Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

# Asymmetrical-mapping based methodology: Constructing historical datasets for Portuguese census tracts

António Manuel Rodrigues amrodrigues@fcsh.unl.pt

Teresa Santos teresasantos@fcsh.unl.pt

Dulce Pimentel dpimentel@fcsh.unl.pt

e-GEO Research Centre for Geography and Regional Planning
Faculdade de Ciências Sociais e Humanas – Universidade Nova de Lisboa
Av. de Berna 26-C
1069-061 Lisboa - Portugal

**Abstract**

The measurement, identification and interpretation of spatial-temporal patterns are often dependent on the geometrical schemes used to divide land into discrete units/objects. Because anisotropy is a common characteristic of spatial surfaces, heterogeneity within artificial boundaries should be assumed. Hence, non-stationarity of datasets often causes aggregation problems and mis-interpretaions, commonly known as ecological fallacy. In multi-temporal analysis, when for different time-periods, data is collected for distinct sets of spatial objects – different schema -, modifiable areal unit problems become also an important issue. This article demonstrates how datasets' coherence in terms of spatial boundaries and time periods can be created using auxiliary geographical information.
The motivation for the present article is the attempt to create a small-area database from Portuguese census data, which will for the first time allow the dynamic analysis of demographic patterns for a specific study-area, the Lisbon local council. Census tracts' data are available for three periods: 1991, 2001 and 2011. Historical data analysis is possible between the two first periods but not for the 2011 dataset; this is so because complex geometries did not allow for a coherent coding of observations. Through the use of Geographical Information Systems (GIS), census data was allocated to common areas taking into account a weighting scheme dependent on asymmetric mapping based on auxiliary data (area of building blocks). It is believed that the results obtained minimize the error associated with spatial aggregation. Finally, the article exemplifies the potential benefits of the methodology used for future research on the evolution of demographic patterns in Portugal.

*Keywords*: Modifiable Areal Unit Problem (MAUP), small-area data, historical datasets

## 1 Introduction

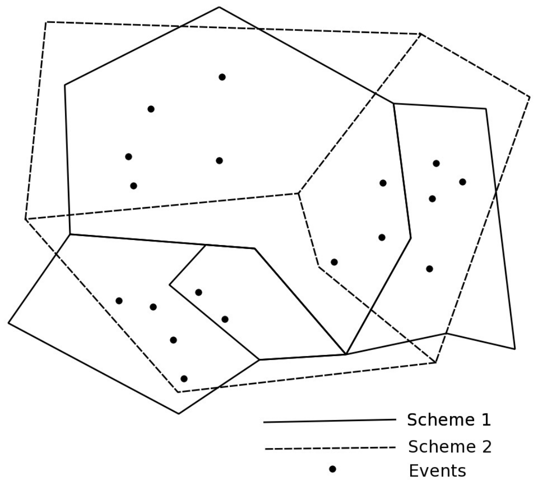
The exercise of compiling socio-economic data into structured datasets implies the division of land into parcels – spatial units or objects, which super-impose artificial boundaries; in fact, with the exception of natural or functional regions, these do not represent barriers to the flow of activity. Particular issues associated with the analysis of these *spatial* datasets are well documented (see for example [1]) and range from the implications of using sums or other measures rather than information for each agent (*ecological fallacy*) to cross-analysis of data from distinct territorial division schema (*Modifiable Areal Unit Problem* - MAUP ).

Census data are one of the most important tools in the analysis of the spatial behaviour/distribution of anthropic activity. Its use is particularly important when the area of interest is sufficiently small as to underline the need for small-area information – data aggregated for census tracts. Moreover, the regularity of national census exercises makes them important sources for spatial-temporal analysis, although technical constraints often serve as barriers to research on dynamic behaviours [2].

When census datasets for different historical moments are aggregated according to different sets of tracts, the exercise of obtaining a coherent multi-temporal database is particular challenging, especially because of MAUP related issues. Also, independent of the administrative level of analysis, the value of observation $i$ from variable $X$ is highly dependent on the size of each spatial unit.

In terms of different spatial aggregation schema, figure 1 exemplifies how count data changes when different territorial borders are used.

