A GIS-BASED UPDATING SYSTEM FOR GRIDDED POPULATION DATABASE OF CHINA

Huang Yaohuan [a], Yang Xiaohuan[b], Wang Jianhua[a], Zhou Qin [b,c]

[a] China Institute of Water Resources and Hydropower Research, Beijing 100044
huangyh@lreis.ac.cn, wjh@iwhr.com
[b] Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101- yangxh@lreis.ac.cn
[c] Graduate University of Chinese Academy of Sciences. Beijing 100039- zhouq.06b@igsnrr.ac.cn

Commission VI, WG VI/4

KEY WORDS: GIS, Human Settlement, Spatial, Land Use, Modelling, Systems

ABSTRACT:

Spatial population data make many related research more convenient. However, its fussy generation process limited its application of the spatial population data. Spatial distribution of population has close relation with land use and land cover change (LUCC) patterns both at regional and global scales, which can be used to redistribute population onto geo-referenced square grids. Since there exist efficient approaches for monitoring LUCC with remote sensing and GIS, geo-referenced population data can also be updated conveniently. The patterns of LUCC, which are the inputting parameters of the Population Spatialization Model (PSM), is gained from MODIS L1B by using Pattern Decomposition Method (PDM) and LUCC-Conversion Model (LUCC-CM). The 1km×1km gridded population data of 2000 is calculated using the PSM, and the result is reliable validated by finer township census data of case county Yishui. Finally, A Spatial Population Updating System (SPUS) is developed to conduct annually updating of China gridded population database at spatial resolution of 1km by integrating three models.

1. INTRODUCTION

The demographical data is one of the most direct indexes of the human being's activities and influences to the planet earth. The information can be used in many areas such as ecosystems assessment, global environmental change, regional sustainable development studies. Census data sets of administrative or statistical reporting units could not meet the needs of these researches because (1) the rough spatial resolution and (2) too long time for data updating. A new concept of "population spatialization" has been presented in a fruitful workshop on Global Demography in 1994, which means redistributing population onto geo-referenced grids instead of political or administrative units (Jerome E. Dobson, 1998). Data on population dynamics and distribution are the key to the understanding of human interactions with the environment and to the consideration of possible responses to global environmental change (Paul Sutton, 1997).

Since Tobler et al. released the first version of Gridded Population of the World (GPW) in 1995 (Tobler et al, 1995), excellent jobs have been done by many scientists and organizations (for example, Paul Sutton, 1997; Lo C P, 1998). Version 2 of GPW of population estimates and densities for 1990 and 1995 was then developed by the Center for International Earth Science Information Network (CIESIN) of USA and other cooperators in 2000, improved input data and a revised gridding methodology were applied to produce a global grid of the distribution of human population at a resolution of 2.5 arc minutes. In 1996, with financial support from the United Nations Environment Programme (UNEP), Environment Canada has created a population database depicting the worldwide distribution of population for 1990 with a 1°*1° latitude/longitude resolution.

During the Workshop on Gridding Population Data at Columbia University May 2000, Conceptual, methodological and technical issues in the production of the grid population were discussed widely. Uniform density within certain administrative unit can be improved based on more input information such as land cover, DEM, transportation, and so on. Then finer resolution of population can be achieved.

Land cover type is a good indicator of population distribution. Oak Ridge National Laboratory (ORNL) has developed the LandScan Global Population Dataset, which is a worldwide population database at 30" X 30" resolution for estimating ambient populations at risk Census data (population). Land cover data together with information about roads, DEM and Nighttime lights were taken in granted for population simulation.

There is a direct relationship between to land use pattern and population density. Generally speaking, when the population density amounts to 2500 persons per squire kilometer, its corresponding land use always represents as resident areas and towns. When the population density reduced by 35 persons per square kilometer, its corresponding land use represents as the original unused situation (Gao Zhiqiang, et al. 1999).

