Research on Population Distribution Model Based on Communication Base Station—Take Shanxi Province as An Example

Bin Yang¹²*

¹ Seismological Bureau of Shanxi Province, Taiyuan 030021, China
² Taiyuan Continental Rift Dynamics National Field Scientific Observation and Research Station, Taiyuan 030025, China
Email: ybsuning@163.com

Abstract. The goal of earthquake prevention and disaster reduction is to scientifically manage the risk of earthquake disasters in order to reduce the impact of earthquake disasters on social economy. The main goal is to reduce casualties. Accurate information of population distribution is an important prerequisite to reduce the impact of natural disasters on population. At present, the population data in the provincial earthquake emergency database are mainly statistical data, which can be divided into population data based on administrative division and population data based on kilometer grid. The population grid data is closer to the actual distribution of the population than the traditional census statistics, but it is also mainly based on static data. Neither can reflect the changes in the distribution of people during the day, nor can it reflect the impact of holidays on the flow of people, so it is very important to study the dynamic population distribution data. This paper analyzes the main methods of the existing population distribution model, compares the advantages and disadvantages of each model, and points out the limitations of the static population distribution model. At the same time, based on the communication base station and based on the population distribution model of the residential area, Shanxi Province’s work day/holiday population distribution model was established. And through data comparison and analysis, the feasibility of this method is verified.

1. Introduction
With the rapid development of social economy, the damage caused by destructive earthquakes is increasing. The goal of earthquake prevention and disaster reduction is to scientifically manage the risk of earthquake disasters in order to reduce the impact of earthquake disasters on social economy. The main goal is to reduce casualties. Accurate information of population distribution is an important prerequisite to reduce the impact of natural disasters on population. In daily preparation, accurately grasping the population distribution can provide a basis for decision-making for the establishment of emergency shelters and the planning of emergency disaster relief material reserves. In disaster response, accurately knowing the population distribution can quickly assess the number of casualties, the location of people buried after the earthquake, the key areas of rescue, the dispatching demand of rescue forces, the distribution of relief materials, etc., providing reliable basis for the government departments to carry out emergency response in time.

The existing population distribution data is social statistical data, which can be divided into population data based on administrative divisions and population data based on kilometer grids. At
present, the population data in the basic earthquake emergency database of each province of the country are of these two types. The population gridded data is closer to the actual distribution of the population than the traditional census statistics, but it is also mainly based on static data. Neither can reflect the changes in the distribution of people during the day, nor can it reflect the impact of holidays on the flow of people, so it is very important to study the dynamic population distribution data.

2. Establishment of Population Distribution Model
Exisiting population distribution models are mainly divided into average distribution method, grid interpolation method, population distribution factor analysis method, etc. [1-3]. Through the study of the existing population distribution model, comparing the characteristics, advantages and disadvantages of each model, it can be seen that the average distribution method is easy to operate, but the results often do not conform to the real situation, and the error is large; Grid interpolation can effectively reduce the workload, but the problem is that the results of different interpolation methods are different, whether it is in line with the actual population distribution needs to be further verified; Using the method of population distribution law, the analysis of urban population is more in line with the actual distribution, but for the suburbs and rural areas, there is a big error with the actual distribution of population. The factor analysis method of population distribution needs to assign weight, which is greatly influenced by human subjective factors [4-9].

The traditional population distribution model is mainly static population distribution model, which cannot reflect the characteristics of population flow. Dynamic population distribution model is the trend of population model development. The penetration rate of mobile phone in China has exceeded 100 per 100 people. The distribution and dynamic changes of mobile phone users are very close to the actual population distribution and dynamic changes in statistical sense [10-12]. It is most valuable and representative to analyze the dynamic distribution of population in time and space by using mobile phone positioning data. The population model based on mobile communication is currently dominated by the population heat map, but the population positioning data of the population heat map is based on the data of the intelligent application APP. The data automatically recorded by the intelligent APP is affected by the popularity of the APP and cannot be fully characterized Population Distribution; The method based on user's sharing location data is limited by user's will, not to mention the population distribution [13-15]. The data based on the communication base station avoids the above restrictions. The signal covers more than 3G, 4G and 5G, which can truly reflect the population distribution. In this paper, the population distribution model will be studied based on the big data of the number of users connected to the communication base station and the data of the residential area.

