

AN AGENT-BASED POPULATION MODEL FOR CHINA

Zhongxin Chen^{1,2,*}, Einar Holm³, Huajun Tang^{1,2}, Kalle Mäkilä³, Wenjuan Li^{2,3}, Shenghe Liu⁴

¹ Key Laboratory of Resource Remote Sensing and Digital Agriculture, Ministry of Agriculture, Beijing 100081 CHINA

² Institute of Agriculture Resources & regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081 CHINA

³ Kiruna Spatial Modelling Centre, Umeå University, Kiruna SE98128, SWEDEN

⁴ Institute of Geographical Science & Natural Resources Research, Beijing 100101, CHINA

* Corresponding author, Address: Room 322, IARRP, CAAS, 12 Zhongguancun Nan Dajie, Beijing, 100081, P. R. China, Tel/Fax: +86-10-68918684, Email: zxchen@mail.caas.net.cn

Abstract: Human is a crucial factor responsible for local and large scale environmental change and is a key factor in rural management. Accurate simulating and predicting population change and its spatial distribution is of great interests to environmental scientists as well as social scientists. In this paper, we introduced the development of a spatial-explicit agent-based population model for China based on a prototype from SVERIGE model. The model was used to predict the population in China. The validation result shows that agent-based model performances well in population prediction in China.

Keywords: Agent-Based Model, SVERIGE, Population, Prediction, China

1. INTRODUCTION

Our planet has suffered great changes at very strong intensity world wide since industrial revolution. There are a lot of evidences from various observations, investigations and scientific researches showing that the amplitude of global change has been enhanced since the mid-term of last century. The causes for global change are various and very complicated, among which anthropogenic factors contribute a lot (Turner et al 1995).

Accurate simulating and predicting population change and its spatial distribution is of great significance in various fields of research and application including global changing, land use/cover change, rural management and so on (Ziervogel, 2005).

The world population is about 6.6 billion in 2007 (US Census Bureau, 2007). The world population has been doubled in the past 40 years, and is still increasing at a very rapid pace. The rapid population growth is not only a demographic problem. It also concerns the social and economical development, food security, environment and sustainability and so on in specific regions as well as worldwide. China contributes more than 1/5 to the world's population. And the ratio has been quite stable during the past half century. The population in China increases rapidly. In order to alleviate the side-effect brought by the population growth, family planning program has been applied in China since late-1970s. So there are other demographic problems in China, such as age structure shift, the increasing of aged people, etc. The rapid population growth have great impacts on the following issues, such as economic development, food security, the demand of educational resources, population migration and urbanization, land use and land cover changes and other environmental and social issues, etc. (Brown et al., 1999)

Although there have been 5 national demographic censuses in China and there are a lot of population monitoring and survey data, yet the demand of a robust, accurate demographic prediction tools is very strong. The tools can be used to predict the future trend of the population in China, and it can also be used to aid the decision-making of the population-related issues, such as population migration management policy, land use planning, educational resources configuration, etc. To meet the urgent demands, we decided to set up the China population simulation and prediction model by employing individual based micro-simulation model, i.e. agent-based model. (Bankes, 2002 ; Bonabeau, 2002 ; Holm et al., 2002 ; Nuppenau, 2002 ; Thomas et al., 2006 ; Rykiel , 1996 ; Michael et al., 2006)

2. DATA PREPARATION

2.1 The national demographical census data in 1990

The 1990 population Census of China was conducted by the State Statistics Bureau of China. This dataset (1% sampling) was prepared taking villages as sampling unit, and contains a record for each household and supplies variables describing the location, type, and composition of the households. Each household record is followed by a record for each

individual residing in the household. Information on individual includes demographic characteristics, occupation, literacy, ethnicity, and fertility.

2.2 Aggregation data of 2000 national demographical census

The 5th National Demographic Census in China was committed on November 1, 2000. The Aggregation dataset was compiled by Population Census Office of The State Council and National Bureau of Statistics of China, and it was published in 2002.

2.3 County-level demographical database

There are the frequently used demographic indices in 1990 for more than 2000 counties from 30 provinces, autonomous regions and municipalities. These indices include: total population, sex ratio, size of family, population in cities, population in towns, the ratio of illiterate and semi-illiterate among the population over 15, the number of college-educated people per 10,000 people, the number of middle-school-educated people per 10,000 people, population aged between 0-14, population aged 15-49, population aged 50-65, population older than 65.

2.4 Provincial migration database

This database summarized the population migration data from 1949. It included the outputs from the 4th national demographic census in 1990, and some other large-scale sampling survey since 1980s, such as the migration survey in 74 cities in 1986, 1‰ national sampling of population in 1987, the national sampling survey on anti-pregnancy in 1988, the sampling on pregnancy and delivery in 1992, and other related data from police's registering system.

2.5 County-level agro-eco-environmental database

The database is from Information Center, Ministry of Agriculture of China. It contains the statistics for agriculture production in 2523 counties in China. The data is for 1986 to 2001. The data items include population, cropland acreage, acreage and yield for each crop, GDP, etc.

