The final publication is available at

https://doi.org/10.1016/j.apgeog.2017.06.021

Additional Information

Dasymetric distribution of votes in a dense city

Jose Manuel Pavía (corresponding author)
Universitat de Valencia
Av Tarongers, s/n, 46022-Valencia (Spain)
email: pavia@uv.es

Isidro Cantarino
Department of Land Engineering, Polytechnic University of Valencia
Escuela de Caminos, Camino de Vera, s/n, 46071-Valencia (Spain)
email: icantari@trr.upv.es

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 journalists and parties' teams 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 dense big city. The case of Barcelona (Spain), a city that recently redrew its census sections' boundaries, is analysed. Although previous studies show the approaches based on dasymetric techniques outperforming simpler solutions interpolating census figures, our results show that improvements in the process of reallocating votes are marginal. This questions 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.

Keywords: Modifiable areal unit problem; Boundary changes; Land Cover and Use Information System (SIOSE); Small-areas; Elections.

Dasymetric distribution of votes in a dense city

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. Nowadays collecting, storing, and transmitting detailed data have become simpler than ever and handling big data sets of microdata in such an easy task that reconstructing population longitudinal series for the new spatial units is just an issue of will and budget for statistical agencies. Furthermore, thanks to popularisation of geo-localization tools in mobile devices, we are walking towards a future in which the storing of the geographical coordinates where each datum is collected will be possible (Larraz, Pavía and Ferrari, 2013).

The above arguments may lead to conclude that no additional research in this area is already 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 for 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 just microdata, do not exist. Moreover, due to confidentiality issues, average analysts hardly ever have access to the microdata files, 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. Indeed, the research in this area has been fertile during 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 and Johnston, 2000) and electoral pollsters (e.g., Pavía and 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 must face.

For the general problem of population reallocation, a large number of methods have been suggested in the literature. They have evolved from simple areal weighting procedures (e.g., Goodchild and Lam, 1980) and point-based areal interpolation methods (e.g., Fisher and Langford 1995) to more complex dasymetric mapping strategies. Dasymetric mapping (Wright, 1936) is a spatial 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 conceived as a disaggregation method, it is employed in the reallocation problem through a two-step process in which first source data is spatial refined to then be aggregated to the

target units. In the refining process, they have been used among other auxiliary variables: land uses (Mennis, 2003; Giordano and Cheever, 2010, night-time lights and road networks (Reibel and Bufalino, 2005), the spatial distribution of built structures (Maantay, Maroko and Herrmann, 2007), residential building volumes (Sridharan and Qiu, 2013), satellite images (Holt, Lo and Hodler, 2004), volunteered geographic information (Bakillah et al., 2014), mobile phone data (Deville et al., 2014), and a combination of spatial methods and Maximum Entropy or the Expectation-Maximization (EM) algorithm (Sridharan and Qiu 2013; Schroeder, and Van Riper, 2013; Buttenfield, Ruther and 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 whether and to what extent a more complex approach is worth it 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 being more robust and accurate for the spatial reallocation of census data, we explore here whether this superiority is maintained for voting data. The comparison between approaches is customary (see, e.g., Ruther, Leyk and Buttenfield, 2015; Buttenfield, Ruther and Leyk, 2015; Pavía and Cantarino, 2016) an relevant because depending on the final objectives of the redistribution process, the analyst should weight if it is worth enough to apply 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 shape 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 and Pattie, 2006; Pavía, Larraz and Montero, 2008), election results show geographical structure and spatial patterns.

Barcelona in Spain, a polycentric and spatially complex city (Catalán, Saurí and Serra, 2008) with a marked social structure (Broner, 2010) and more than 1.6 million of 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 1,482 to 1,061 (defining a total of 6,329 different intersection polygons). Every four years local elections are held in Spain to elect mayors. Therefore, having

available (an approximation of) the results recorded in the 2007 local election in the census sections (the smallest geographical units for which votes are declared in Spain) corresponding 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 stated 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 to four dasymetric procedures that have proven to be superior to the simplest methods using exactly the same breakdowns for the customary problem of reallocating census data (Pavía and Cantarino, 2016). 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 approaches 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 are common interpolation methods (point-based spatial interpolation and areal interpolation without ancillary data, or areal weighting), classical and quite simple. They are used as baseline to gauge the value of four other alternatives based on dasymetric techniques, and therefore more complex (areal interpolation with ancillary data). We have followed Pavía and Cantarino (2016) to choose the dasymetric methods considered in this research. For exactly the same breakdowns analysed in this paper, Pavía and Cantarino (2016) 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 interpolation 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 Naciona*l, we have downloaded the 2009 Land Cover and Land Use Information System of Spain (SIOSE), which is a unique database of Spain produced at 1:25,000 scale combining topographic maps, satellite imagery, aerial photography and cadastral registers, with a minimum mapping unit (MMU) of 1 ha for the urban fabric areas. This dataset contains information about heterogeneity of land use within any given polygon in Spain (Goerlich and Cantarino, 2013). SIOSE comprises 2.5 million polygons with nearby 820,000 different land cover categories, obtained after combining with different weights the land cover elements (Cantarino *et al*. 2014). In the SIOSE model each polygon is defined by a land cover. Each land cover can be simple (uniform over the entire polygon) or composite. A variety of simple or even composite covers within a polygon defines a composite coverage (Cantarino *et al*., 2014). Four different attributes are used to identify types of inhabited buildings in SIOSE. They correspond to compact apartment blocks, isolated apartment blocks, terraced houses, and detached houses.

From the European Environment Agency (EEA), we have downloaded the Urban Atlas database that provides pan-European comparable land use and land cover data for Functional Urban Areas. Its date of production is 2009 and it is available at a 1:10,000 scale with a position accuracy of +/-5 m. We have used the layer "Urban Fabric" from this database. This layer classifies the polygons into categories according to their percentage of soil sealing or sealing levels. Assuming that the larger the percentage of soil sealing in a polygon the larger the percentage of residential area in the polygon is, we have set as proportions of residential areas in each polygon the average of soil sealing of its class: 0.05, 0.20, 0.40, 0.70 and 0.90, respectively, for categories "Discontinuous Very Low Density Urban Fabric", "Discontinuous Low Density Urban Fabric", "Discontinuous Medium Density Urban Fabric", "Discontinuous Dense Urban Fabric" and "Continuous Urban Fabric" and "Isolated Structures".

From *Dirección General del Catastro*, we have downloaded the Spanish cadastral database composed of shape and ASCII (*.cat) files. This administrative register contains the description of all the rustic, urban and special feature properties in Spain. The shape files define the boundaries of each cadastral parcel. The ASCII files contain exhaustive information about the use (residential or other types), year of building and living area per dwelling.