Data Center for Resources and Environmental Sciences (RESDC) of Chinese Academy of Sciences (CAS) has developed China gridded population dataset of 1995 and 2000 at spatial resolution of 1km in 2002. China gridded population

Huang Yaohuan: China Institute of Water Resources and Hydropower Research, Beijing 100044, Email- huangyh@lreis.ac.cn

dataset addresses residential population rather than ambient populations of Land Scan database, which integrates diurnal movements and collective travel habits into a single measure. This paper focused on the establishment of Spatial Population Updating System (SPUS), which can accomplish the auto-updating of gridded population database annually based on LUCC data derived from MODIS images. Validation and future directions of improvement were also discussed.

2. DATA AND METHODOLOGY

2.1 Data Sources

The data of the research includes Chinese census data, land use data and ancillary data detailed in table 1. According to the differences in data sources and data types, the criterion and precision also vary. Standardization data pre-processing has been done before all the data used in the research including satellite images correction, projection system transformation and attribution data standardization processing.

Data type	time	Data sources	Scale or resolution
Census data	1995-2002	State Bureau of Statistics of China	China at county level
MODIS L1B	2002	EOS website	500m
Statistical social and economic data	2000	Chinese Statistic Yearbooks	County level
Land use	2000	RESDC	1：100000
Land use	2000	RESDC	County level
DEM data	2000	State Bureau of Surveying and Mapping of China	1：250000
ETM+	2000	RESDC	30m
Boundary of county	2000	RESDC	1：100000
Resident map	2000	State Bureau of Surveying and Mapping of China	1：250000
Statistic population of Yishui County	2002	Statistic Yearbook of Yishui County	township level

Table 1. List of the Data Sources

2.2 Methodology

2.2.1 Method of land-use/land-cover (LUCC) data updating

The Chinese land use types were used as primary index in the spatial population model of this paper. Pattern Decomposition Method (PDM) is applied to obtain the land use / land cover patterns with Moderate Resolution Imaging Spectroradiometer (MODIS) data of NASA's Earth Observing System (L1B data of 1-7 bands of 500m resolution). The vegetation, water and soil coefficients are extracted by PDM for each MODIS pixel and composed with LUCC-Conversion Model (LUCC-CM). MODIS images are classified into different land use types based on the relation between the spectral coefficients and land use structure.

2.2.1.1 Pattern Decomposition Method (PDM)

LUCC information can be shown by three standard spectral patterns: vegetation, water and soil (Muramatsu et al., 2000). 7 bands MODIS data could be transferred to 3 dimensional data based on spectral pattern decomposition.

Firstly, the standard spectrum patterns are extracted. Surface albedo of MODIS L1B data within the chose experimental area is been normalized processing to avoid the disturbance of absolute spectrum value. The normalization formula is shown as below.

$$B_i = \frac{A_i}{\sum_{j=1}^{7} A_j} \quad (1)$$

After choosing the sample pixels of pure water, soil and vegetation in test area, a 3 × 7 matrix P is gained by making average of 7 bands' Bi, which represents MODIS data's spectrum pattern of water, vegetation and soil from band 1 to 7. Taking Shandong Province as the example, the standard spectrum patterns shows as table 2.

	Water pattern P_w	vegetation pattern P_v	Soil pattern P_s
Band1	0.305889	0.089826	0.091817
Band2	0.237651	0.083721	0.089176
Band3	0.172095	0.067106	0.097329
Band4	0.114738	0.245933	0.178803
Band5	0.076467	0.248804	0.205277
Band6	0.057306	0.172512	0.191812
Band7	0.035854	0.092098	0.145785

