Applied Geography 50 (2014) 99–107

Contents lists available at ScienceDirect

Applied Geography

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

A fine-scale spatial population distribution on the High-resolution Gridded Population Surface and application in Alachua County, Florida

Peng Jia [a,*], Youliang Qiu [b], Andrea E. Gaughan [c]

[a] Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA, USA
[b] Department of Geography, University of Florida, Gainesville, FL, USA
[c] Department of Geography and Geosciences, University of Louisville, Louisville, KY, USA

Keywords:
Dasymetric mapping
Parcel
Population
GIS
Land Cover
Superfund

A B S T R A C T

Geospatial techniques, using Geographic Information Systems and remote sensing data, have become more commonly used with dasymetric modeling of fine-scale demographic data. In this study, we apply a dasymetric approach using the Heuristic Sampling Method for 2010 parcel data to disaggregate population counts from the 2010 U.S. Census into a quadrilateral grid composed of 30 × 30 m cells covering the Alachua County, Florida. The final output, termed the High-resolution Gridded Population Surface (HGPS), is compared to a land cover-based population product (LCPP) and the detail of each product is assessed. Results suggest that the HGPS provides increased spatial heterogeneity and more detail in the boundaries of populated areas over the use of census blocks or land cover lots. For an example of the final output, we use a case study at the Cabot–Koppers Superfund Site to demonstrate the advantages of the HGPS over the LCPP. The HGPS is expected to serve as a more accurate input in various research fields, such as public health, crime analysis, and climate change. The approach outlined provides an improved means of producing spatially-explicit population grids where fine-scale ancillary data, such as parcel data, is available.

Introduction

As cities develop and their inner structures get more complex, the study of population distribution within city boundaries becomes increasingly important. The location of people, compared to traditional statistical numbers, is critical to many aspects of risk assessment and policy decision-making, especially in response to natural and artificial disasters. For instance, emergency management plans need evacuation routes and the ability to quickly allocate resources in areas where people live (Chakraborty, Tobin, & Montz, 2005; Pal, Graettinger, & Triche, 2003). Environmental health assessments need to spatially correlate environmental risk factors to health outcomes based on the concentration of people in given areas (KuÈnzli et al., 2000; Maantay, Maroko, & Herrmann, 2007). Surveying and analyzing the distribution of demographic characteristics need to be done prior to any optimal site selection with the purpose of private profits or public service (Luo & Wang, 2003; Tayman & Pol, 2011). Therefore, there is demand for population products that accurately identify the spatial distribution and density of the population to meet the growing demand for immediate and well-informed decision making (Krunić, Bajat, Kilibarda, & Tošić, 2011; Maantay et al., 2007).

Census data are still the primary data source for a variety of demographic research endeavors, but they are usually released to the public about once every 10 years, in an aggregated form for confidentiality and administrative purposes (Wu, 2006). The aggregated form, known as a choropleth representation (Mennis, 2003), is composed of a series of vector polygons representing administrative units and summed values depicting the aggregated attributes over those units. Martin (Martin, 2011) highlights major limitations of conventional choropleth maps from census data, including the Modifiable Areal Unit Problem (MAUP), time lag between data collection and publication and susceptibility to periodic changes of administrative boundaries. The MAUP is a matter of prime importance, referring to the fact that the scale at which data are aggregated and how areal units are grouped can cause substantial changes in the results of spatial data analysis (Openshaw, 1984). As such, the continuous and dynamic nature of human population is poorly described by discrete, aggregated choropleth-

* Corresponding author.
E-mail addresses: pjia3@lsu.edu (P. Jia), youliang@gmail.com (Y. Qiu), aegaughan@gmail.com (A.E. Gaughan).

http://dx.doi.org/10.1016/j.apgeog.2014.02.009
0143-6228/

oriented data products that often rely on static modeling approaches (Martin, 2011). An additional shortcoming of using data in a choropleth form is the masking of spatial heterogeneity within given administrative units by replacing a range of values with one aggregated value and also being susceptible to abrupt changes at the boundary of artificially defined units that do not necessarily relate to any natural or artificial phenomenon (Pfeiffer & Hugh-Jones, 2002).

Dasymetric mapping is becoming much more standard with the increase and incorporation of spatial technologies and data sets (Langford, 2007; Lin, Cromley, & Zhang, 2011; Saporito, Chavers, Nixon, & McQuiddy, 2007). Dasymetric mapping is a particular cartographic process by which a traditional choropleth map or irregular partitioning of data can be transformed into a regular and continuous gridded surface with values re-distributed across a given spatial unit rather than relying on the aggregated values (Martin, 2011; Mennis, 2009). Ancillary data are a key element for dasymetric mapping and utilizing Geographic Information Systems (GIS) along with remotely-sensed data sets provides a large pool of ancillary data to inform the disaggregation of data (Wang & Wu, 2010).