The spatiotemporal behavior of the population has similarities: on working days and during the day (8:00 a.m. to 8:00 p.m.), people will have diverse behaviors such as commuting and commuting to school, at night (8:00 p.m. to 8:00 a.m.) People mainly focus on home behavior; On rest days or holidays, except for travel, people are mainly home-based behavior day and night, which is similar to the night of working days. Different behaviors of people correspond to different urban functional areas, and residential areas reflect the spatial attributes of people well, and can be used as an important basis for population spatial classification [16-20]. Based on the above principles, this article determines the following research methods:

1) Using communication base stations as the modeling medium, the key is to find the matching relationship of “population-weekdays/holidays-communication base station user connection data”. First, the user data of the communication base station is divided according to the characteristics of the working day/holiday population behavior, and the “population-working day/holiday” relationship model is constructed to estimate the total population of working days and holidays in the study area. Then, according to the matching relationship between the population and the communication base station, the relationship model between the population number and the communication base station is constructed for working days and holidays.

2) Spatialize the population onto residential plots. The relationship model of “population-weekdays/holidays-communication base station user connection data” realizes the
decomposition of working day/holiday population to different communication base stations, but, the coverage of different base stations is different, so it is too rough to study the spatial distribution of population simply based on base stations. In order to ensure the width and range of mobile communication signals, the base station mostly adopts the cellular cell structure. In order to better reflect the characteristics of the cellular cell covered by the base station, the topic will build a Tyson polygon to represent the influence range of the mobile base station, and introduce the residential area into the population distribution model, so as to realize the spatial distribution of the population of different base stations and different cellular areas to the residential land.

(3) The spatial distribution of population in different periods is a ternary structure of time, space (base station distribution) and attribute (population). It needs to integrate user connection data, base station location data and residential area data through scale conversion and spatial topology calculation. According to the law of people’s life, combined with the actual situation of Shanxi Province, the working day is divided into four time periods to simulate the distribution of population flow, namely 8:00 a.m., 15:00 a.m., 20:00 p.m. and 3:00 a.m. Afterwards, the “population-working day/holiday-communication base station” model construction can be carried out.

\[ P_j = \sum_{i=1}^{n} \left( \frac{S_{ij}}{S_i} \times P_i \right) \]  

(1)

\( P_j \) population in the jth statistical unit; \( S_{ij} \) Area weight of the i-th Tyson polygon in the j-th residential area; \( S_i \) Area weight of the i-th Tyson polygon; \( P_i \) Total number of people in the i-th Tyson polygon site.

\[ P_i = \frac{A_i}{A} \times P \times W_i \]  

(2)

\( A_i \) Area of the i-th Tyson polygon; \( A \) Study area; \( P \) Study area population; \( W_i \) Tyson polygon area weight.

After bringing equation (2) into equation (1), we get equation (3):

\[ P_j = \sum_{i=1}^{n} \left( \frac{S_{ij}}{S_i} \times \frac{A_i}{A} \times P \times W_i \right) \]  

(3)

However, the model fails to fully reflect the population distribution characteristics of the base station cell. Considering that the number of base station user connections is allocated to the Tyson polygon, the working day population distribution model is obtained

\[ P_{ji} = \frac{s_{ji}}{S_{iq}} \times P_{iq} \]  

(4)

\[ P_{iq} = \frac{s_{iq}}{S_q} \times P_q \times W_i \]  

(5)

\[ P_j = \sum_{q=1}^{n} \sum_{i=1}^{n} P_{ji} = \sum_{q=1}^{n} \sum_{i=1}^{n} \left( \frac{s_{ji}}{S_{iq}} \times P_{iq} \right) = \sum_{q=1}^{n} \sum_{i=1}^{n} \left( \frac{s_{ji}}{S_{iq}} \times \frac{s_{iq}}{S_q} \times P_q \times W_i \right) \]  

(6)

\( P_j \) Number of working days in the jth statistical unit; \( P_{ji} \) Number of working days in the i-th residential area in the j-th statistical unit; \( S_{ji} \) Area of i residential area in the jth statistical unit; \( S_{iq} \) Area of the i-th residential area within the q-th Tyson polygon; \( P_{iq} \) Number of people in the i-th settlement within within the qth Tyson polygon; \( S_q \) Area of qth Tyson polygon; \( P_q \) Number of working days for the qth base station; \( W_i \) Weight of the i-th residence.

Similarly, the holiday population distribution model is:

\[ H_j = \sum_{q=1}^{n} \sum_{i=1}^{n} H_{ji} = \sum_{q=1}^{n} \sum_{i=1}^{n} \left( \frac{s_{ji}}{S_{iq}} \times P_{iq} \right) = \sum_{q=1}^{n} \sum_{i=1}^{n} \left( \frac{s_{ji}}{S_{iq}} \times \frac{s_{iq}}{S_q} \times H_q \times W_i \right) \]  

(7)

\( H_j \) Holiday population in the jth statistical unit; \( H_{ji} \) Holiday population in the i-th residential area in the j-th statistical unit; \( H_q \) Holiday population of the qth base station.
The population distribution model at different periods focuses on obtaining the population of each base station at different periods. The population distribution model is as follows:

\[ T_j = \sum_{q=1}^{n} \sum_{i=1}^{n} T_{ji} = \sum_{q=1}^{n} \sum_{i=1}^{n} \left( \frac{S_{ji}}{S_{iq}} \times P_{iq} \right) = \sum_{q=1}^{n} \sum_{i=1}^{n} \left( \frac{S_{ji}}{S_{iq}} \times \frac{S_{iq}}{S_q} \times T_q \times W_i \right) \]  

(8)

\( T_j \) Population in the jth statistical unit at different time periods; \( T_{ji} \) Number of people in the i-th residential area in different time periods in the j-th statistical unit; \( T_q \) Number of people in the qth base station at different times.

