Applied Geography 63 (2015) 220–230

Contents lists available at ScienceDirect

Applied Geography

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

Incorporating spatial non-stationarity to improve dasymetric mapping of population

Kasper Cockx*, Frank Canters

Cartography and GIS Research Group, Department of Geography, Vrije Universiteit Brussel, Building F, Pleinlaan 2, B-1050 Brussel, Belgium

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form
3 July 2015
Accepted 4 July 2015
Available online xxx

Keywords:
Dasymetric mapping
Population estimation
Spatial non-stationarity
Rule-based regression
GWR
Flanders–Brussels region

ABSTRACT

Population data are traditionally obtained through censuses and aggregated up to the level of administrative units for reasons of privacy. This way, however, detailed information on the spatial distribution of the population within these units is masked. Dasymetric mapping techniques were developed to disaggregate population to a finer spatial level using ancillary data. However, a frequently recurring problem in dasymetric mapping studies relates to the overestimation of low-population-density areas and the underestimation of high-population-density areas. To tackle this issue, this research proposes a novel dasymetric mapping approach explicitly dealing with spatial non-stationarity. For this purpose, a comparative model building framework was set up. The impact of spatial non-stationarity on model performance was investigated by comparing global (OLS), regional (rule-based) and local (geographically weighted) regression. Also, the impact of model complexity was considered through stepwise inclusion of information on address type and location, household size and demographic and residential characteristics in the dasymetric model. The approach was tested in the highly complex environment of the Flanders–Brussels region. It was found that the regional model that incorporates address type and household size information performs best and overcomes the structural over- and underestimation issue in dasymetric mapping.

1. Introduction

Information on the spatial distribution of population forms an important input for a wide range of analyses on resource management, facility allocation, land-use planning, natural hazard and environmental risk, infection control, disaster relief/mitigation, quality of life assessment and society–environment interactions (Dong, Ramesh, & Nepali, 2010; Li & Weng, 2005; Liao, Wang, Meng, & Li, 2010; Qiu, Sridharan, & Chun, 2010; Upegui & Viel, 2012). Population data are traditionally obtained through censuses and afterwards aggregated up to the level of administrative units for reasons of privacy. This way, detailed information on the spatial distribution of the population is masked, creating the unrealistic impression that population is distributed homogeneously throughout each administrative unit (Mennis, 2003; Wu & Murray, 2005). Also, results of spatial data analysis may be severely affected by data aggregation (referred to as the modifiable areal unit problem (MAUP) by Openshaw (1983)).

Given the limitations of aggregated census data, researchers have widely adopted the principles of dasymetric mapping to obtain population estimates at finer spatial scales. In dasymetric mapping (Wright, 1936), ancillary information that provides clues to the actual population distribution is used to subdivide a source zone for which the population is known into target zones, after which that population is disaggregated to these finer spatial units. In the literature, a distinction can be made between 'lower-resolution' (generalised information on the physical and functional characteristics of the built environment) and 'higher-resolution' ancillary data (detailed, local information on street networks, buildings and addresses) for use in dasymetric mapping. The most widely used type of lower-resolution ancillary data is land-use information. In the literature, many approaches have been proposed to estimate representative population densities per land-use category from available land use and population data and to use that information for disaggregating population to finer spatial units (e.g. Briggs, Gulliver, Fecht, & Vienneau, 2007; Eicher & Brewer, 2001; Holloway, Schumacher & Redmond, 1999; Langford, 2007;

---

* Corresponding author.
E-mail addresses: kacockx@vub.ac.be (K. Cockx), fcanters@vub.ac.be (F. Canters).

http://dx.doi.org/10.1016/j.apgeog.2015.07.002
0143-6228/

Langford, Magnire, & Unwin, 1991; Langford & Unwin, 1994; Li & Corcoran, 2011; Mennis, 2003; Mennis & Hultgren, 2006; Yuan, Smith, & Limp, 1997). Other types of lower-resolution ancillary data like spectral and/or textural metrics obtained from remotely sensed data (Harvey, 2002; Li & Weng, 2005; Liu, Clarke, & Herold, 2006) or demographic information and distance-to-services metrics (Deng, Wu, & Wang, 2010) have also been applied, but the results obtained in these studies suffered from low accuracies. Impervious surface fractions, on the other hand, have proven to perform equally well as or even better than land-use data as a source for disaggregating population data (Lu, Weng, & Li, 2006; Wu & Murray, 2007; Zandbergen & Ignizio, 2010). Recently, nonparametric modelling has been applied to disaggregate population based on a large number of remotely sensed and other geospatial variables (Patel et al., 2015; Stevens, Gaughan, Linard, & Tatem, 2015).