Figure 1: Example of count data aggregation



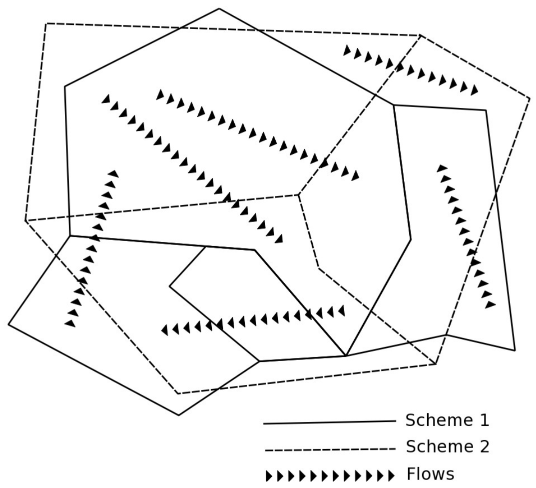

Using two distinct sets of spatial units, the set of counts would vary as follows:

$$Scheme\ 1=\{8,4,2,4\}\ ;Scheme\ 2=\{5,7,6\} \quad (1)$$

For complex territorial structures, with a large number of territorial units, differences arising from distinct schema do not necessarily disappear; on the contrary, different geometries induce systematic variations in terms of observed patterns in the data.

Using the same abstract structures, figure 2 exemplifies the effect of MAUP when inter-regional flow data is collected. If each arrow represents point-origin and destination, it is clear that flows which are captured by scheme 1 are not by scheme 2 and vice-versa.

Figure 2: Example of flow data aggregation




As mentioned above, MAUP related issues may undermine the analysis of dynamic behaviour of human activity along a given time-span. There are different methods which may be used in order to overcome such issues. The most common - the kind dealt within the scope of this article, are those which involve the use of Geographical Information Systems (GIS) in order to obtain a common geometrical scheme and to weight count data according to this new division of a given spatial surface. These *basis change* methods [3] vary in complexity and are determined by fixed assumptions.

One method of overcoming such issues is through the use of asymmetric mapping methodologies using (or not) control zones [3,4]. Asymmetric mapping differs from dasymetric mapping [5] since in the latter there is not necessarily and overlapping conflicts between layers, as in the case of the former.

An asymmetric map results when regions do not coincide in shape and area. If data is to be aggregated according to a common boundary scheme and datasets are assumed to be stationary within each of the original objects, then proportions may be obtained in order to serve as weights in the re-distribution of data. If however non-stationarity is assumed, then auxiliary geographical datasets may be used in order to minimize to error associated with the basis change exercise. Nonetheless, distribution of data within each object belonging to the control zones layer is assumed to be homogeneous.

The need to use basis-change methodology when studying demographic dynamics for Portuguese small-area data - census tracts, arises from the fact that, due to changes in land-use, it was impossible for the National Statistical Institute to construct a code which would allow data aggregation for common areas [6]. This is true for the last census exercise, the results of which were published in June 2011 [7].

The main objective of this work is to create a coherent database which will allow the historical analysis of small-area data corresponding to three census operations: 1991, 2001 and 2011. In order to achieve this goal, auxiliary data (control zones) will be used: buildings' footprints and their respective area. This methodology and resulting datasets open a large number of research possibilities since so far it was not possible to perform this type of analysis due to technical issues described and which this study allowed to overcome.

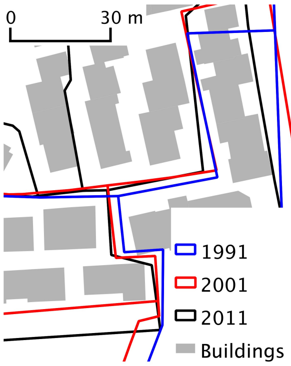
The remaining of the article is organized as follows: the next section presents the datasets used, exemplifying the problems associated with spatial-temporal analysis of small-area Portuguese census data. Next, in section 3, the methodology used will be described in detail. This will be followed by a results section where density of population per built-up area is used to serve as an example of the potential of the final database. Section 5 concludes.

## 2 Data

The geographical data used in this study includes two planimetric sets: the census tracts and the buildings' footprint.

The census tracts in Portugal are supported by cartography and are available in vector format for the years 1991, 2001 and 2011. For 1991, the geographic base is in analogical format and it is called “Geographic Spatial Referencing Base” (BGRE 1991). This map was then digitalized and made available in a Geographic Information System (GIS). For the following years, it is named “Geographic Information Referencing Base” (BGRI 2001, BGRI 2011) and is already based on a GIS.

Figure 3: Example of geometric inconsistency




The buildings' footprint was obtained from the city of Lisbon. This layer is available in vector format, for the year 2006, and has a numeric scale of 1:1000.

Figure 3 shows a snapshot of the city with the four layers represented. There are differences in terms of geometry between all referencing bases; however, for historical comparisons between 1991 and 2001, the National Statistical Institute provided a common area code, allowing for a common geometry to be built. This is a snapshot of the asymmetric paradigm.