3. Comparative Analysis of Data

After the model is established, the model results need to be compared with the existing population data to determine the accuracy and availability of the model results. In this paper, through the comparison of three levels of province, city and village, the population distribution model built is analyzed and studied.

At the provincial level, this paper mainly compares the model results with the night light remote sensing data. It can be seen from the comparison that the results of the base station population distribution model are basically consistent with the night light remote sensing data in the overall trend, but the data is more delicate. In fact, there is still a population distribution in many shadows of the night light. The comparison results are shown in figure 1.

![Figure 1. Comparison of night light remote sensing data and working day population.](image)

At the municipal level, the results of the model will be compared with the statistical data of the Sixth Census and the population data of the kilometer grid (existing data of the basic earthquake emergency database). The population comparison of cities is shown in table 1.

It can be seen from the comparison of the population of 11 cities in the province that in terms of total population in figure 2, the population of km grid is higher than that of Sixth Census, because the population of km grid is based on the data produced by the latest statistical yearbook, the data increased slightly. The total number of population calculations based on the base station is higher than that of the young and middle-aged population of the Sixth Census population, and lower than the number of people in the grid and the statistical population of the Sixth Census population. This is because some of the elderly and children do not use mobile phones in the overall population, and this part of the population is not reflected in the number of base station user connections, and the census
data is all population data. From the perspective of horizontal comparison between cities, the population of Taiyuan on working days is significantly larger than that on holidays, while the population of Jinzhong is more on holidays, which reflects the population adsorption capacity of Taiyuan as a provincial capital city between working days and the population resettlement role of Jinzhong as an area close to Taiyuan.

Table 1. Comparison of population in cities.

|                          | Taiyuan | Datong | Shouzou | Xinzhou | Yangqu | Lvyian | Jinzhong | Changzhi | Jincheng | Linfen | Yuncheng |
|--------------------------|---------|--------|---------|---------|--------|--------|----------|----------|----------|--------|----------|
| Working day population   | 3783    | 2659   | 13156   | 2335    | 11013  | 2802   | 2533     | 2581     | 1898     | 3600   | 40566    |
| (person)                 | 536     | 615    | 73      | 956     | 70     | 457    | 051      | 799      | 460      | 103    | 40       |
| Holiday population       | 3505    | 2537   | 12932   | 2389    | 10973  | 2890   | 2695     | 2571     | 1861     | 3700   | 40816    |
| (person)                 | 200     | 260    | 36      | 642     | 99     | 327    | 393      | 570      | 909      | 666    | 53       |
| Total Population of Sixth | 4201    | 3318   | 17148   | 3067    | 13685  | 3727   | 3249     | 3334     | 2279     | 4316   | 51347    |
| Census Statistics (person)| 592     | 054    | 57      | 503     | 02     | 068    | 425      | 565      | 146      | 610    | 79       |
| Grid population          | 4355    | 3423   | 17660   | 3148    | 14009  | 3857   | 3346     | 3435     | 2319     | 4456   | 53061    |
| (person)                 | 198     | 094    | 44      | 946     | 34     | 172    | 768      | 617      | 983      | 465    | 71       |
| Young and middle-aged    | 3301    | 2491   | 12811   | 2244    | 10441  | 2735   | 2425     | 2520     | 1750     | 3245   | 38599    |
| population of Sixth      | 458     | 232    | 81      | 445     | 27     | 881    | 230      | 821      | 588      | 937    | 26       |
| Census (person)          |         |        |         |         |        |        |          |          |          |        |          |

Figure 2. Comparative analysis of population in cities.

At the village level, this paper adds the actual number of people to the comparative analysis. The comparison of village population is shown in table 2.

