Applied Geography 86 (2017) 22–31

Contents lists available at ScienceDirect

Applied Geography

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

Dasymetric distribution of votes in a dense city

Jose M. Pavía [a, *], Isidro Cantarino [b]

[a] Universitat de Valencia, Av Tarongers, s/n, 46022, Valencia, Spain
[b] Department of Land Engineering, Polytechnic University of Valencia, Escuela de Caminos, Camino de Vera, s/n, 46071, Valencia, Spain

ARTICLE INFO

Article history:
Received 1 June 2016
Received in revised form
11 March 2017
Accepted 14 June 2017
Available online 23 June 2017

Keywords:
Modifiable areal unit problem
Boundary changes
Small-areas
Elections
Barcelona

ABSTRACT

A large proportion of electoral analyses using geography are performed on a small area basis, such as polling units. Unfortunately, polling units are frequently redrawn, provoking breaks in their data series. Previous electoral results play a key role in many analyses. They are used by political party workers and journalists to present quick assessments of outcomes, by political scientists and electoral geographers to perform detailed scrutinizes and by pollsters and forecasters to anticipate electoral results. In this paper, we study to what extent more complex geographical approaches (based on a proper location of electors on the territory using dasymetric techniques) are of value in comparison to simple methods (like areal weighting) for the problem of reallocating votes in a large, dense city. Barcelona is such a city and, having recently redrawn the boundaries of its census sections, it is an ideal candidate for further scrutiny. Although previous studies show the approaches based on dasymetric techniques outperforming simpler solutions for interpolating census figures, our results show that improvements in the process of reallocating votes are marginal. This brings into question the extra effort that entails introducing ancillary sources of information in a dense urban area for this kind of data. Additional research is required to know whether and when these results are extendable.

1. Introduction

Thanks to the IT revolution, the problem of reallocating census data from a set of geographical administrative units onto another overlapping but non-hierarchical set of spatial units could be an in-practice solved problem in Western countries. Nowadays collecting, storing, and transmitting detailed data is simpler than ever and handling big data sets of microdata has become such an easy task that reconstructing population longitudinal series for the new spatial units is potentially just an issue of resolve and budget for statistical agencies. Furthermore, thanks to the popularisation of geo-localization tools in mobile devices, we are moving towards a future in which the storing of the geographical coordinates of each datum collected will be possible (Larraz, Pavía, & Ferrari, 2013).

The above arguments may lead to conclude that no additional research in this area is now needed as census aggregate data could be reconstructed from individual records. That of course would be a mistake. On the one hand, it is difficult to conceive a situation in which ballot secrecy disappears from political issues and where the votes are not collected and declared geographically aggregated. On the other hand, there are many historical data sets for which microdata with spatial marks, or simply microdata, do not exist. Moreover, due to confidentiality issues, the average analyst hardly ever has access to the microdata files, even when available. In these cases, the unique solution is to use geographical and statistical methods to estimate the non-available data in one spatial breakdown from the available data in the other spatial breakdown. The research in this area has been indeed very promising over the last few years.

We focus on the problem of the spatial redistribution of votes among two census section breakdowns. This is an interesting problem due to the key role played by small-area past election results in political modelling. For example, they are routinely used by political geographers (e.g., Pattie & Johnston, 2000) and electoral pollsters (e.g., Pavía & Larraz, 2012). The aim of the paper is therefore to utilize geographic approaches to resolve not the customary problem of population reallocation, but a problem with a spatial dimension as is the reallocation of votes that many political and electoral analysts face.

For the general problem, a large number of methods have been suggested in the literature. They have evolved from simple areal weighting procedures (e.g., Goodchild & Lam, 1980) and point-

* Corresponding author.
E-mail addresses: pavia@uv.es (J.M. Pavía), icantari@trr.upv.es (I. Cantarino).

http://dx.doi.org/10.1016/j.apgeog.2017.06.021
0143-6228/

based spatial interpolation methods (e.g., Fisher & Langford, 1995) to more complex dasymetric mapping strategies. Dasymetric mapping (Wright, 1936) is a disaggregation procedure that tries to incorporate in an intelligent fashion ancillary data in the refinement process. Dasymetric techniques seek to define homogeneous areas based on the actual spatial distribution of the variable of interest, rather than on administrative or other arbitrary units (Mennis, 2009). Although dasymetric mapping was initially conceived as a disaggregation method, it is employed in the reallocation problem through a two-step process in which source data is first spatially refined and then aggregated to the target units. Auxiliary variables used in the refining process have been among others: land uses (Giordano & Cheever, 2010; Mennis, 2003), nighttime lights and road networks (Reibel & Bufalino, 2005), the spatial distribution of built structures (Maantay, Maroko, & Herrmann, 2007), residential building volumes (Sridharan & Qiu, 2013), satellite images (Holt, Lo, & Hodler, 2004), volunteered geographic information (Bakillah, Liang, Mobasheri, Arsanjani, & Zipf, 2014), mobile phone data (Deville et al., 2014), and a combination of spatial methods and Maximum Entropy or the Expectation-Maximization (EM) algorithm (Sridharan & Qiu, 2013; Schroeder & Van Riper, 2013; Buttenfield, Ruther, & Leyk, 2015). A new avenue of research based on the spatiotemporal interpolation of flow variables is likewise emerging (Mennis, 2016).