Table 2. Standard spectrum patterns in Shandong Province

Secondly, pixels spectrum decomposition is conducted based on the standard spectrum pattern. Taking each band's albedo as linear combination of reflection of LUCC, use the matrixes P and Bi to simulate water, vegetation and soil component coefficients. The formula is shown as below:

$$A_i = C_w P_{iw} + C_v P_{iv} + C_s P_{is} + R \quad (2)$$

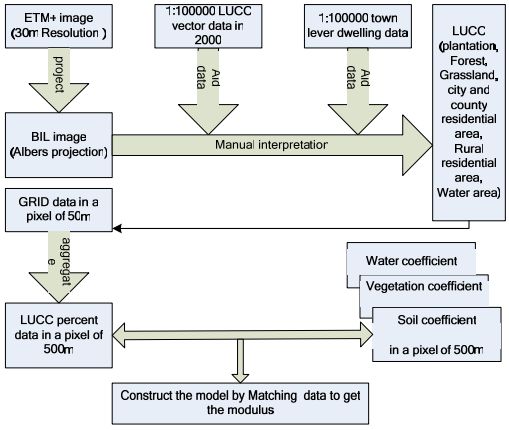
Where Ai is the surface albedo of band i. P_{iw}, P_{iv}, P_{is} are the standard spectrum patterns of water, vegetation and soil. Cw, Cv,Cs indicate the decomposition coefficients, which are positive number and the max value are equal to pure samples. According to PDM principle, following formula can get three types of land cover proportion in each pixel:

$$r_k = \frac{C_k / S_k}{C_w / S_w + C_v / S_v + C_s / S_s}, k = w, v, s \quad (4)$$

Where rw, rv, rs indicate three matrixes expressing water, vegetation and soil proportion in a pixel of 500*500m.

2.2.1.2 LUCC-Conversion Model (LUCC-CM)

The result in decomposition step can not be distinguished directly as land-use type while three proportions in each pixel display certain land use structure. This paper uses the decomposition result and LUCC-CM to derive land use types. The manual interpretation data of test area's land use proportion at 500m is combined with decomposition index into correlation analyses to get the parameters of land use distinguishment. The whole process is shown as Figure 1.



Finger 1. Process of getting the parameters of LUCC

LUCC-CM is built up with the parameters as below.

$$\begin{cases} C_1 \square X_1 = Y \\ Y = [C_2 C_3][r_w r_v r_s]^T \end{cases} \quad (5)$$

Where C1 , C2 , C3 , X1 , Y are matrixes, where X1 = [x1 , x2 , x3 , x4 , x5 , x6], expresses different percentages of cropland, forest, grassland, urban residential, rural residential, and water area in each grid. C1 , C2 , C3 are modulus of model. and constant part. rw , rv , rs represent the facts of decomposition.

2.2.2 Method of population spatialization

The paper adopts the method of population spatialization model based on the relation between statistics demographical data and land use types and redistributes population onto 1km grid (Yang, et al., 2002). The Data Center for Resources and Environmental Sciences of the Chinese Academy of Sciences (RESDC, CAS) has applied this method to build the China gridded population database of 2000.

2.2.2.1 Technical flow of population spatialization

The general steps for redistributing China population are listed as figure 2.

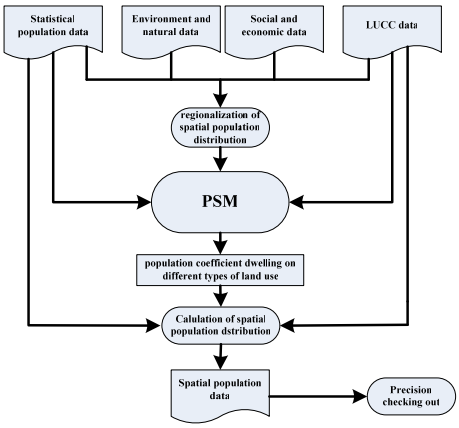


Figure 2. Flow chart for producing China Gridded Population Dataset

2.2.2.2 Regionalization of Spatial Population Distribution

China is a large country with different population density and land cover patterns from west to east. According the Fifth Census of China, the average population densities of the east, middle and west of China are 452.3, 262.2 and 51.4 (person/km2) respectively. To obtain more reliable result, we constructed a three-dimensional feature space based on provincial population density core and regionalization index to divide whole country into 8 regions using minimum distance rule (Hu Huanyong, 1983) (Figure 3). Population models were established within every region.