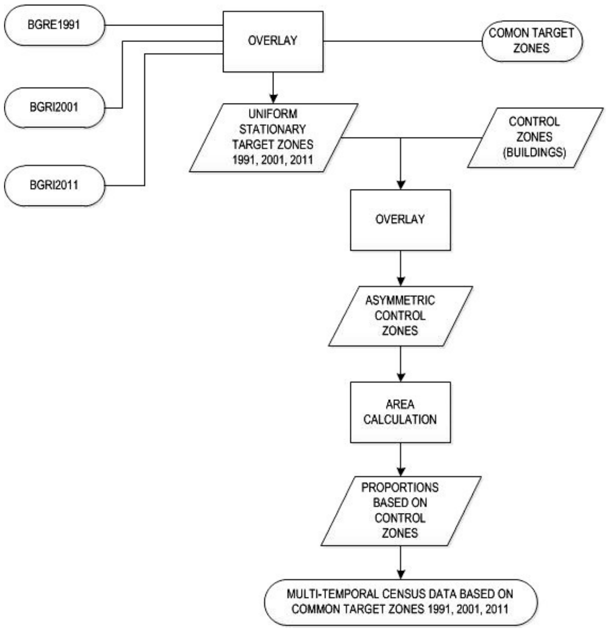
## 3 Methodology

Because of the geographical challenges facing the multi-temporal analysis of census data, ideally, efforts should continue towards the point-base geo-referencing of primary sources of data retrieval [8]. However, when such point-data is not available, asymmetric mapping using control zones is an appropriate solution [3].

Figure 4 describes the different steps which lead to the final datasets. The exercise of building a coherent multi-temporal census database for Portugal involved a basis-change exercise where the target areas - or common target zones - result from the aggregation of the BGRI 2001 according to the coding of minimum common areas between 1991 and 2001 [6]. Overlaying all geometries allowed for a common code to be attributed to each small spatial block. This itself would allow the base-change exercise to be completed if target zones' densities were assumed to be uniform [3].

One potential hazard of assuming constant behaviour over a spatial object, using the objects' total area as the source for basis-change is the fact that non-stationarity is common since within unit variation is generally not constant [9] . If however stationarity is not assumed, control zones should be used. As mentioned above, in this study, densities are assumed to be uniform, but for the buildings' footprint – the control zones. In practice, this means that for each block/polygon resulting from the first overlay exercise, total buildings' area was calculated. Later, these values were summed up for each census unit for all three time periods. This served to calculate weights which were used to assign proportions of census data which were finally aggregated for the common target zones.

Figure 4: Flowchart



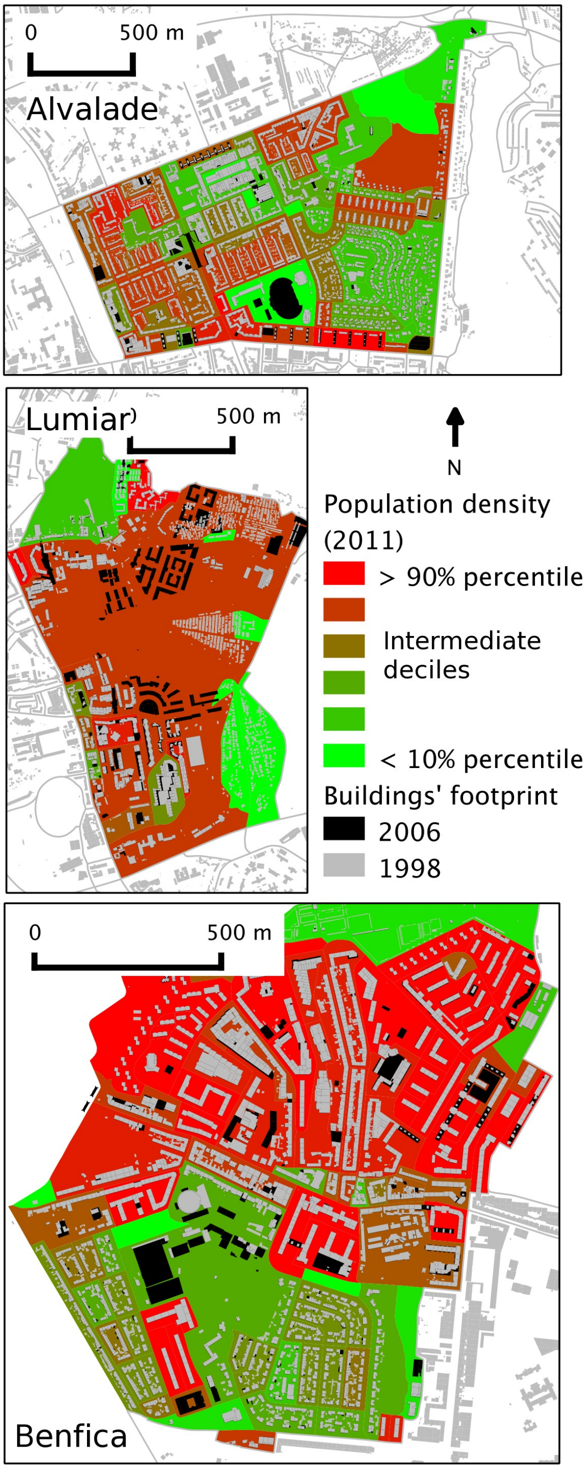

## 4 Results

Based on the methodology described above, several maps of Lisbon were created for indicators such as the population density, growth rate per built-up area between 1990 and 2011 and population density per built-up area in some of the neighbourhoods (see figure 5). These maps reflect the possibilities of a bigger scale analysis, highlighting the diversity of urban dynamics.