2.2. Reallocating approaches

This subsection offers the details of the six reallocating procedures considered in our scrutiny. The point-based areal interpolation and areal weighting solutions just require the files providing the polygon attribute tables of the spatial breakdowns to be executed. As point-based procedure (M1), we have used natural neighbour (NN) interpolation (Sibson, 1981), as default in ArcGis® 10.2, to derive the density surface. The point-based method does not verify the pycnophylactic condition of volume preservation (Tobler, 1979). All the other five approaches implemented are volume-preserving and share approach. All of them use as allocation function an estimator of the form given by equation (1), where $\hat{V}_{jk}$ are the allocated votes for party k in target unit j, V_{ik} are the recorded votes for party k in source unit i, N is the total number of source units and w_{ij} is the weight assigned to source unit i in estimating the votes of target unit j. The methods just differ in the way the weights w_{ij} are calculated.

We follow the classical terminology (Goodchild and Lam, 1980) of calling source units (or polygons) to the units in which the variable of interest is available and target units to the units where the reallocated data are required.

$$\hat{V}_{jk} = \sum_{i=1}^{N} w_{ij} V_{ik} \qquad (1)$$

The areal weighting approach (M2), also known as polygon overlay (Markoff and Shapiro 1973), estimates the values in the target areas using a geometric approach. It handles source and target polygons as geometric figures and computes weights, w_{ij}, as the ratio between the area of the intersection between source unit i and target unit j and the total area of unit i. This approach assumes that the variable of interest is distributed uniformly within each source unit and that source polygons with null intersection have no effect on the predicted value. A total of 6,329 polygons were obtained by intersecting 2007 and 2011 polygons.

Areal weighting and point-based interpolation are simple methods. They not require any additional geographical data besides source and targets polygons. Dasymetric mapping adds additional information that provides insights on how the population is distributed within each unit of analysis. The most popular methods employ two-dimensional (2-D) areal measures to locate where population lives in each source unit. From richer databases, three-dimensional (3-D) volume measures are also computed. There are also examples in the literature using one-dimensional (1-D) length measures (e.g. the street network in Reibel and Bufalino, 2005).

The simplest 2-D dasymetric mapping methods are based on a binary classification of land uses: residential and nonresidential. In the binary approach, nonresidential land use areas are observed as unpopulated and consequently as zero weighted. In the SIOSE approach (M4), we use SOISE residential areas to compute weights. In a similar fashion that in areal weighting, source and target units are overlaid but restricted to residential areas and weights, w_{ij}, are determined by the ratio between the corresponding area of the intersection between the source unit i and the target unit j and the residential area of unit i.

Land use data are useful to separate residential and non-residential areas, but they do not discriminate by residential attributes. In 2-D polycategorical dasymetric procedures, residential areas are grouped by categories and a different weight per area unit is assigned to each type attending to their relative population density. In the Urban Atlas procedure (M5), we use the five Urban Fabric categories defined in the Urban Atlas database using the average of soil sealing of the corresponding class as density (see below). In particular, denoting by d_t the relative soil sealing surface of residential polygons of type t and by w_{ijt} the area of the intersection among the source unit i, the target unit j and the residential polygons of type t, we compute the w_{ij} weights dividing $\sum_{t=1}^{5} d_t w_{ijt}$ by $\sum_{t=1}^{5} d_t w_{it}$, where w_{it} is the area of the intersection between the source unit i and the residential polygons of type t.

As alternative to 2-D polycategorical dasymetric procedures, the 3-D approaches consider the vertical distribution of the population. 3-D methods account for the height, volume or total area of the residential buildings. In the Cadastral method (M6), we compute weights exploiting the more detailed information available in Spain about households. Using the information about the postal address, floor, uses and square meters of each existing property in Spain, we compute in each cadastral polygon its total housing area to then calculate the w_{ij} weights as the ratio between the total housing area placed within the intersection of the source unit i and the target unit j and the total housing area of unit i.

For the sake of completeness, we have also implemented a 1-D procedure. In 1-D length methods, the density of population across a polygon is assumed to be directly related to the density of road, streetlights or power network segments across the polygon. In the Callejero method (M3), we have used the street and road network available in *Callejero*, and after constructing a buffer of 10 meters centered in each road/street line we have proceeded as in the 2-D binary method (SIOSE approach) by viewing the buffer polygons as residential areas.

All the spatial computations have been completed in ESRI ArcGIS® 10.2 (ESRI, 2014), using its geo-processor ArcPy to create scripts. The calculation routines have been performed with Python 2.7.3 (Python Software Foundation, 2014).

3. The secret ballot challenge

Once votes (including abstentions) have been reallocated from source units to target units using whatever of the techniques described in section 2, a problem emerges when we try to assess the different approximations. The votes (proportions) imputed to each target unit refer to an estimate of which would have been its 2007 local election outcomes if it had existed at that time. However, the observed outcomes available for target units are those recorded in the 2011 local election and they cannot be directly compared to the 2007 imputed values because, even assuming stationary electorates, there are always swings from one party or candidate to another between elections. Likewise, due to secret ballot, individual votes are not available and therefore a standard of comparison cannot be obtained from single records.

Fortunately, both over time and around the world, there is evidence that strong relationships persist among polling units in the outcomes harvested by the different electoral options in consecutive elections (see, e.g., Mosteller *et al.,* 1949; Pavía-Miralles, 2005; Curtice et al., 2011). On the one hand, in voting units where, say, Republican support is relatively high in an election, this will still be relatively high in the next election. On the other hand, the so-called swings, or vote changes between options that occur between elections, tend to be spread throughout the set of voting units, with quite stable constant rates of change among them.

These patterns also arise in our case of study, as can be observed in Figure 2. Our strategy to construct a standard of comparison is based on exploiting these patterns. To do

that, however, we need (i) to specify a statistical model that permits us to *transfer* the 2007 imputations to 2011 and (ii) to decide how to estimate fairly the model in order to apply it without favouring any particular approach. Then, the closer the transferred allocations to the actual recorded values are, the better the spatial imputation approach will be considered.

In subsection 3.2 we revise the relevant literature and specify a model that relates current and past vote proportions. In subsection 3.1 we compare the geography of 2007 units and 2011 target units to search for a set of unchanged census sections that (i) assist us in the process of selecting a proper model and, mainly, (ii) can be used to estimate the parameters of the transfer model using their actual outcomes recorded in the 2007 and 2011 Barcelona local elections.

3.1.   *Equivalent polygons*

Despite the profound reorganization of census sections that Barcelona's cartographic authorities performed in 2009, some of the sections remained basically unchanged. Identifying what they are is interesting for assessing purposes. Unfortunately, almost no polygon remained exactly the same between 2007 and 2011. Polygon boundaries are usually placed on streets and some small shifts just occurred by a slight translation of the axis within their surface without entailing any population shift. We consider two sections as equivalent when they keep the same residential buildings.