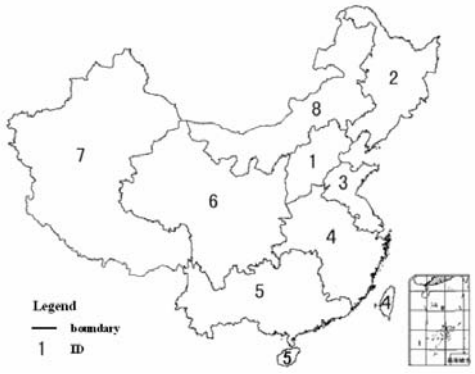


Figure 3. Regionalization Map of Population Spatial Distribution in China

(1) Calculation of spatial population characteristic index

According to population density, economic development, land use structure, transportation network and river density at county level, design the model of spatial population characteristic index (equation 5) and calculate indices for the distribution of population division.

$$Ip = (P + GDP) / S / (1\text{-}Ic)/(1\text{-}Ird)/(1\text{-}Irl)/(1\text{-}Ir) \quad (6)$$

Where Ip is the spatial population characteristic index; P is the total population; GDP is the gross domestic product; S is the

total land area; Ic is index of planting density ; Ird for road density ;Irl is the railway density; Ir is residential density.

(2)Determining the center of provincial population distribution
According to characteristics of a provincial population distribution index (z) and the spatial distribution of population centers (x, y) , construct a three-dimensional space to calculate inter-provincial population distribution distance(d),.Using the minimum distance method divide the first level 8 population regions.

(3) Secondary population regions are divided by the county-level terrain and population density.

2.2.3 Population Spatialization Model(PSM)

The research shown that population distribution is of high degree of correlation with land use / land cover. We built a population spatialization model to redistribute population of a county into different types of land use as formula below:

$$P_i = \sum_{j=1}^{nf} a_i x_{ij} + B_i \quad (7)$$

Where Pi stands for total population of the ith county， aj is the population density of the jth land use type，and xfj is the total area of ith land use type in f section (km2)，nf is the number of land use types existing in this section . According to the rule of "no resident area, no population", the intercept Bi is set to zero.

It is necessary to control total population predicted equal to census statistical data within each administrative unit. Population density of each land use type should be adjusted by the ratio of the predicted population (Pi) and the census count (Pi0). Adjust initial coefficient aj is as below:

$$a_{ij} = \frac{p_i^0}{p_i} a_j \quad (8)$$

Where aij stands for modified population density for the jth land use type within the ith administrative unit;. Pi and Pi0 stand for predicted population and census count of the ith administrative unit respectively.
After steps mentioned above, population can be estimated from cell to cell. We calculate population of each cell by linking population density coefficient aij to land use grid using formula 8 to create spatialized population data of China.

$$P_i = \sum a_{ij} x_j, \; j = 1, 2, ..., n \quad (9)$$

3. SPUS SYSTEM DESIGN AND DEVELOPMENT

SPUS is an operational software integrated PDM, LUCC-CM, PSM. It supports generation of gridded population data automatically, result display, analysis and database management. The system is developed in modularization under the environment of Windows and is easy to use by users. It can update gridded spatial population dataset after inputting census data, MODIS data and other ancillary data necessary.

3.1 Data Processing Flow and Framework

3.1.1 Data Processing Flow

Figure 4 shows the data processing flow of SPUS. The input data includes the statistical population data at county level, MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid, ancillary data such as administrative boundary maps. All the data of the system is stored in the attribute and spatial database.