2.6 County-level administration region's map

The map is from Columbia University's Center for International Earth Science Information Network (CIESIN). The original scale for this administrative map is 1: 1 000 000. The ID for each county for each dataset is based on "Codes for the Administrative Divisions of the People's Republic of China" 1990 version (GB 2260-90).

3. MODEL DESIGN AND IMPLEMENTATION

The prototype for the China population model is the Swedish SVERIGE (System for Visualising Economic and Regional Influences in Governing) model, which was developed by Spatial Modelling Centre (SMC), Department of Social and Economic Geography, Umeå University (Holm et al 2002). The developing tools for the China population model include Microsoft Visual C++, other data preparation tools include SPSS, SAS, ESRI ArcGIS, etc.

3.1 Overall Structure of the Model

The China population simulation model is an individual based model. The basic agent in this model is an individual person, but with a household (or a family) as the basic operational unit. There can be one or more individual person(s) in one family. The reason for us choosing household as the basic organizational unit is that the members in a same household usually share a lot of resources such as housing, income, etc. But the household is not static but dynamic. It means that either the members can leave or join a household by migration, death, birth, etc. Even the household itself can migrate, emerge, or disappear. Then each household has an attribute of geographical location. Herein, a county code is assigned to it to represent its geographical location. The household as well as the household members can share the resources and limitations in this geographical region.

3.2 The Attributes of Individual Agent

The attributes for a household and an individual person can be very complex. In this paper, we only consider the critical and essential features for them. The following is the data structure for an individual person in the model.

```
typedef struct chinastruct { // -----
    unsigned FID           :24;
```

```

    unsigned EducationLevel    :3; // 0-7
    unsigned EducationSector   :3;
    unsigned Relat             :2;
    // Relation in the family //1=Head,2=Spouse,0=Child

    unsigned Sex               :1;
    // 0 = male, 1 = female
    unsigned BirthYear         :9;
    // Base year is 1870
    unsigned OutOfLabor        :1;
    unsigned InEducation       :1;
    // In school or university
    unsigned Working           :1; // Working
    unsigned Unemployed        :1;
    // Unemployed
    unsigned MaritalStatus     :2;
    // (0-3) 0=Single, //1=Married,2=Widowed,3=Divorced
    unsigned LARegion          :12;
    //region identifier 1-2500
    unsigned RuralUrban        :1;
    unsigned School            :3;

    unsigned EthnicGroup       :6;
    unsigned ffiller           :26;
} ChinaBuffer;

```

3.3 Parameterization

The demographical attributes of the individuals as well as households are correlated with other demographical parameters and environmental factors. Herein, the China population simulation model is running in a simple way. The parameters for the model are estimated mainly by probability tables and rules. The probability tables are from the census data or other investigation data. The rules are either based on the demographical theories or related studies or investigations. The parameterization for each process is being explained as follows.

(1) Aging

The simulation step for the China population simulation model is 1 year. In the beginning of each simulation step, each active agent in the model will automatically gain one year in age. Age is also a basis to determine the parameters for other processes, such as education, marriage, migration, giving birth to a child, death, etc.

(2) Fertility

The fertility of the women older than 15-years-old and younger than 50-years-old is estimated based on the 1990 census data. The conditional probabilities of the women in each age falling in this age group were given. We considered the following conditions: (a) whether the woman lives in the

rural area or the urban area (city or town), (b) different marital status, and (c) if she has already given birth to a certain number of child(ren).

Also the ethnic groups of the wife and the husband can affect the fertility. We used rules in this regard. If either of the couples is from a minority ethnic group, they can have more than one child. If the couples live in rural area and they have a 4-year-old child, they can have the second child.

(3) Mortality

The mortality module in the China population simulation model is employed to terminate the lives of individual agents. A lot of factors affect the mortality. But in order to keep the model simplicity at the beginning, age, gender, and residential type (either in rural area or in urban area) are considered as the main impact factors to determine the mortality for a specific person in the model. The mortality probability is parameterized based on the 1990 and 2000 census data, and the mortality database of China.

(4) Education

The education is estimated based on simple rules and probability. The rules are as follows: (a) A child begins to go to primary school at 6 or 7 at a probability; (b) He or she will spend 6 years in primary school, after that he or she goes to junior high school at a probability; (c) He or she will spend 3 years in junior high school, after that he or she goes to senior high school at a probability; (c) He or she will spend 3 years in senior high school, after that he or she goes to Junior college or university at a probability. The probabilities at the critical ages are from statistical data.

(5) Marriage

Marriage module is used to set up families in the model. It affects the fertility and the number of families. An agent tries to find a partner to get married after certain age. In this model, age, gender, education level, the residential type, and geographical location are considered as the main factors to determine the probability for the potential partners to get married.

(6) Migration

The migration rates between provinces are used to estimate the probability of migration for a specific agent in the model. The data is from 1990 and 2000 national census data.