Deciding if two polygons are equivalent is not an easy task (Kidner, 1996) and, in our case, identifying them is hampered by the small shifts faced by the boundaries of the equivalent polygons. For example, for a rectangular polygon of one hectare, a displacement of just one meter in every direction could result in a variation in surface up to 4%. Therefore, whatever the measures of comparison chosen, some margin of discrepancy should be allowed.

Given the large number of polygons to be compared and the computational cost that would entail making over than 1.5 million of comparisons, as initial screening we consider two polygons as potentially equivalent when the distance between their centroids is less than 50 meters. This is a really conservative condition that guarantees the detection of any pair of equivalent polygons. Once we have a set of potential equivalent polygons, to resolve their equivalence a battery of six measures of similarity has been calculated for each potential pair, (S_i, T_j), where S_i refers to source polygon j and T_j to target polygon j: (i) the rate of total surfaces, $\Delta(S_i)/\Delta(T_j)$; (ii) the rate of residential surfaces, $\Delta(S_i \cap RS)/\Delta(T_j \cap RS)$; (iii) the rate between the 2011 surface and the intersection of 2007 and 2011 surfaces, $\Delta(T_j)/\Delta(S_i \cap T_j)$ (always ≥1); (iv) the rate between the 2007 surface and the intersection of 2007 and 2011 surfaces, $\Delta(S_i)/\Delta(S_i \cap T_j)$ (always ≥1); (v) the rate of elongation shape factors, $\Delta(S_i)E(T_j)^2/\Delta(T_j)E(S_i)^2$, where $E(S)$ denotes the maximum elongation of surface S obtained from its convex hull and $1.27\Delta(S)/E(S)^2$ is unitary for circular polygons; and (vi) the rate of circularities, $\Delta(S_i)P(T_j)^2/\Delta(T_j)P(S_i)^2$, where $P(S)$ is the perimeter of surface S and

$4\pi\Delta(S)/P(S)^2$ is unitary for circular polygons. Two polygons are considered equivalent if all the six measures are inside the range $[0.95, 1.05]$.

A total of 122 polygons were identified as equivalent. As example, the zone around Catalonia Square, the one with the highest density of equivalent polygons, is depicted in Figure 1. The polygons identified as equivalent in both breakdowns have been shaded. The values recorded for the relevant variables of 2007 and 2011 Barcelona local elections in the equivalent polygons are likewise portrayed in Figure 2. The strategy followed to identify the equivalent polygons is reinforced by the existence of strong statistical patterns in Figure 2.

Figure 1. Extract of Barcelona (Spain) division in census sections during the 2007 (left panel) and 2011 (right panel) local elections. The same zone (area around Catalonia Square) depicted in both figures. The polygons identified as equivalent are shaded in both panels.

3.2.   A temporal transfer model

Twenty-two and twenty-five parties presented candidatures in Barcelona in, respectively, 2007 and 2011 local elections. Among them, only five parties or coalitions—CiU, a right-wing coalition of regional parties; PSC, the Socialist Party; PP, the Conservative Party; ICV, a national coalition of left-wing; and, ERC a left-wing coalition of regional parties—surpassed in both elections the 5% threshold to reach representation. We focus exclusively on these parties and on an additional artificial party (Others), which adds the remaining voted options, including null and blank votes.

As can be observed in Figure 2 and in line with previous findings in the literature, strong relationships link the outcomes of 2007 and 2011 Barcelona local elections at census sections. Strong linear patterns relate turnout rates and the proportions of votes gained by the main political parties. The results recorded for ICV (a left-wing national coalition) and ERC (a left-wing regional coalition) show however a volatility relatively higher around their underlying straights. This seems to be a reflection of both statistical and political issues. On the one hand,

these proportions are closer to zero. On the other hand, these parties have strategies of coalitions more unstable and electorates showing more pronounced tactical behaviours. Moreover they compete in more crowded electoral spaces. For example, in the 2007 election ERC concur in coalition with just AM. However, in 2011 and as a response to the rise of new political actors (SI, CUP, Ri.Cat, DCat) contesting its electoral space, ERC extended its strategy of agreements and presented candidature together with more partners.

Despite the strong relationships displayed in each panel of Figure 2, it is not possible simply to consider separate regression models for the change in each party's share of the vote without violating the sum-to-1 constraint. It is clear that, within any individual section, the sum of the share of the votes is unitary and that likewise the outcomes of each of the parties are not independent of the results gained for the rest of parties. Therefore, the model proposed should be multivariate and guarantee the sum-to-1 constraint.
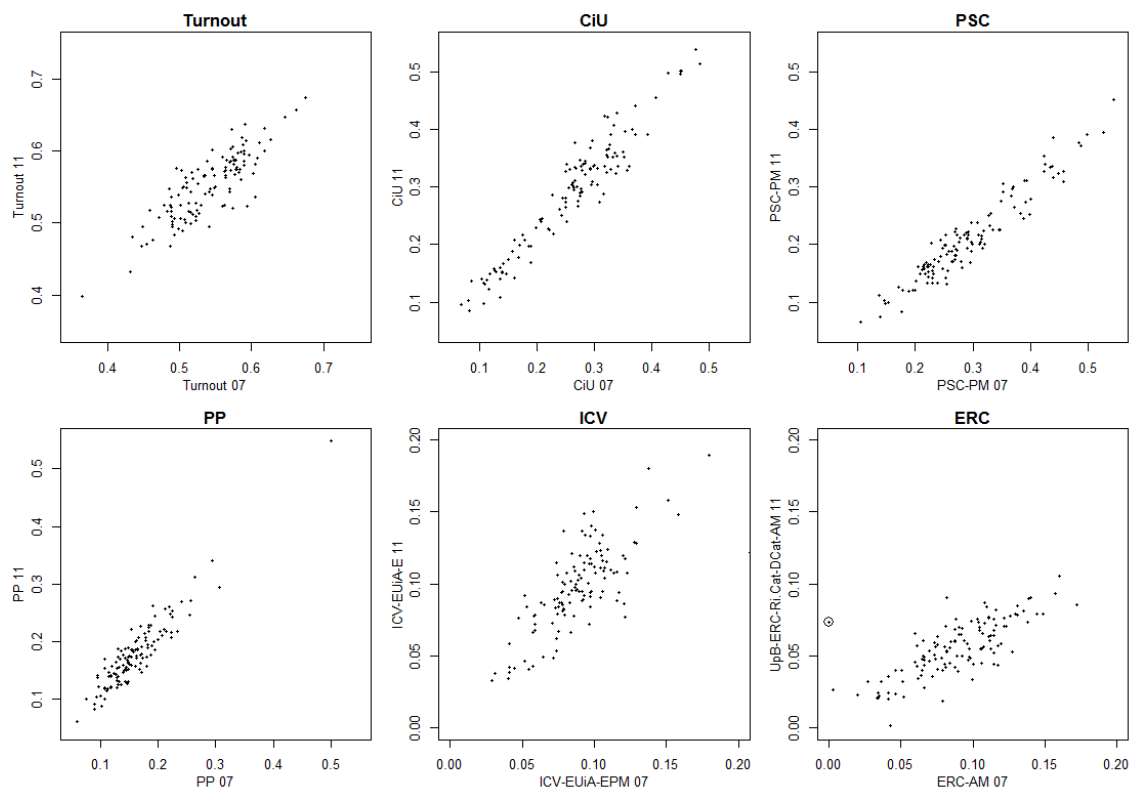