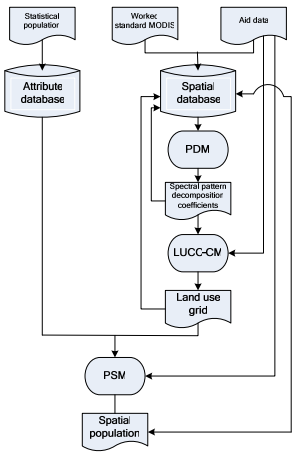


Figure 4. Data process flow of SPU

3.1.2 Framework of SPUS

SPUS is developed to conduct generation, management and display of spatial population dataset. It consists of Presentation Layer, Application Layer, Service Layer and Data Layer (Figure 5).

(1) Presentation Layer: provides the Graphical User Interfaces (GUI). Users achieve the functions such as authorization; display, browse, maintenance and query of system data; RS image processing; land use data processing and statistical population spatializing.

(2) Application Layer: supports inner model processing of the system. The layer provides the practical implement of the users' request from GUI.

(3) Service Layer: it is a middle tier of the system providing the professional models of PDM, LUCC-CM and PSM which can be considered as repository of SPUS. Moreover, the layer including the spatial data engine (ArcSDE) which can be used to accessing the massive amount of geographical data stored in RDBMS.

(4) Data Layer: it consists of all SPUS required database.

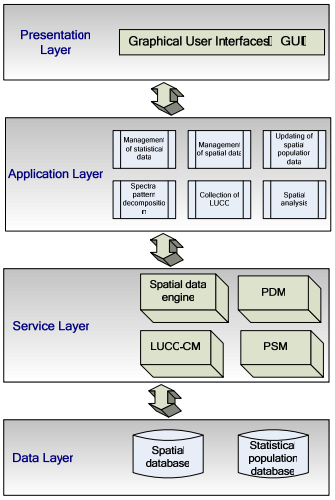


Figure 5. Framework of SPUS

3.2 System Function Development

Figure 6 shows the schematic representation of system function modules which is composed of 6 parts.

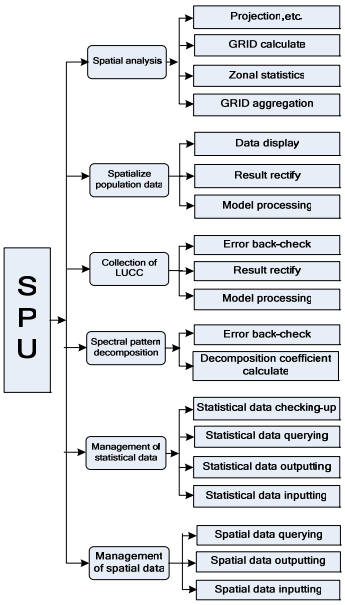


Figure 6. Framework of function modules

3.2.1 Management module of spatial data

The module is a spatial database tool that manages all the SPUS referring spatial data based on ArcSDE. It can be used to import, export and query spatial data including grid database and vector database.

3.2.2 Management module of statistical data

The module manages all kinds of statistical data including population, socio-economic data and ancillary attribution data with Oracle 9i. The functions include data input,. storing, output with common format such as Excel, XML, MDB, DBF etc., data checking and quality control, data querying.

3.2.3 Spectral pattern decomposition module

The module executes the PDM to conduct the processing of standard MODIS data to generate coefficients ASCII file referring to the spectral pattern decomposition using IDL program in the environment of .net. The ArcEngine is used to transform ASCII file to GRID by the ArcGIS spatial analysis technique. The module is a mixed program of .net, IDL and ArcEngine.

3.2.4 LUCC-CM module

It is a functional module that executes LUCC-CM. At first, train the example areas selected according to the regionalization to get the modulus which can be considered repository of SPUS as the basic information of LUCC. Secondly, apply the LUCC-CM to convert the result of PDM to six different land use types defined as cropland, forest, grassland, city, rural residential area and water in a pixel of 500 by 500 meters. Finally, check the data of LUCC and import it into spatial database. Considering the advantage of IDL in scientific computation, LUCC-CM was developed by IDL. The IDL program generates the ASCII file contains the percent of six LUCC types in each pixel, and then applies ArcEngine to create LUCC grids for spatial analysis.