From the comparison of village level population in figure 3, it can be seen that the working day population is significantly lower than the household registration number due to factors such as migrant work and schooling. During holidays, the number of people who go out returns, so the number of people on holidays is similar to the number of people who are investigated.
Table 2. Comparison of village population.

|                                | Yingli Village | Nianjiao Village | Dashuitou Village | Maitian Village | Menjiagou Village | Longmen Village | Niuta Village |
|--------------------------------|----------------|------------------|-------------------|-----------------|-------------------|----------------|--------------|
| Working day population (person)| 1862           | 588              | 684               | 285             | 416               | 363            | 356          |
| Holiday population (person)    | 1783           | 1200             | 884               | 596             | 478               | 680            | 566          |
| Household registration number   | 2223           | 2060             | 1198              | 747             | 701               | 965            | 718          |
| Permanent youth population      | 1566           | 801              | 749               | 516             | 479               | 644            | 450          |
| Number of survey                | 1920           | 1364             | 948               | 662             | 533               | 785            | 629          |

Figure 3. A comparative analysis of the village population in Changning town.

4. Conclusion
This paper analyzes the main methods of the existing population distribution model, compares the advantages and disadvantages of each model, and points out the limitations of the static population distribution model. At the same time, a population distribution model based on the big data of communication base station and residential area is proposed. The model uses the communication base station as a medium, and uses Tyson polygons to represent the cellular range covered by the base station, and at the same time intersects with the residential area, and finally realizes the spatialization of the population of different base stations and different cell areas on the residential plot. Finally, the population of different base stations and different cellular areas is spatialized to the residential area, and the temporal, spatial and demographic attributes of the population distribution model are combined to form a dynamic population distribution model. After the model is established, the feasibility of the method is verified by data comparison and analysis.

Acknowledgements
This paper is supported by The Earthquake Science and Technology Spark Plan Project (NO. XH18009); Key projects of earthquake disaster scenario construction in large and medium cities (2016QJGJ05); Seismological Bureau of Shanxi Province Project (NO.SBK-2040); Earthquake Industry Special Scientific Research Projects (NO. 201208018).
References

[1] Liu S H, Deng Y and Hu Z 2010 Research on classification methods and spatial patterns of the regional types of China’s floating population Acta Geographica Sinica 65 (10) 1187-1196.

[2] Mao X, Xu R and Li X 2010 Fine grid dynamic features of population distribution in Shenzhen Acta Geographica Sinica 65 (4) 443-453.

[3] Yang H, Tian Y and Wang B 2011 Simulation of urban population distribution during daytime based on the high resolution of RS and GIS Journal of Anhui Agricultural Sciences 39 (2) 1129-1132.

[4] Kang W 2006 On the Character of Downtown Daytime Population and Its Impacts on Regional Develop: A Case Study in Luwan District Shanghai (East China Normal University).

[5] Collins A and Greaves S 2007 Daytime population tracking for planning and pollution exposure assessment Road & Transport Research Journal 16 (1) 55-68.

[6] McPherson T N and Brown M J 2004 Estimating daytime and nighttime population distributions in US cities for emergency response activities The American Meteorological Society.

[7] Zhao Y 2010 International studies on “daytime population” and its implication for China South China Population 25 (6) 24-31.

[8] Chen K and Xie Y 1981 The research on day-night population distribution and convection in Taipei area The Taipei City Government Research and Development Appraisal Commission.

[9] Meng Bin, Wang Jinfeng 2005 A review on the methodology of scaling with geo-data Acta Geographica Sinica 60 (2) 277-288.

[10] Ye Y, Liu G and Feng X 2006 Presentation of spatial distribution of population and its application Journal of Geo-Information Science 8 (2) 59-65.

[11] Calabrese F, Lorenzo G D and Liu L 2011 Estimating origin-destination flows using mobile phone location data IEEE Pervasive Computing 10 (4) 36-44.

[12] Gao Z, Liu J and Zhuang D 1999 The relations analysis between ecological environmental quality of Chinese land resources and population Journal of Remote Sensing 3 (1) 66-70.

[13] Zhang T and Wang Y 1986 A preliminary study on the relationship between the distributions of population and earthquake in China Earthquake Research in China 2 (3) 82-87.

[14] Tian Y, Chen S and Yue T 2004 Simulation of Chinese population density based on land use Acta Geographica Sinica 59 (2) 283-292.

[15] Kuang W and Du G 2011 Analyzing urban population spatial distribution in Beijing proper Journal of Geo-Information Science 13 (4) 506-512.

[16] Yin J, Zhang G and Shan X 2007 Simulating population density based on TM Images: Taking Beijing as an example Journal of Catastrophology 22 (1) 31-35.

[17] Collins A and Greaves S 2007 Daytime population tracking for planning and pollution exposure assessment Road & Transport Research Journal 16 (1) 55-68.

[18] Zhang Z, Zhou Yand Li Q 2010 An estimation method of dynamic population within an urban local area Journal of Geo-Information Science 12 (4) 503-509.

[19] Guo X 2008 Urban public security and risk governance: An analysis of population migration and social inclusion Urban Problem 60 (1) 6-11.

[20] Jiang D, Yang X H and Wang N 2002 Study on spatial distribution of population based on remote sensing and GIS Advance in Earth Sciences 17 (5) 734-738.