Remote Sensing of Environment 210 (2018) 269–281

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Mapping population density in China between 1990 and 2010 using remote sensing

Litao Wang, Shixin Wang, Yi Zhou*, Wenliang Liu, Yanfang Hou, Jinfeng Zhu, Futao Wang

Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Mapping population density
Remote sensing
China
Population spatialization

ABSTRACT

Knowledge of the spatial distribution of populations at finer spatial scales is of significant value and fundamental to many applications such as environmental change, urbanization, regional planning, public health, and disaster management. However, detailed assessment of the population distribution data of countries that have large populations (such as China) and significant variation in distribution requires improved data processing methods and spatialization models. This paper described the construction of a novel population spatialization method by combining land use/cover data and night-light data. Based on the analysis of data characteristics, the method used partial correlation analysis and geographically weighted regression to improve the distribution accuracy and reduce regional errors. China's census data for the years 1990, 2000, and 2010 were assessed. The results showed that the method was better at population spatialization than methods that use only night-light data or land use/cover data and global linear regression. Evaluation of overall accuracies revealed that the coefficient of correlation R-square was > 0.90 and increased by > 0.13 in the years 1990, 2000, and 2010. Moreover, the local R-square of over 90% of the samples (counties) was higher than the adjusted R-square of the general linear regression model. Furthermore, the gridded population density datasets obtained by this method can be used to analyse spatial-temporal patterns of population density and provide population distribution information with increased accuracy and precision compared to conventional models.

1. Introduction

In sociology, population refers to a collection of humans in a certain region within a certain period. According to the Concise Report on the World Population Situation in 2014, the world's population reached 7.2 billion in 2014, and its state can be described as unprecedentedly diverse (United Nations, 2014). The distribution of the world's population and population density vary widely (Small and Cohen, 2004). Understanding the spatial pattern of population distribution has a very important and fundamental role in research related to environmental changes, urbanization, regional planning, public health, and disaster management (Clark and Rhind, 1992; Dilley et al., 2005; Garb et al., 2007; King et al., 1994; Rabelo et al., 2006; Su et al., 2010; Vitousek et al., 1997).

Demographic census is an important administrative data source on population distribution (Zeng et al., 2011). However, census data are usually delimited by administrative boundaries (such as counties) and reported at the centroids of the census polygon (Bhaduri et al., 2007). The census statistics represent entire populations of the census units and cannot highlight the spatial distribution of residents in different parts of the units. However, for many applications, population distribution in the interior of a census unit is necessary. For example, ‘population at risk’ is the estimation of population in an area declared as under risk (such as an inundated region) and usually does not follow census boundaries. Furthermore, the administrative boundaries in a region may change over time, rendering temporal comparison of census statistics difficult (Briggs et al., 2007; Holt et al., 2004). However, gridded population density datasets can overcome these disadvantages as they can reflect the spatial-temporal characteristics of population distribution.

Spatial interpolation methods for gridded population density have been extensively studied over the last few decades (Bhaduri et al., 2007). The areal weighted interpolation is the simplest method that can be used when the population in a census unit is assumed to be uniformly distributed (Bhaduri et al., 2007; Goodchild and Lam, 1980; Goodchild et al., 1993; Pan and Liu, 2002). However, some inherent problems limit the accuracy and precision of this method despite its advantage of easy execution (Langford, 2003). Firstly, the areal method does not account for the potential variations in population density due to scale and boundary effects (Holt et al., 2004). Furthermore,

* Corresponding author.
E-mail address: zhouyi@radi.ac.cn (Y. Zhou).

https://doi.org/10.1016/j.rse.2018.03.007
Received 12 October 2017; Received in revised form 27 February 2018; Accepted 4 March 2018

population density maps based on this method indicate abrupt changes at the boundaries of the administrative areas. Some researchers have employed smoothing functions to model population distribution as a continuous surface (Martin, 1989; Tobler, 1979). However, population patterns in most areas are significantly disjunct and concentrated into smaller populations that are separated by larger areas of dispersed populations (Briggs et al., 2007). Dasymetric methods that utilize ancillary spatial data with fine spatial resolution and variability (e.g., land use/cover) to redistribute the population into census units are currently used widely (Fisher and Langford, 1995; Goerlich and Cantarino, 2013; Holt et al., 2004). The dasymetric model assigns weights to density and has two aspects: representing the difference among different land use/cover classes; and representing the differences of the same class in different regions. Previous works have used techniques such as the limiting variable method, regression, multi-dimensional adaption, and smart interpolation (Eicher and Brewer, 2001; Gallego et al., 2011; Mennis, 2003; Mennis and Hultgren, 2006).

Remotely sensed data, especially night-light imageries from the Defense Meteorological Satellite Program (DMSP) Operational Linescan System (OLS) (Elvidge et al., 1997), have provided useful indicators for assigning weights to population densities and constructing population density models (Amaral et al., 2005, 2006; Briggs et al., 2007; Zhang et al., 2015; Zhuo et al., 2009). Night-time imagery can make spatialized population density "reflect the population distribution character more explicitly and in greater detail" (Zeng et al., 2011). However, using only the night-light data to downscale the population information is unreliable (Briggs et al., 2007).