Figure 2. Relationships between actual outcomes recorded during 2007 (horizontal axes) and 2011 (vertical axes) Barcelona local elections in the equivalent census sections. The variables drawn are turnout rates (upper left panel) and proportions of votes gained by the five parties that reached representation: CiU, a right-wing regional coalition (upper middle panel); PSC, the Socialist Party (right upper panel); PP, the Conservative Party (lower left panel); ICV, a left-wing national coalition (lower middle panel); and, ERC, a left-wing regional coalition (lower right panel). Highlighted with a circle an outlier with significant leverage. (Source: Own elaboration from Spanish Home Office data, http://www.infoelectoral.interior.es/min/areaDescarga.html).

Different proposals can be found in the literature. Pavía-Miralles and Larraz-Iribas (2008) suggested (i) using an independent model for turnout rates and (ii) jointly modelling

proportions of votes by a multivariate linear Gaussian model with zero-mean-distributed random disturbances and correlations constant between parties and independent between polling units. Bernardo (1997) advocates for the superiority of working with logit transformations and, in a Bayesian framework, proposed (i) employing logit transformations for both current and past proportions and (ii) using directly previous recorded turnouts as weights. And Curtice and Firth (2008), in a context of exit polling, employed a linear hierarchy 'nested' regression for modelling the change in each party's share of the vote. They also assume no change in the pattern of turnout.

Inspired by the above three models, we propose to use a two-block model to temporally transfer the spatial imputations; a route which incidentally mimics the decision-making process of voters. First electors decide on whether or not to vote and then for which to vote (Aldrich, 1997). A model is suggested to time transfer turnout rates and a separate model is proposed to transfer share of votes. In this way, for each individual section, the product of (i) the people entitled to vote in the target section, (ii) the transferred turnout and (iii) the vector of the transferred share of the vote leads to a time transferred prediction of the vector of votes gained in the section by the different parties. This can be compared to the outcomes actually recorded in the census section and aggregating by sections also compared to the results of the whole constituency.

To guarantee the congruence of the transferred proportions, we consider a binomial response for the number of voters and a multinomial response for the vector of votes. As first block, we employ a logit regression to transfer turnouts. The logit transformation of previous turnout is used as predictor. In the second block, we propose a multinomial logistic regression model to transfer vote proportions, using the logit transformations of previous vote proportions as predictors.

Mathematically, the first equation states that:

$$logit(t_1) = \alpha_t + logit(t_0)\,\beta_t \qquad (9)$$

where, for a target section, t_1 denotes the turnout registered during 2011 election, t_0 the turnout in the section in 2007 and α_t and β_t are unknown parameters.

On the other hand, the multinomial logistic regression transfer model can be stated as:

$$log\frac{\pi_{i1}}{\pi_{I1}} = \alpha_i + logit(\pi_{10}, \pi_{20}, \ldots, \pi_{i0}, \ldots, \pi_{I0})\,\beta_i' \quad i = 1, 2, \ldots, I-1 \quad (10)$$

where $(\pi_{11}, \pi_{21}, \ldots, \pi_{i1}, \ldots, \pi_{I1})$ is the vector of share of the votes recorded in 2011, $logit(\pi_{10}, \pi_{20}, \ldots, \pi_{i0}, \ldots, \pi_{I0})$ is the vector of logit transformations of the vote proportions recorded in the same section in 2007, and the (α_i, β_i) are unknown parametric vectors of length $I + 1$, being $I = 6$ the number of main parties (CiU, PSC, PP, ICV, ERC plus Others, an artificial party grouping the rest of voted options).

We use the actual values observed in 2007 and 2011 in the equivalent sections (excluding the section for which the outlier identified in the lower right panel of Figure 2

belongs to) to fit the models and estimate the parameters by maximum likelihood. The electoral censuses, EC, of the target sections are the sizes of the binomial distributions and the number of votes registered, V, the sizes of the multinomial distributions. In particular, it is assumed that $V_j \sim Bi(EC_j, t_{1j})$ and $(V_{j1}, \ldots, V_{jk}, \ldots, V_{jI}) \sim Mn(V_j, \pi_{11j}, \ldots, \pi_{i1j}, \ldots, \pi_{I1j})$, where j indexes the target section and V_{jk} are the votes gained by party k in the section j. Once estimated, the parameters of equations (9) and (10) were applied to each set of spatial imputations to translate them to 2011. All the computations have been performed in the version 3.3.1 of free statistical software R (R Core Team, 2016) with the assistance of versions 0.9-1 and 7.3-6 of, respectively, packages VGAM (Yee, 2013) and nnet (Venables and Ripley, 2002).

As a matter of assessing the transfer strategy, we have also applied it to predict 2015 elections from 2011 outcomes, using the equivalent sections to estimate the parameters of the corresponding model. Between 2011 and 2015, there were almost no geographic shifts in the boundaries of Barcelona section breakdown, what makes it unnecessary to reallocate votes, but a political earthquake took place in Spain (see, e.g., Pavía, Bodoque, and Martín, 2016), with some new strong parties emerging. Despite this, the robustness of the transfer strategy is manifested in the fact that (i), at the aggregate level, the approach is able to generate forecasts three times more accurate than when (the intermediate step of) reallocating votes is necessary and (ii), at the level of sections, it produces estimates in average two times more precise.

4. Assessing imputations

Once the six sets of spatial allocations corresponding to the values recorded in 2007 were temporally translated to 2011, we carry out several analyses to assess their relative performance. The time transferred imputations have been compared, both separately at section level and jointly aggregated, to actual results. In this section, we just present the outcomes of reallocating votes in absolute numbers, very similar results are obtained when we focus on reallocating proportions.