(7) Retirement

The retirement model is used to let the agents to quit from the labor markets. It is a rule-based module. The women agents retire at 55, while the men agents retire at 60.

4. RESULTS AND DISCUSSION

After running of the model, we can get the year-by-year population distribution map of China like Fig. 1.

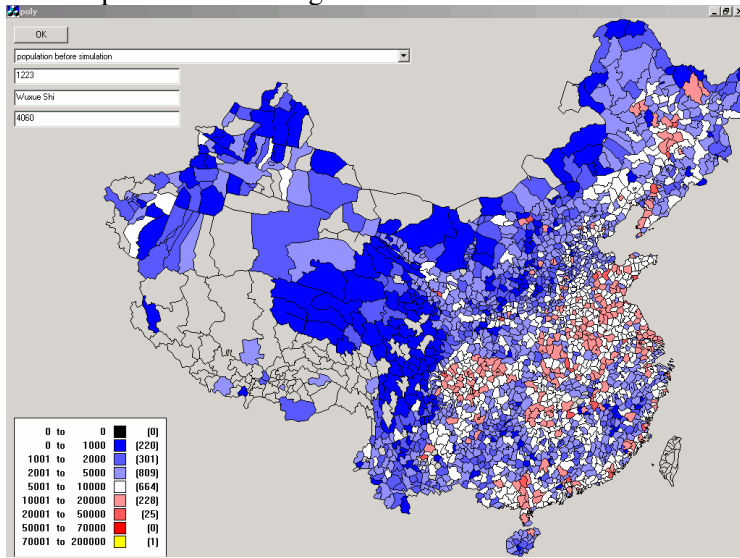


Fig. 1. The population distribution map from the model

The model outputs are evaluated by the existing population datasets during 1990-2001, which includes 2000 census data, provincial population dataset from 1990. The preliminary validation of the total population is listed in the following table. The projected population for China is very close to the statistical figure. The relative errors are less than 0.64%. We can conclude that the model is quite good overall.

Table 1. The validation of the model

| Year | Statistic Population (Million) | Projected Population (Million) | Relative Error (%) |
|------|-----------------------------------|-----------------------------------|-----------------------|
| 1991 | 1158.23 | 1165.70 | 0.64 |
| 1992 | 1171.71 | 1177.78 | 0.52 |
| 1993 | 1185.17 | 1190.04 | 0.41 |
| 1994 | 1198.50 | 1202.48 | 0.33 |
| 1995 | 1211.21 | 1215.14 | 0.32 |
| 1996 | 1223.89 | 1227.67 | 0.31 |
| 1997 | 1236.26 | 1240.14 | 0.31 |

We succeeded in setting-up of the spatial-specific population mode for China with employing agent-based technique. The validation result shows

that it can accurately simulate the population in China. But there are still a lot works to be done. These works include:

1. More accurate estimation of the parameters for the processes in the agent. The multiple regression method could be used to investigate the relationships which control the key demographic factors, such as fertility, mortality, migration and education, as what have been done with the SVERIGE model.
2. Collecting more data for the model. We think that more information is demanding for more accurate estimation of the parameters and rule building.

ACKNOWLEDGEMENTS

The research in this paper is funded by a National Key Technologies R&D Program of China (No. 2006BAD10A06) and a National High Technology Research and Development Program of China (863 Program No. 2006AA12Z103)

REFERENCES

- Bankes, SC, Agent-based Model: Revolution? PNAS, 2002, 99(spp3):7199-7200.
- Bonabeau, E, Agent-based modeling: Methods and techniques for simulating human systems. PNAS, 2002, 99(spp3):7280-7287.
- Brown, LR, GT Gardner, B Halweil, Beyond Malthus. Norton. 167pp. 1999.
- Holm, E., K Holme, K Mäkilä, M Mattsson-Kauppi, G Mötvik, The SVERIGE Micro-simulation Model. Dept. of Soc. & Econo. Geography, Umeå University. 55pp. 2002.
- Michael M, A Miguel, C Baird, Coupled human and natural systems: A multi-agent-based approach. Environmental Modelling & Software, 2006, 1-8.
- Nuppenau, E, Towards a genuine exchange value of nature: interactions between humans and nature in a principal-agent-framework. Ecological Economics, 2002, 43:33-47.
- Rykiel J, Testing ecological models: the meaning of validation. Ecological Modelling, 1996, 90:229-244.
- Thomas B, S Pepijn, W Johannes, Multi-agent simulation for the targeting of development policies in less-favored areas. Agricultural Systems, 2006, 88:28-43.
- Turner II BL, D Skole, S Sanderson, G Fischer, L Fresco, R Leemans, Land-Use and Land-Cover Change (LUCC), Science/Research Plan. IHDP Report 7/IGBP Report 35, Stockholm and Geneva. 1995.
- US Census Bureau, <http://www.census.gov>. 2007.
- Ziervogel, GM Bithell, R Washington, T Downing, Agent-based social simulation: a method for assessing the impact of seasonal climate forecast applications among smallholder farmers. Agricultural Systems, 2005, 83: 1-26.