The aim of this paper is to develop a spatialization method for population census data using spatial analysis techniques. The population census data from the fourth (1990), fifth (2000), and sixth (2010) National Population Censuses of the People's Republic of China were used to produce population grid density datasets based on the described method. These datasets evaluate the precision of the proposed method and analyse spatial-temporal patterns in the population density of China since 1990s. During this period, China's population distribution has changed notably with the rapid development of society and economy. These datasets will play an important role in analysing rapid urbanization, environmental changes, and other research fields.

2. Data and pre-processing

2.1. Population census data and administrative boundaries

Census data are the best available baseline data on populations (Briggs et al., 2007). After the formation of the People's Republic of China, six national population censuses have been conducted in the years 1953, 1964, 1982, 1990, 2000, and 2010. Total resident population in county-level administrative units was obtained from the public data of the national population censuses conducted in 1990, 2000, and 2010, and used for population spatialization. However, the data for Hong Kong, Macao, and Taiwan were not used in this study.

County-level administrative boundary maps (scale of 1:4,000,000) were downloaded from the website of the National Fundamental Geography Information System (NFGIS) (http://nfgis.nsdi.gov.cn/ (accessed April 1, 2009)).

During the association of multi-temporal census data with their respective administrative units, the census datasets of 1990, 2000, and 2010 exhibited some inconsistencies with the county-level administrative boundary maps from NFGIS due to administrative changes such as the change of county name and adjustment of county range. We ensured that the census population was consistent with the area from the administrative boundary maps through data revision and checking. Subsequently, census population was joined spatially with the corresponding administrative boundary vector element (Fig. 1).

2.2. Land use/cover data

The land use/cover data were obtained from the National Land Use/Cover Database of China (NLUD-C), created under a project to map the national land use/cover using remote sensing and geographic information system techniques (Liu et al., 2005). The project was conducted by the Chinese Academy of Sciences (CAS) that employed eight research institutions and approximately 100 scientists in the late 1990s. The NLUD-C is generated on a 1:100,000 scale from remotely-sensed monitoring data and covers five periods of land use/cover data, including the late 1980s, 2000, 2005, 2008, and 2010 (Zhang et al., 2014). This database has been the main source of authoritative data on land use/cover from the late 1980s in China (Liu and Deng, 2010; Liu et al., 2014, 2009, 2005, 2003).

The land use/cover data from the NLUD-C dataset has a hierarchical classification system comprising 25 land use/cover classes and 25 raster files. Each raster file represents a particular land cover type, and the value of each grid-cell in the raster file denotes the percentage of the specific land cover type in the grid-cell. The data we obtained had been transformed into raster data with 1 km resolution using cell-based encoding for percentage breakdown.

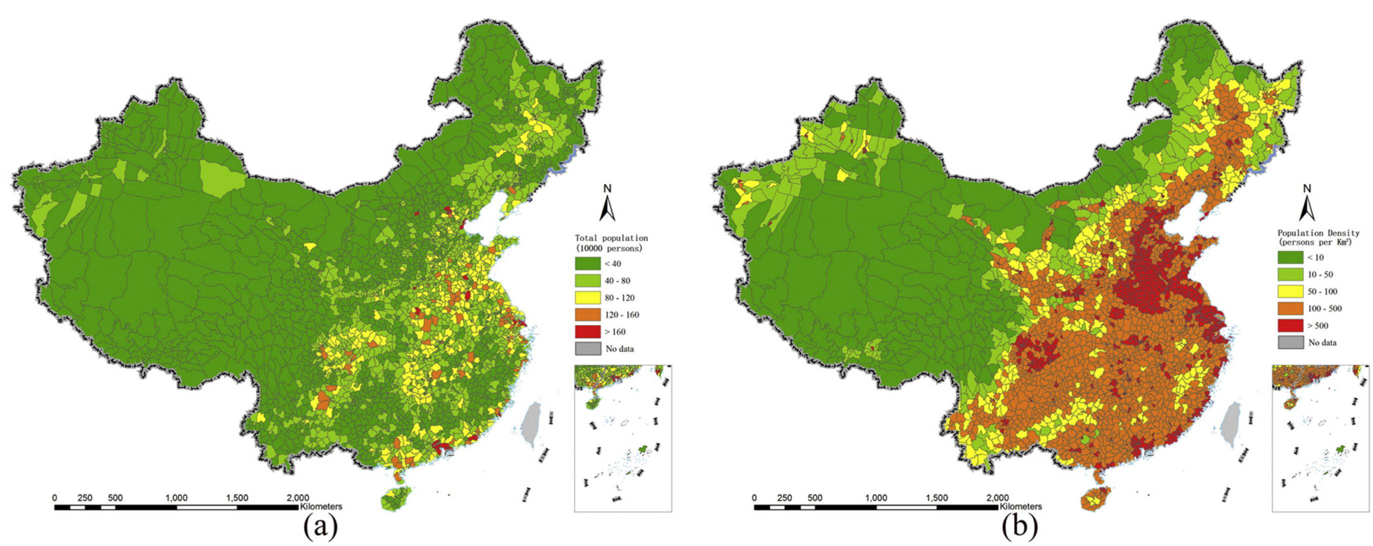

Fig. 1. Total population (a) versus population density (b) in the year 2010.