At census section level, we performed χ^2-goodness-of-fit tests to decide whether the transferred estimates can be observed as random variations of the actual data. Likewise, the 2011 actual and time transferred proportions are graphically compared in Figure 3 for all the spatial approaches. At 1% significant level, the percentage of census sections for which the imputed votes can be considered as generated by the actual section distributions reached 75.4, 78.7, 79.0, 79.7, 79.2, and 79.8 for, respectively, NN point-based areal interpolation (M1), areal weighting (M2), Callejero (M3), SIOSE (M4), Urban Fabric (M5) and Cadastral (M6) approaches. Despite the differences between the approaches being not too high, a clear order emerges among the spatial approaches considered: M6 ≻ M4 ≻ M5 ≻ M3 ≻ M2 ≻ M1.

Although null hypothesis aceptance rates around 79% (which drop to around 55% and 65% for, respectively, 10% and 5% significant levels) can lead to consider imputations as not

sufficiently suitable, it should be noted that a portion of the deviations between actual results and approximations are not attributed to the spatial approach followed to allocate votes. Rather, it is a consequence of (i) the obvious fluctuations that naturally happen between elections and of (ii) the statistical model specified to transfer proportions temporally. In fact, we find that the rates of null acceptance for the time transferred votes in the equivalent sections are 58.3%, 70.8% and 85.8% for, respectively, 10%, 5% and 1% significant levels, despite them having been used to fit equations (9) and (10). On the other hand, as we will discuss in the last section, the closeness between temporal transferred allocations and actual results also depends on a part of the model specified.

In many applications more important than closely approximating actual values is to dispose of a variable really alike in correlation terms. This variable could be used in place of the unobserved variable in a regression. Overall, almost all the approaches produce transferred imputations closely correlated to actual figures. The correlations fluctuate between a minimum of 0.9694 for NN interpolation and a maximum of 0.9724 for Cadastral dasymetric. Certainly, the graphical comparison in Figure 3 shows a good level of agreement for all the approaches between actual proportions and estimates.
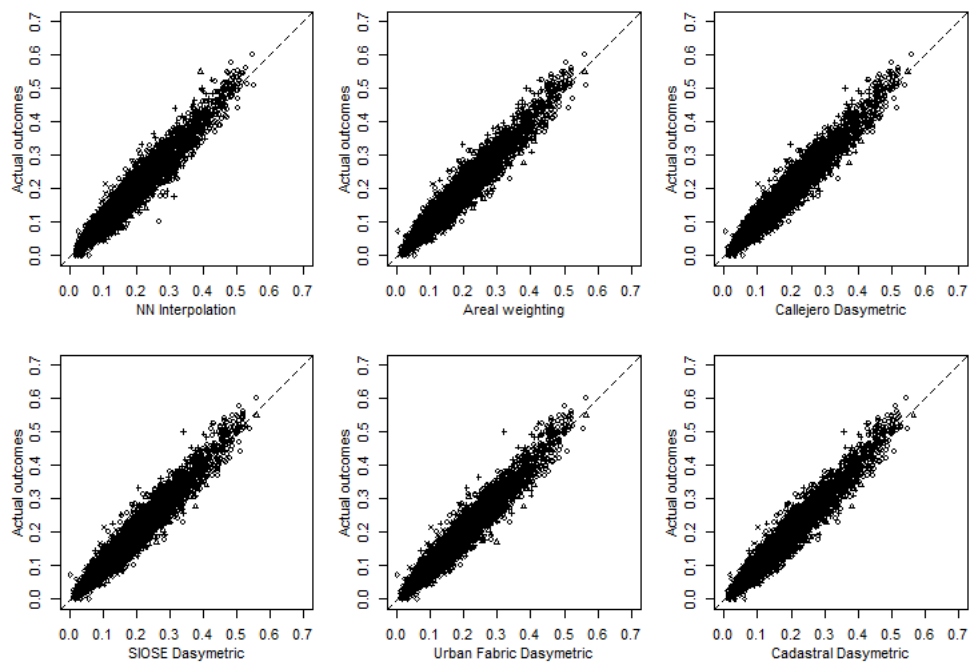

Figure 3. Comparing real proportions (vertical axes) and time transferred imputations (horizontal axes) at census section level for Barcelona (Spain) 2011 local elections. From left to right, top to bottom: NN interpolation, areal weighting, Callejero, SIOSE, Urban Fabric and Cadastral approaches. The distance from the 45º line indicates how far apart time transferred allocations and outcomes are. The number of data points in each scatterplot is 6,366 (1,061 sections per 6 political options: ○ right-wing coalition of regional parties (CiU), + the Socialist Party (PS), Δ the Conservative Party (PP), × national coalition of left-wing parties (ICV), ◊ left-wing coalition of regional parties (ERC), and ∇ Others (OT)). (Source: Own elaboration from Spanish Home Office data, http://www.infoelectoral.interior.es/min/areaDescarga.html).

Regarding the comparison in aggregated terms, Table 1 presents the combined results obtained after assigning a 2007 vote history to each target section of the city of Barcelona and transferring them to 2011 using the two-block process specified by equations (9) and (10). As can be observed, despite the swings registered between 2007 and 2011 Barcelona local elections, the forecasts obtained with all the spatial approaches are very accurate. All the aggregate party approximations obtained are quite precise, the aggregate global fit being impressive. The average absolute error per party is just around 0.1%, as clearly shows the computed $Error = \frac{100}{6}\sum_{i=1}^{6}\left|\hat{\pi}_i^X - \pi_i\right|$ statistic—where $\hat{\pi}_i^X$ represents the aggregate proportion of votes estimated for party i using spatial approach X (with X = M1, M2, M3, M4, M5, M6) and π_i the actual party i share of the vote. Table 1 moreover shows the average of the χ^2 distances, computed at census sections, between recorded votes and time-transferred allocations.

Table 1. Estimates for 2011 Barcelona local election after time transferring spatial imputations.

Approach[a]	urnout	Percentages[b]						ean[d]		Councillors				
		iU	SC	P	CV	RC	thers	rror[c]	χ²	iU	SC	P	CV	RC
2007 results	9.62	5.32	9.75	5.52	.30	.77	1.34	.75		2	4			
M1	1.95	8.14	1.70	7.21	0.14	.45	7.36	.10	4.41	4	1			
M2	2.32	8.05	1.62	7.30	0.17	.46	7.40	.14	3.04	4	1			
M3	2.33	8.04	1.62	7.31	0.17	.46	7.40	.14	3.00	4	1			
M4	2.35	8.02	1.64	7.32	0.17	.46	7.39	.15	3.13	4	1			
M5	2.30	8.06	1.63	7.29	0.17	.46	7.39	.14	2.99	4	1			
M6	2.35	8.02	1.64	7.31	0.17	.46	7.40	.14	2.91	4	1			
2011 results	2.99	8.24	1.77	6.95	0.21	.50	7.33			4	1			

[a] M1: NN point-based areal interpolation, M2: areal weighting, M3: Callejero dasymetric, M4: SIOSE dasymetric, M5: Urban Fabric dasymetric, and M6: Cadastral dasymetric.