Currently there are a few well-modeled, high-quality population products covering the United States (U.S.), which include the Gridded Population of the World (GPW) (Balk & Yetman, 2004), Global Rural Urban Mapping Project (GRUMP) (CIESIN, 2004), LandScan Global (Dobson, Bright, Coleman, Durfee, & Worley, 2000) and LandScan USA model (Bhaduri, Bright, Coleman, & Urban, 2007). These regional to global scale products incorporate various ancillary data sets such as land use/land-cover (LULC), nighttime lights, transportation network, various landmarks, elevation, slope, etc., to generate final population data sets. However, the underlying modeling frameworks of these products often rely on coarser-scale, more generalizable input data or have modeling algorithms that provide more flexibility for incorporating regional and global scale data sets (Gaughan, Stevens, Linard, Jia, & Tatem, 2013; Tatem, Campiz, Gething, Snow, & Linard, 2011).

For interests at a more local scale, parcel boundaries provide detailed information which may more accurately portray the underlying population distribution than coarser, generalized data sets. This is due to the nature of parcel data and its specific orientation towards population density for a given point in space, especially in rural areas (Tapp, 2010). Parcel data also provides potentially important information, because while manmade structures can be identified through analysis of high-quality remote sensing data, functional information associated with parcel data, such as property types, rarely can be reliably extracted by any known algorithm (Xie, 2006). Although parcel data have been used in a few number of studies (Maantay et al., 2007; Tapp, 2010; Xie, 2006) and likely others, to the authors' knowledge, there has not been any effort on examining the relationship between population density and parcel type.

Nineteen states in the U.S. have digitized at least 80% of their parcels (Stage & Von Meyer, 2006), of which the State of Florida maintains an annual statewide digital parcel data set containing the boundaries of parcels in all 67 counties with associated tax roll information from the Florida Department of Revenue's tax database (alternatively referred to as tax parcel data) (Florida Department of Revenue, 2006). We chose the county of Alachua in north-central Florida which is a moderate size and includes mixed land use types for this study in order to initiate an effort of parcel-based population modeling.

Various techniques have been used to increase the accuracy and detail of dasymetric modeling approaches (Maantay et al., 2007; Mennis, 2009; Wu, Qiu, & Wang, 2005). A well-cited approach, the Heuristic Sampling Method (HSM) (Mennis, 2003), was developed based on a "grid three-class" method described by Eicher and Brewer (Eicher & Brewer, 2001). Mennis proposed empirical sampling and area-based weighting techniques to mitigate the subjectivity of the grid three-class method which relies on a predetermined percent of population assigned to raster land-use classes. However, land cover lots used in Mennis' study still fail to adequately distribute population in a detailed enough manner for purposes of identifying location-specific population for disaster-mitigation and environmental risk assessments. Tax parcel data is an appropriate alternative ancillary data set instead of land cover that may provide the necessary increased level of spatial detail in population counts. In this study, we present an approach to improve upon the HSM using parcel data hereafter termed the High-Resolution Gridded Population Surface (HGPS) for population redistribution.

In the following sections, we outline the data and methodology for producing a 30 × 30 m HGPS in Alachua County, Florida. Then we compare with other approaches and assess the accuracy of the HGPS. Lastly, we present an application of HGPS in public health in Alachua Country, visualizing the spatial distribution of the population potentially exposed to the contamination at Cabot-Koppers Superfund Site in the northeastern Gainesville, Alachua County.

Study region, datasets and methods

The State of Florida is located in the southeastern United States and Alachua County is located in the north-central part of Florida (Fig. 1). The case study is conducted in the urban area of Gainesville, the county seat and largest city in Alachua County, with a population of 176,096 in 2010, also home to the University of Florida (UF).

Datasets

The 2010 U.S. Census data consist of three levels that include census tract, block group and block. These data are collected every 10 years for all three levels, recording the total population over areal units and population within each category for race and ethnicity. The 2010 Tax Parcel data contain the boundaries of parcels in all 67 counties of Florida with associated tax information including the property types of parcels (Florida Department of Revenue, 2010).

Modeling approach

Empirical sampling

The HGPS uses tax parcel data and a novel approach to improve upon the HSM in redistribution of population counts. While spatially overlaying the administrative boundary and parcel layers for empirical sampling, an issue was identified in which rare census units, even at the finest scale blocks, are 100% covered by only one type of parcel, also called a property type, due to the nature of mixed property types within census blocks. Even in areas with only one type of parcel, there are streets and open space among houses. Those areas do not include any population on them, so are supposed to be excluded when counting residential area. This is different from the situation while using land cover data as a land-cover class easily covers an entire block and is counted as an entire residential area including streets and open spaces. As a result of taking into account the increased detail of specific population location there is increased complexity in calculating the population density for each property type by empirical sampling.

To adapt the HSM to the new context, a concept of an eligible mono-type block is proposed. An eligible mono-type block is defined as a census block within which the area proportion of only