12	5	14	14	102	10	26	–		470
	Area of one-storey houses (ha)		x_6	18.75	33.94	31.63	36.50	18.31	22.69	24.56	1.06		1.44
	Distance to commercial center (km)		x_7	3.40	1.73	2.80	2.45	5.95	2.93	4.85	5.03		1.13
Population($\times10^3$)			Y	53.24	41.27	30.68	34.56	35.99	28.98	44.23	25.63		34.46
				For the latter case of regression									
Variables of remote sensing indicators	Number of entrances to buildings	5–6-storey	x_1	180	20	8	–	209	11	59	252		28
		4-storey	x_2	202	64	2	–	78	6	135	68		12
		2–3-storey	x_3	156	48	21	20	199	23	59	88		587
	Area of one-storey houses (ha)		x_4	18.75	33.94	31.63	36.50	18.31	22.69	24.56	1.06		1.44
	Distance to commercial center (km)		x_5	3.40	1.73	2.80	2.45	5.95	2.93	4.85	5.03		1.13
Population ($\times10^3$)			Y	53.24	41.27	30.68	34.56	35.99	28.98	44.23	25.63		34.46

entrances to 2- and 3-storey buildings (x_3), the area of one-storey houses (x_4), and the distance to the central commercial district—the Quanye bazaar (x_5). As a result of the above-mentioned two regressions the structure forms of the regression equations are obtained with a microcomputer, respectively, as follows:

For the former case:

$$y=15.8732+0.5446x_6+0.1008x_3+0.0552x_4+$$
$$0.0219x_5+0.0329x_2-0.9022x_7-0.0261x_1 \qquad (7)$$
$$R^2=0.94$$

where y is the number of people in each subdistrict of Tianjin, x_1 is the number of entrances to 6-storey buildings (imagery), x_2 is the number of entrances to 5-storey buildings (imagery), x_3 is the number of entrances to 4-storey buildings (imagery), x_4 is the number of entrances to 3-storey buildings (imagery), x_5 is the number of entrances to 2-storey buildings (imagery), x_6 is the area of one-storey houses (imagery), x_7 is the distance to the central commercial district.

For the latter case:

$$y=14.07568+0.5554x_4+0.0328x_3+0.1158x_2+$$
$$0.0250x_1-0.7090x_5 \qquad (8)$$
$$R^2=0.94$$

where y is the number of the people in each subdistrict of Tianjin, x_1 is the number of entrances to 5- and 6-storey buildings (imagery), x_2 is the number of entrances to 4-storey buildings (imagery), x_3 is the number of entrances to 2- and 3-storey buildings (imagery), x_4 is the area of one-storey houses (imagery), x_5 is the distance to the central commercial district.

First of all, the results of implementing those regressions demonstrate the importance of independence variables on population changes in subdistricts of the city. In the former regression equation the area of one-storey houses (x_6) is the most important factor influencing the population changes in subdistricts of the city. As the variable x_6 changes by one unit, the number of population changes by 0.5446 units. Thus, it's obvious that the importance of the residential building of lower storeys to population changes in subdistricts of the city is greater than that of higher storeys. The variables, indicators of population, are directly proportional to the dependent variable. In cases where a type (storey level) of building makes up less than 50% the total, the former is inversely proportional to the latter, e.g. the variable x_1, the number of the entrances to 6-storey buildings, and the variable x_7, the distance from each subdistrict to the central commercial district are inversely proportional to the numbers of the population in each subdistrict of the city. As the variable x_7 increases by one unit, the number of population decreases by 0.9022 units.

The results of the latter regression analysis are similar to those of the former one. Therefore, the buildings of lower storeys are still better indicators of population in comparison with the buildings with higher storeys.

The results of the calculation of regression effect indexes and variance analysis on both of the above-mentioned regressions indicate that the influence of the variables x_1-x_7 in the former regression and the variable x_1-x_5 in the latter one on the dependant variable y has reached very significance level. Because of using remote sensing data of 72 subdistricts (it is ten times as much as the number of independent variables)

in the equations of population estimation of Tianjin City, the sample size is quite large. The deviation of the estimated population (with the probability of 95%) from the real one is ±5.95 thousand in the former regression, ±6.60 thousand in the latter regression.

From the above analysis it is concluded that both of the regression equations have really reflected the objective relation of remote sensing variables to the population. By both of the mathematical models the regional population may be estimated from remote sensing variable values with high accuracy. The number of the independent variables involved in the latter regression equation is somewhat smaller and the collection of remote sensing data is somewhat easier, but the deviation is a little larger.

The facts show that it is entirely possible and extremely effective to estimate the regional population of the city by remote sensing technology, so long as the aerial remote sensing photography exists. The regional population may be quickly estimated with high accuracy and the population dynamic information may be obtained according to the mathematical models of population estimation proposed in this paper.

4 CONCLUSION

In summary, it may be seen that the valuable population dynamic information in most regions, both in cities and in the countryside, may be attained by remote sensing data. The application of aerial remote sensing information to the estimation of the entrances to residential buildings and population is very effective, especially in cities and regions with a high rate of population increase.

Until today most results of application of remote sensing to demographic research are encouraging.

However, because there are a few studies in this respect, a series of problems are as yet unsolved (ZHANG, 1987), such as selection of scales of remote sensing imagery, reduction of influence of plant cover on remote sensing data, stability of estimation relations, correction of building count, etc. Among them, in the author′s opinion (ZHANG, 1991), more attention should be paid to the following two respects, the accuracy (error) of building count and the stability of the correlation between population and remote sensing indicator variables. If a clear understanding of these two problems may be made, a small regional model of population changes can be created by remote sensing information and on this basis even the effect of demographic policy can be estimated.

All existing analyses indicate that it is entirely possible to develop population change models by remote sensing technology and the application of remote sensing to the demographic study and urban planning has vast prospects.

REFERENCES

BARRETT E, CURTIS L, 1986. *Introduction to Environmental Remote Sensing* [M]. London: Chapman and Hall, 238−241.

HORTON F, 1974. Remote sensing techniques and urban acquisition [A]. In: ESTES J, SENGER L (eds.). *Remote Sensing Techniques for Environmental Analysis* [C]. Santa Barbara: Hamilton Publishing Co., 243−276.

Tianjin Statistical Bureau, 1991. *Tianjin Statistical Yearbook* [M]. Beijing: China Statistics Press, 157−179. (in Chinese)

ZHANG Bao-guang, 1987. Urban population estimation by remote sensing [J]. *Population and Family*, 15(6): 13−15. (in Chinese)

ZHANG Bao-guang, 1991. *Fundamentals of Geographical Forecasting* [M]. Beijing: High Education Press, 155−160. (in Chinese)