[b] Result rows show the proportions of votes won by each political party. Approach rows portray the forecasts that would have been obtained after transferring the corresponding 2007 imputation to 2011.

[c] The error column displays the average sum of the differences in absolute values between percentages of predictions and actual outcomes: $Error = \frac{100}{6}\sum_{i=1}^{6}\left|\hat{\pi}_i^X - \pi_i\right|$.

[d] This column shows the average of the χ^2-distances in each census section between its actual votes and its corresponding time-transferred allocations.

As can be inferred, the average distances measured using the χ^2-statistic show an order of preference among the spatial approaches analysed that as a rule favours dasymetric strategies (M6 > M5 > M3 > M2 > M4 > M1). As expected, the introduction of more information in the allocation process tends to produce better imputations. The differences between the approaches is nevertheless really small and the increases of accuracy obtained with the more complex approaches compared to the baseline methods (areal interpolation

and areal weighting) are dubious enough to reward the extra effort that dasymetric approaches require.

In the presentation of the results performed so far, we have not yet tackled the spatial analysis of the reallocation errors. The issue is that as a rule no great differences exist among the spatial distributions of the different errors attained. The more relevant result is that, as expected, the relative worse performance of the NN point-based areal interpolation approach at census level has its origin in the relative worse imputations that this method produces in the border sections. In any case, the gains in accuracy of dasymetric refinement regarding to areal weighting are, for the example here studied, just marginal. More likely, the great similarity that areal weighting and dasymetric methods show is a consequence of the uniformity in the distribution votes in neighbouring areas ("the homogeneity assumption" cited in Goodchild & Lam, 1980). For instance, the spatial autocorrelations for the two big political parties in 2007, measured on the Moran index of the share of votes, reach 0.50 for CiU (p-value ¼ 0.0000) and 0.48 for PSC (p-value ¼ 0.0000).

As an example of a spatial distribution of errors, Fig. 4 shows the relative absolute errors obtained at census section level using the dasymetric Cadastral approach. The one producing the better allocations. The general picture displayed for this approach is common to all the other procedures. As we can observe, the errors present a peripheral or radial distribution, which defines a statistically significant non-random spatial distribution, although probably not significant in practical terms (Gelman and Carlin, 2017
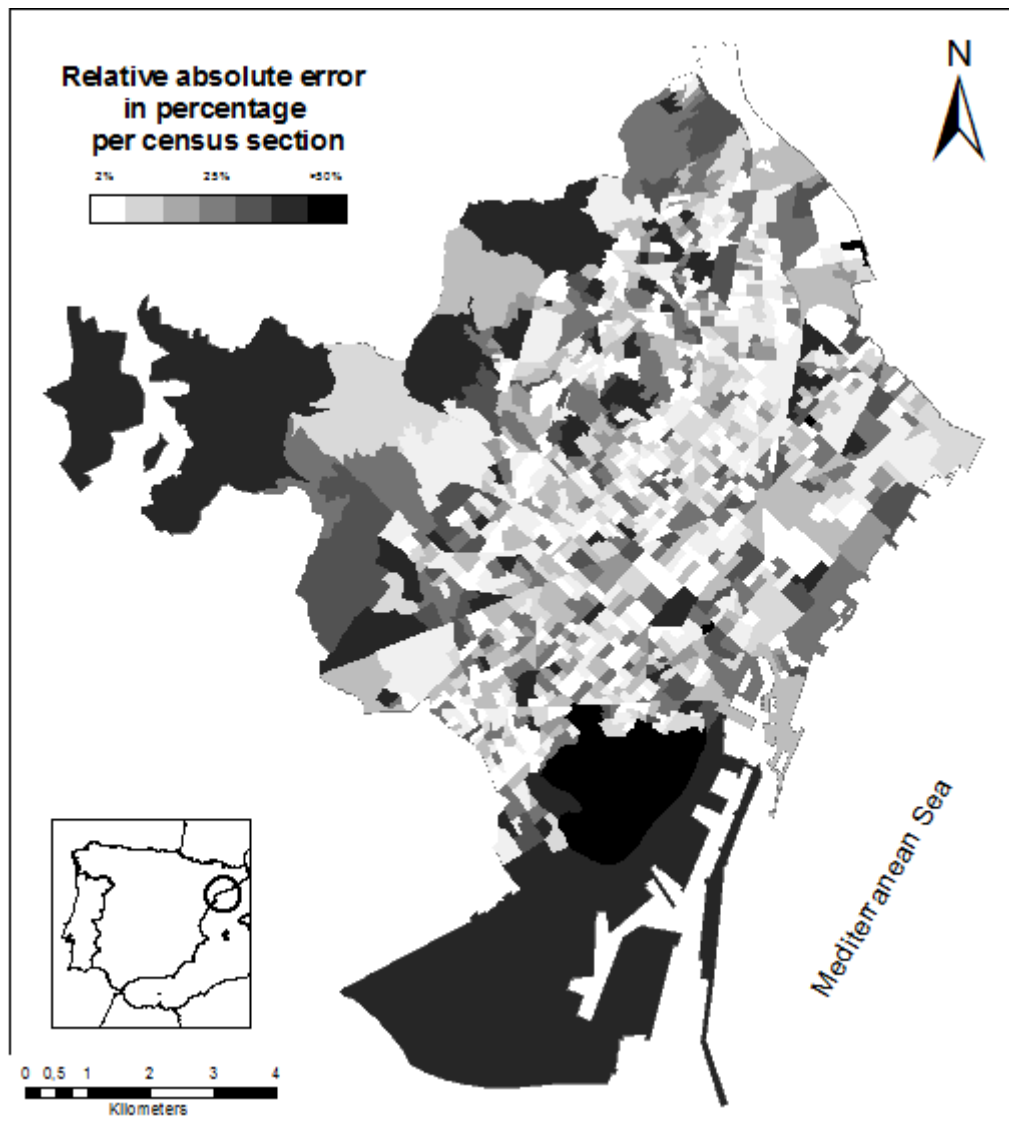
Figure 4. Spatial distribution of relative absolute errors at census section level after time-transferring the reallocated absolute number of votes using dasymetric cadastral-based areal interpolation.