3.2.5 Population data Spatialization module

The module is the core of SPUS. It executes the PSM to redistribute the statistical population data on grid by combining the LUCC grid, province vector data, county vector data and other ancillary data. It is a GIS system developed with ArcEngine consisting of 3 functions.

Town name	2002 statistical data	calculated population	residual	relative error
Dao tuo	39360	36831	-2529	-6.43%
Xia wei	54898	52967	-1931	-3.52%
Gao zhuang	51507	50914	-593	-1.15%
Long jiajuan	61231	47691	-13540	-22.11%
Cu ijiayu	33380	39101	5721	17.14%
Huang shanpu	50252	37842	-12410	-24.70%
Xu jiahu	81431	76713	-4718	-5.79%
Yuan dongtou	29191	36185	6994	23.96%
Sishi lipu	66312	61245	-5067	-7.64%
Zhao dianzi	44766	40533	-4233	-9.46%

Table3 Validation of town-lever population in Yishui County

(1) Data display. The main interface can add and display the spatial data and provide the functions of querying, panning, zooming in, zooming out, selection and saving.

(2) Model processing. Create spatial population grid based on PSM.

(3) Results verification. Adjust the primary results according to total population control within county level unit to generate the final result of gridded spatial population dataset.

3.2.6 Spatial analyses module

SPUS provides some common functions of GIS spatial analysis such as grid calculation, projection transformation, aggregate, buffer, overlay (union, intersect, erase), zonal, neighborhood and so on. Users can combine these functions easily to obtain some new spatial index such as regional sum population, population growth rate, degree of population aggregation

4. RESULTS

The most important product of SPUS system is updating gridded population data. Figure 7 shows results achieved gridded population data of Shandong Province of China in 2002 with MODIS data. The spatial resolution of grid is 500m and the maxim value of population in 0.25 sq km grid is 3094 persons in Shandong Province. There are some high-value areas on image within county-level administrative unit showing the city area with high population density.

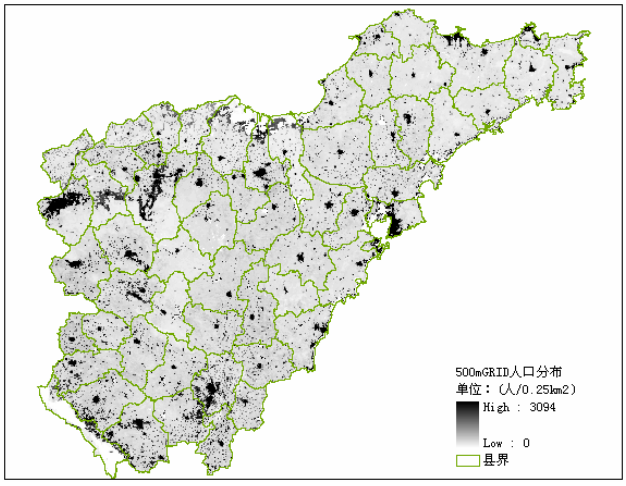


Figure 7. 500m gridded population data of Shandong Peninsula in 2002

Verification of redistributed China population data is time consuming mainly because of the difficulty of establishing a suitable reference database for purposes of comparison. It is difficult to get actual census counts for 1km × 1km grids. However, census data of sub-county unit, which is town in China, could be obtained in some provinces. A substitute approach has been designed for verification based on these population data. The following steps were carried out for towns with census data: (1) Create vector boundary maps of towns with census data; (2) Overlay gridded population data with boundary maps ;(3) Accumulate population of all cells within these towns;(4) Compare population estimated with census data. We have collected census data of 19 towns in Yishui county. Verifications have been done with 2002 statistical data (table 3). Relative errors between predicted population and actual census counts varies from 1% to 25.6% and 8 towns are lower than 10%, 8 higher than 20%, the rest 3 between 10%-20%. The relative error result is acceptable.