Areal weighting and point-based spatial interpolation methods are conceptually simple and do not require an in-depth comprehension of spatial methods. They are quite intuitive and can be implemented easily in free software like R. Dasymetric techniques are notably more complex and demand a higher understanding and ability in the use of GIS tools. They entail the employment of more spatial layers and combining data from several sources.

This work examines to what extent the use of a more complex approach is worthwhile in the spatial vote redistribution problem. In particular, we pick up the gauntlet thrown down by Pavía and López-Quilez (2013, p. 663), who point out that it would be worth testing "how the use of dasymetric mapping and related techniques would enhance the quality of approximations". Although dasymetric mapping has been proved to be more robust and accurate for the spatial reallocation of census data, we explore here whether this superiority is preserved for voting data. The comparison between approaches is customary (see, e.g., Ruther, Leyk, & Buttenfield, 2015; Buttenfield et al., 2015; Pavía & Cantarino, 2017) an relevant because depending on the final goals of the redistribution process, the analyst should weigh up if it is worth applying more time consuming and complex methods or just to consider simple approaches.

We have performed the assessment for the case of reallocating votes among really small areas of a big city, a situation in which the more complex approaches are valuable a priori. On the one hand, small areas' boundaries are routinely shaped with no special meaning in terms of the underlying spatial distribution of the variable of interest, an issue implicitly assumed by simpler methods. On the other hand, according to Poulsen and Kennedy (2004), dasymetric techniques are particularly useful with data that have spatial patterns dictated by an underlying geographically distributed structure, either physical or social; and, as it is known (see, e.g., O'Loughlin, 2002; Johnston & Pattie, 2006; Pavía, Larraz, & Montero, 2008), election results show geographical structure and spatial patterns.

Barcelona in Spain, a polycentric and spatially complex city (Catalán, Saurí, & Serra, 2008) with a marked social structure (Broner, 2010) and more than 1.6 million inhabitants, has been taken as a case of study and the outcomes corresponding to its 2007 local election as target variable. At the beginning of 2009 the way Barcelona (the second-largest city in Spain) was broken up in small areas was redrawn and its number of precincts (called census sections in Spain) dropped from 1482 to 1061 (defining a total of 6329 different intersection polygons). Every four years local elections are held in Spain to elect mayors. Therefore, having available (an approximation of) the results that had been recorded in the 2007 local election in the census sections (the smallest geographical units for which votes are declared in Spain) that correspond to the 2011 local elections is of interest for many agents, including political analysts, sociologists, economists, geographers, politicians, survey pollsters and electoral forecasters.

The rest of the paper is organised as follows. Section 2 describes the spatial methods analysed in this research. In addition to the baseline approaches of point-based spatial interpolation and areal weighting, four additional methods based on dasymetric techniques, are proposed as alternatives. In particular, we compare them to four dasymetric procedures that have proven to be superior when dealing with the customary problem of reallocating census data using exactly the same breakdowns (Pavía & Cantarino, 2017). Section 3 focuses on the challenge that individual secret ballots entails in terms of disposing of a standard for validation and how this can be solved. In Section 4 the different reallocations obtained after applying the methods described in section 2 and section 3 are compared and their relative merits assessed. Section 5 discusses and summarizes findings.

2. Methods and data

In this section, the particular details of the reallocating approaches assessed in this research are described. The first two proposals (point-based spatial interpolation and areal weighting) are classical and quite simple. They are used as a baseline to gauge the value of four other, more complex alternatives based on dasymetric techniques. We have followed the example of Pavía and Cantarino (2017) when choosing the dasymetric methods considered in this research. For exactly the same breakdowns analysed in this paper, Pavía and Cantarino (2017) have shown that dasymetric refinements clearly outperform simple methods when dealing with the customary problem of reallocating census figures. They find the expected hierarchy: 3-D procedures producing the better outcomes, followed by multiclass 2-D methods, binary 2-D approaches, and areal weighting and 1-D algorithms, with the point-based procedures generating by far the worst estimates. As dasymetric procedures of comparison, we have chosen within each one of the classes (1-D, binary 2-D, multiclass 2-D and 3-D) one representative. In what follows, we first introduce the sources and variables employed for the dasymetric refinements. Afterwards, we present the procedures considered.

2.1. Geographic sources of information

In addition to the election figures, we have managed geographical files provided by four institutions: the City Council of Barcelona (*l'Ajuntament de Barcelona*), the Spanish Geographic Institute (*Instituto Geográfico Nacional*), the European Environment Agency (EEA) and the Spanish Cadastral Agency (*Dirección General del Catastro*). From the official website of cartographic information of *l'Ajuntament de Barcelona*, we took (i) the shape (shp) files corresponding to the 2007 and 2011 section breakdowns of Barcelona and (ii) the Barcelona city street map (*Callejero*). From the *Instituto Geográfico Nacional*, we have downloaded the 2009 Land Cover and Land Use Information System of Spain (SIOSE), which is a unique database of Spain with a scale of 1:25,000. SIOSE combines topographic maps, satellite imagery, aerial photography and cadastral