Indeed, the Moran indexes of the errors just fluctuate between 0.04 and 0.05; although they increase to 0.09 if only the city centre is considered. In summary, given that no large differences have been found in the census section errors among the different methods (excluding NN spatial interpolation), we could conclude that their origin comes more from the fluctuations of political tendencies themselves than on the spatial methods, all of which produce really similar allocations. An analysis of the differences/similarities among the allocations obtained before time-transferring them indeed corroborates this idea.

5. Discussion and final remarks

Many social statistics are offered in a really small-area basis, like census blocks or precincts. The successive computation revolutions, with their subsequent improvements in processing power and storage capacity, have enormously eased their inclusion in an increasing number of economic and social applications, from determining locations of new business settlement to support decision making in government programs. These are not obviously the only issues for which they can be of interest. They are also useful for drawing electoral boundaries, assessing public program effectiveness, offering litigation support, monitoring infectious diseases, determining consumer profiles or deciding sales forecasting, to name but a few.

Small areas are however more exposed than established big areas to experiencing changes in their boundaries. Therefore, it is not uncommon that most frequent breaks occur in their series of data, making their inclusion in dynamic analysis difficult. Hence, many methods have been proposed in the literature to try to allocate data between two spatial divisions of the same territory: from a set of source units to a set of target units.

Given that a precise location of population is the key to answering many practical questions of social interest, the introduction in the reallocation process of new ancillary data through dasymetric techniques is the route commonly followed to improve spatial allocations. However, according to Zandbergen and Ignizio (2010, p. 199) "the performance of dasymetric methods varies substantially among study areas, and no single technique consistently outperforms all others." In this paper, we take as a case of study the city of Barcelona in Spain, a big city with more than 1.6 million of inhabitants that has recently redrawn its small areas' boundaries, and we analyse how different reallocation approaches based on dasymetric mapping perform on the issue of redistributing votes. The problem selected is particularly interesting because a solution based on individual records is unimaginable for this kind of data.

Although in urban areas and mainly in big cities, it is expected that dasymetric approaches shows limited improvement in accuracy over simple areal weighting approaches, in a recent paper Pavía and Cantarino (2016) have shown for the same breakdowns handled in this research that dasymetric refinement clearly beats simple methods for the customary problem of reallocating census figures. This paper analyses whether this is also true for the case for reallocating votes.

As expected, at the small unit level our analyses show that the approaches based on dasymetric techniques outperform the simpler solutions of point-based areal interpolation and, in a lesser extent, areal weighting. The improvements can be labelled as marginal, however. There are no great differences between the accuracy of the approximations achieved with NN point-based areal interpolation and areal weighting and those reached with other more complex techniques. The biggest differences are concentrated for the NN point-based areal interpolation approach in the

border sections, where the NN method harvests its worst allocations. In light of our results, we doubt it worth the extra effort of geo-locating population more precisely introducing ancillary sources of information in the process of reallocating votes (or share of votes) in an urban area, at least when votes show strong spatial patterns in cities with a relatively compact population structure. Indeed, in response to Pavía and López-Quilez (2013) inquiry stated in the introduction, our recommendation would be not to complicate things and to rely on the simplest methods (areal weighting), mainly when the aim of the reallocation process is aggregate forecasting. For other kind of problems, such as ecological inference (see, e.g, King, 1997), when disposing of accurate small area reallocations is certainly quite relevant, it would be interesting to explore other approximations. It could be interesting to include in a dasymetric procedure to reallocate votes other variables that could be available at census level more closely related to the individual determinants of vote, such as the average income or the structure by ages of the population.

Despite the geographical literature showing as a rule more sophisticated and data-demanding methods producing better allocations (e.g., Sridharan and Qiu 2013; Ruther, Leyk and Buttenfield, 2015; Buttenfield, Ruther and Leyk, 2015; Pavía and Cantarino, 2016; Mennis, 2016), the results of this research come to point out some limits to the improvements that can be reached with more complex dasymetric procedures. The tradeoff cost-effectiveness of each solution is case-dependent: it depends on both the target variable and the particular geography. More research should be, therefore, performed to decide whether and when our conclusions could be extended to other instances. It should be interesting to know to what extent more complex methods improve allocations and what is their dependence on the type of variable to be allocated, the ultimate goals of the imputation process and/or the kind of urban structure under study. It is foreseeable that in other cities with more disperse populations (as a consequence of, for example, urban sprawl) and/or a fewest spatial autocorrelation between the political trends, the dasymetric methods will produce vote allocations with improvements relatively higher.

To end, we would also add some remarks obtained as by-product of our analyses about the specification and estimation of the temporal transfer model employed that could be of interest for pollsters and forecasters. In addition to the strategy presented in our exposition, we have also tested other specifications. In this sense, it should be remarked that equivalent results would have been obtained if (i) no logit transformations had been performed to the predictors, (ii) a joint model for turnout and share of votes had been used and/or (iii) the outlier section had also been included to fit the model. In all the cases, the spatial allocation approaches would have shown the same order of preferences, with similar global relative distances (although higher) among them. We have opted for presenting the specification producing the temporal translations globally closer to the actual outcomes given that the aim of the temporal transfer model is just to build a standard of comparison. During the process,

however, we have learned a couple of issues that we think could be of interest for political scientists, survey pollsters and electoral forecasters. First, it is considerably worse to model jointly turnout and vote proportions (i.e., consider abstention in the same level as voting for a particular party) than to use a two-block model. Second, the specifications employing logit transformations of the predictors are more sensitive to outliers and in their presence may lead to worse outcomes. In light of these two points, in a real situation in which the objective is predicting the constituencies' outcomes of an electoral process, we would recommend using the two-block equation approach with raw proportions as predictors.

References

Aldrich, J.H. (1997). When is it rational to vote? In *Perspectives on Public Choice: A Handbook*, (ed D. Mueller). Cambridge: Cambridge University Press, 373–398.

Bakillah, M., Liang, S., Mobasheri, A., Arsanjani, J.J and Alexander Zipf, A. (2014) Fine-resolution population mapping using OpenStreetMap points-of-interest. *International Journal of Geographical Information Science*, 28(9), 1940-1963.

Buttenfield, B.P., Ruther, M. and Leyk, S. (2015). Exploring the impact of dasymetric refinement on spatiotemporal small area estimates. *Cartography and Geographic Information Science*, 42(5), 449459.

Bernardo, J.M. (1997). Probing public opinion: the State of Valencia experience. In *Bayesian Case Studies* (eds C. Gatsonis, J.S. Hodges, R. E. Kass, R. McCulloch, P. Rossi and N.D. Singpurwalla). New York: Springer, 3–21.

Broner, S.J. (2010). *Análisis Espacial de Datos Electorales. Aplicación al Municipio de Barcelona*. Universitat Politécnica de Catalunya: PhD Dissertation.