5. CONCLUSION

This paper described research idea of the statistical population spatialization updating using MODIS data as main input data source, and developed the Spatial Population Updating System (SPUS). SPUS integrates GIS, RS, DB and mathematical statistical method as one professional application system, provides various spatial analyses functions and realizes the gridded population data updating automatically.

Further research may lead in two directions. One is to improve model accuracy with more factors related to the spatial distribution of population especially inside city and validate population model with more actual census counts at finer resolution available. Another one is to extend SPUS functions for wider and easier use such as adding new complex population index to help users building spatial population database.

ACKNOWLEDGEMENTS

The study was supported by the project of National Natural Science Foundation of China (40471112) and partly CAS Knowledge Innovation Project (KZCX2-308). We would like to thank Professor and for their help in the research.

REFERENCES

A. W. Caruthers., 1985. Mapping the population census of Scotland. *The Cartographic Journal*.,22:83-87

Christopher Small, and Joel E. Cohen. Physiography, Climate and the Global *Distribution of Human Population*. 2000. URL: sedac.ciesin.org/plue/gpw/workshop.html

Clark C. Urban., 1951. population densities. *Journal of the Royal Statistical Society* . 114:490-496

Gai YingChun, FENG Min, GUO Jianwen1, SHANG Qingsheng., 2005. Study on Communication Mechanism Between IDL and. Net Environment. *Remote Sensing Technology and Applications*. 6(3):350-354 (in Chinese)

Gao Zhiqiang, Liu Jiyuan and Zhuang Dafang., 1999. Study on relation between ecological environment quality of China land resources and population distribution based on remote Sensing and GIS. *Journal of Remote Sensing*. 3(1): 79-83. (in Chinese)

Gregory Yetman, Uwe Deichmann, and Deborah Balk, 2000, Creating a Global Grid of Human Population, URL: http://sedac.ciesin.org/plue/gpw

Hu Huanyong. 1983. On population distribution of China. China Science Press, Beijing, China, 518p. (in Chinese)

Jerome E. Dobson. 1998. LandScan Global Population 1998 Database. URL: http:// www.ornl.gov /gist /projects /LandScan

K. MURAMATSU, S.FURUMI, N.FUJIWARA., 2000. Pattern decomposition method in the albedo space for Landsat TM and MSS data analysis. *INT.J.REMOTE SENSING*, (1): 99-119.

Langford M., and D. J. Unwin,, 1994. Generating and mapping population density surfaces within a geographical information system, *The Cartographic Journal*, 31:21-26

Liu Zheng. Questions of population theories. Beijing: China Social Science Press. 1984 (in Chinese)

Lo C.P. 1998. Application of Landsat TM data for quality of life assessment in an urban environment. Computer *Environment and Urban Systems*, 21(3/4): 259-276.

Mi Hong, Ji Guoli, Lin Qican., 1999. Study on systematic theory and evaluation methods of sustainable development based on harmonic development among people, resources and environment on county level in China. *Population and Economy*, 6: 17-24. (in Chinese)

Paul Sutton. 1997. Modeling population density with nighttime satellite imagery and GIS. *Computer Environment and Urban System*, 21(3/4): 227-244

Robert B., Matlock, J.R and John B. Welch. 1996. Estimation population density per unit area from mark, release, recapture data. *Ecological applications*. 6(4): 1241-1253

YANG Xiaohuan, JIANG Dong, WANG Naibin, LIU Honghui., 2002. Method of Pixelizing Population Data. *Journal of Geographical Sciences*,57

Yi-Fan Li. 1996. Global Population Distribution Database. URL：grid2.cr.usgs.gov/globalpop/l-degree