While many of the studies above relied on ancillary information obtained from remotely sensed data, the resolution of the imagery used was often too low to obtain accurate disaggregation results, especially in heterogeneous urban environments. With the advent of high-resolution remote sensing, image resolution has become less of a constraint in dasymetric mapping (see e.g. Ural, Hussain, & Shan, 2011). Nevertheless, given the high cost, limited coverage and narrow time range of this data source, other high-resolution data types have also proven interesting. Local infrastructure information recently has become a popular ancillary data source for disaggregating population to a finer spatial level, including street length or street density (Reibel & Bufalino, 2005; Su, Lin, Hsieh, Tsai, & Lin, 2010), ‘points of interest’ correlated with high population densities (Bakillah, Liang, Mobashen, Arsanjani, & Zipf, 2014) or building area or volume extracted from (a) LIDAR (Dong et al., 2010; Sridharan & Qiu, 2013; Upegui & Viel, 2012), (b) high-resolution satellite or aerial imagery (Lung, Lübker, Ngochoch, Schaab et al., 2013; Ural et al., 2011) or (c) parcel data (Jia, Qiu, & Gaughan, 2014; Maantay, Maroko, & Herrmann, 2007; Xie, 2006). To obtain accurate disaggregation results from building data, a distinction is best made between residential and non-residential buildings, and preferably between different residential types (houses and apartments) as well (Jia et al., 2014; Ural et al., 2011). Address data have also been used for dasymetric mapping purposes, e.g. in a study by Tapp (2010), who multiplied the number of addresses in the target zone's residential parcels by the average household size in the source zone. Zandbergen (2011) compared using residential address points with dasymetric mapping techniques based on land use, impervious surface cover, light emission and road density. Both Tapp (2010) and Zandbergen (2011) concluded that methods using address points outperform more traditional dasymetric mapping techniques.

Despite the aspiration to arrive at a finer approximation of the actual spatial distribution of the population through ancillary data with ever-increasing detail, a frequently recurring problem in dasymetric mapping remains the overestimation of low-population-density areas and the underestimation of high-population-density areas (Eicher & Brewer, 2001; Gallego, 2010; Gaughan, Stevens, Linard, Patel, & Tatem, 2014; Harvey, 2002; Li & Corcoran, 2011; Li & Weng, 2005; Lu et al., 2006; Mennis & Hultgren, 2006; Su et al., 2010; Upegui & Viel, 2012; Ural et al., 2011; Yang, Yue, & Gao, 2013). This can be attributed to a lack of accounting (sufficiently) for spatial non-stationarity: the derivation of global parameters (be they interpolation weights or regression coefficients) imposes an averaging effect on the disaggregation that masks intrinsic variation in population distribution characteristics, which manifests itself particularly at the population density extremes. This is especially true in medium-scale to small-scale studies covering locations along the rural-urban continuum, which are characterised by a high range of population densities. Rural, low-density areas tend to have a more variable population distribution than medium-to-high-density areas with urban or suburban development, which may be relatively homogeneous (Zandbergen, 2011). Therefore, in research focussing on large-scale, metropolitan or intra-urban areas, the impact of using globally defined parameters may be less pronounced (see Murakami & Tsutsumi, 2011; Qiu et al., 2010; Wu & Murray, 2005, 2007).

Few strategies have been proposed to address spatial non-stationarity in the population estimation field. One suggestion involves dividing the study area into subregions and performing a separate population redistribution within each subregion, as has been done in Li and Corcoran (2011), Mennis (2003) and Yuan et al. (1997). However, these studies neglect the rather arbitrary nature of administrative boundaries, which are unlikely to delineate areas with homogeneous population distribution characteristics. Therefore, local regression approaches that estimate separate coefficients for each population record have been tested through (i) quantile regression (Cromley, Hanink, & Bentley, 2012) and (ii) geographically weighted regression (Dong et al., 2010; Lin, Cromley, & Zhang, 2011; Lo, 2008). While the application of these techniques showed an improvement upon the global regression approach, it still did not seem to solve the classic problem at the population distribution's extremes though.

In this research, a dasymetric mapping approach is proposed that explicitly deals with spatial non-stationarity for accurately predicting fine-grained population in spatially complex, strongly heterogeneous areas. The strategy for dasymetric mapping presented builds on three key features: 1) functional differentiation based on land-use information, 2) population model calibration at the source zone level and 3) volume-preserving disaggregation of the population up to the level of individual address points. To develop our approach, a comparative model building framework was set up accounting for increasing model complexity and spatial non-stationarity.

2. Study area

The approach proposed in this paper was tested on the Flanders region and the Brussels Capital Region, which cover the northern part of Belgium and together form one of Europe's most densely populated areas, with on average 550 inhabitants per km^2 (based on FOD Economie, 2014) (Fig. 1). Characteristic of this area is the complex urban pattern consisting of a few big and several regional cities and smaller settlements that are all connected through a dense road network, along which vast elongated areas of so-called “ribbon-building” occur as a result of strong suburbanisation during the last decades (Meeus & Gulinck, 2008). Consequently, morphological and functional urban relationships are defined by expanded city regions consisting of (a) an urban agglomeration, which comprises the continuous fabric of the core city and the urban fringe, and (b) suburban zones with a rural morphology, but urban functionality (Van Hecke, Halleux, Decroly, & Mérenne-Schoumaker, 2009). The Flanders–Brussels area includes twelve city regions, in which high population densities occur within the core city and – to a greater (Brussels, Antwerp, Ghent) or lesser extent (most of the regional towns) – within the urban fringe; the suburban zones generally show a mix of low and medium population densities (Fig. 1).

To perform meaningful studies in highly sprawled regions like Flanders and Brussels, the need for detailed, high-resolution data has been repeatedly emphasized (Meeus & Gulinck, 2008).