Catalán, B., Saurí, D. and Serra, P. (2008). Urban sprawl in the Mediterranean? Patterns of growth and change in the Barcelona Metropolitan Region 1993–2000. *Landscape and Urban Planning*, 85(3-4), 174–184.

Cantarino, I., Torrijo, F.J., Palencia, S. and Gielen, E. (2014). Assessing residential building values in Spain for risk analyses – application to the landslide hazard in the Autonomous Community of Valencia. *Natural Hazards and Earth System Science*, 14(11), 3015–3030.

Curtice, J. and Firth, D. (2008). Exit polling in a cold climate: the BBC–ITV experience in Britain in 2005 (with discussion). *Journal of the Royal Statistical Society* A, 171(3), 509–539.

Curtice, J., Fisher, S.D. and Kuha, J. (2011). Confounding the commentators: How the 2010 exit poll got it (more or less) right. *Journal of Elections Public Opinion and Parties*, 21(2), 211–235.

Deville, P, Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D. and Tatem, A.J. (2014). Dynamic population mapping using mobile phone data. *PNAS*, 111(45), 15888–15893.

ESRI (2014). *ArcGIS Desktop: Release 10.2*. Redlands, CA: Environmental Systems Research Institute.

Fisher, P.F. and Langford, M.J. (1995). Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, 27(2), 211–224.

Giordano, A. and Cheever, L. (2010). Using dasymetric mapping to identify communities at risk from hazardous waste generation in San Antonio, Texas. *Urban Geography*, 31(5), 623–647.

Goerlich, F.J., and Cantarino, I. (2013). A population density grid for Spain. *International Journal of Geographical Information Science,* 27(12), 2247–2263.

Goodchild, M. and Lam, N.S.N. (1980). Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*, 1, 297–312.

Holt, J.B., Lo, C.P. and Hodler, T.W. (2004). Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science*, 31(2), 103–121.

Johnston, R. and Pattie, C. (2006). *Putting Voters in Their Place: Geography and Elections in Great Britain*. Oxford: Oxford University Press.

Kidner, D.B. (1996). Geometric signatures for determining polygon equivalence during multi-scale GIS update. In *Proceedings of Second Joint European Conference on Geographical Information*. Barcelona, Spain: IOS Press, 238–247.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.

Larraz, B., Pavía, J.M. and Ferrari, G. (2013). Weighting elementary prices in consumer price index construction using spatial autocorrelation. *Communications in Statistics-Theory and Methods*, 42(24), 4460–4475.

Maantay, J.A., Maroko, A.R. and Herrmann, C. (2007). Mapping population distribution in the urban environment: the Cadastral-based Expert Dasymetric System (CEDS). *Cartography and Geographic Information Science,* 34(2), 77–102.

Markoff, J. and Shapiro G. (1973). The linkage of data describing overlapping geographical units. *Historical Methods Newsletter,* 7(1), 34–46.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55, 31–42.

Mennis, J. (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass*, 3/2, 727–745.

Mennis, J. (2016). Dasymetric Spatiotemporal Interpolation. *The Professional Geographer*, 68(1), 92-102.

Mosteller, F., Hyman, H., McCarthy, P.J., Marks, E.S. and Truman, D.B. (1949). *The Pre-election Polls of 1948*. New York: Social Science Research Council.

O'Loughlin, J. (2002). The electoral geography of Weimar Germany: exploratory spatial data analysis (ESDA) of Protestant support for the Nazi party. *Political Analysis*, 10(3), 217–243.

Pattie, C. and Johnston, R. (2000) People who talk together vote together: an exploration of contextual effects in Great Britain. *Annals of the Association of the American Geographers*, 90, 41–66.

Pavía, J.M., Larraz, B. and Montero, J.M. (2008). Election forecasts using spatiotemporal models. *Journal of the American Statistical Association*, 103(483), 1050–1059.

Pavía, J.M. and Larraz, B. (2012) Nonresponse bias and superpopulation models in electoral polls. *Revista Española de Investigaciones Sociológicas*, 137, 237–264.

Pavía, J.M. and López-Quilez, A. (2013). Spatial vote redistribution in redrawn polling units. *Journal of the Royal Statistical Society A*, 176(3), 655–678.

Pavía J.M., Bodoque, A. and Martín, J. (2016). The birth of a new party: Podemos, a hurricane in the Spanish crisis of trust. *Open Journal of Social Sciences*, 4, 67-86.

Pavía J.M. and Cantarino I. (2016). Can Dasymetric Mapping Significantly Improve Population Data Reallocation in a Dense Urban Area? *Geographical Analysis*, online available.

Pavía-Miralles, J.M. (2005). Forecasts from non-random samples: The election night case. *Journal of the American Statistical Association*, 100(472), 1113–1122.

Pavía-Miralles, J.M. and Larraz-Iribas, B. (2008). Quick counts from non-selected polling stations. *Journal of Applied Statistics*, 35(4), 383–405.

Poulsen, E. and Kennedy, L.W. (2004). Using dasymetric mapping for spatially aggregated crime data. *Journal of Quantitative Criminology*, 20(3), 243–262.

Python Software Foundation (2014). *Python Language Reference, version 2.7.3*. http://www.python.org

R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/

Reibel, M. and Bufalino, M.E. (2005). Street weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37, 127–39.

Ruther, M., Leyk, S., and Buttenfield, B.P. (2015). Comparing the Effects of an NLCDderived Dasymetric Refinement on Estimation Accuracies for Multiple Areal Interpolation Methods. *GIScience & Remote Sensing*, 52(2), 158–178.

Schroeder, J.P. and Van Riper, D.C. (2013). Because Muncie's Densities Are Not Manhattan's: Using Geographical Weighting in the Expectation-Maximization Algorithm for Areal Interpolation. *Geographical Analysi*s, 45(3), 216–237.

Sibson, R. (1981) A brief description of natural neighbour interpolation. In *Interpolating Multivariate Data* (ed. V. Barnett). New York: Wiley, Ch. 2.

Sridharan, H. and Qiu, F. (2013). A Spatially Disaggregated Areal Interpolation Model Using Light Detection and Ranging-Derived Building Volumes, *Geographical Analysis*, 45(3), 238–258.

Tobler, W. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 519–30.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. 4th Edition. New York: Springer.

Wright, J.K. (1936). A method of mapping densities of population with Cape Cod as an example. *The Geographical Review*, 26(1), 103–110.

Yee, T.W. (2013). *VGAM: Vector Generalized Linear and Additive Models*. R package version 0.9-1. http://CRAN.R-project.org/package=VGAM

Zandbergen, P.A. and Ignizio, D.A. (2010). Comparison of dasymetric mapping techniques for small-area population estimates. *Cartography and Geographic Information Science*, 37(3), 199–214.