Lisbon has lost 115,763 of its inhabitants (17,45% of its population), particularly from the downtown historical neighbourhoods and their extensions to the west and immediate surroundings in the past 20 years. However, the city gained new inhabitants in other areas. This process of increasing population density occurred specifically in the peripheral areas and in the middle section (see appendix 1). This increase was driven by the Lisbon's functional structure evolution and by the urban regeneration of the eastern (Parque das Nações district) and northern areas (Alta de Lisboa district).

The maps shown represent some of the areas with the biggest population density values for 2011 (see figure 5). The first (Alvalade) is located in the middle section and the others (Alta de Lisboa in Lumiar and Benfica) in the periphery.

Figure 5: Analysis of final datasets




These results confirm the attractiveness of the referred areas and seem to include gentrification processes. The changing uses and residents in central areas of the city are favoured by a number of factors, such as public policies, new lifestyles, buildings with architectural quality and housing dimension.

## 5 Concluding remarks

This paper showed how auxiliary geographical data can help to overcome geometrical inconsistencies between different (historical) territorial boundary schemes. Buildings' footprints were used as control zones in order to properly weight demographic data according to common minimum-area schemes.

The weighting methodology used opens new research possibilities in different scientific areas in Portugal, since for the first time, a coherent historical small-area database is available considering three national censuses. For example, in fields such as geo-demographics and urban planning, this data allows a detailed analysis of human dynamics as a response to urban policies. Also, trend-analysis is a key element of sustainability studies, since the distribution of population in closed urban systems should account for physical constraints [10].

Future work is planned which will focus on several themes: first, the use of different control zones in order to test the robustness of the present results; also, future availability of sample datasets consisting of information for each building will allow a precise evaluation of datasets' quality.

## References

[1] Rogerson, P; Statistical Methods for Geography. SAGE Pubications, 2001.

[2] Anselin, L; The Future of Spatial Analysis in the Social Sciences. *Geographic Information Sciences*. 5(2): 67-76, 2009.

[3] Goodchild, M. F, Anselin, L. and Deichmann, U. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25(3): 383-397, 1993.

[4] Freire, S; C. Aubrecht; Towards improved risk assessment - Mapping spatio-temporal distribution of human exposure to earthquake hazard in the Lisbon Metropolitan Area. In: Boccardo, P. et al. (Eds.): *Remote Sensing and Geo-Information for Environmental Emergencies*. International Symposium on Geo-information for Disaster Management (Gi4DM). CD-ROM. Torino, Italy, February 2-4, 2010.

[5] Mennis, J; Generating Surface Models of Population Using Dasymetric Mapping *The Professional Geographer*, 55(1): 31–42, 2003

[6] Santos, A; Development and Use of Geocoding: Portugal approach. *United Nations Expert Group Meeting on Contemporary Practices in Census Mapping and Use of*

*Geographical Information Systems.* United Nations, New York, 2007.

[7] Instituto Nacional de Estatística, I.P; Censos 2011 – Resultados Provisórios. INE, Lisboa, 2011.

[8] Martin, D; Last of the censuses? The future of small area population data. *Transactions of the Institute of British Geographers.* 31(1): 6-18, 2006.

[9] Lloyd, C. R; Local Models for Spatial Analysis (2nd Edition). CRC Press, Boca Raton, 2011.

[10] Santos, T., Freire, S., Tenedório, J. A, Fonseca, A.. Using satellite imagery to develop a detailed and updated map of imperviousness to improve flood risk management in the city of Lisbon. Computational Vision and Medical Image - Processing VipIMAGE 2011, J.M. Tavares, R.M. Natal Jorge (Ed.), CRC Press, 2012.

[11] UNITED NATIONS - Department of Economic and Social Affairs Statistics Division; Handbook on Geographic Databases and Census Mapping (Draft version). *Proceedings of the United Nations Expert Group Meeting on Measuring the Economically Active Population in Censuses,* New York, 2008.

[12] Rodrigues, A; Tenedório, J; The Use of Spatial Weights Matrices and the Effect of Geometry and Geographical Scale. *European Colloquium on quantitative and Theoretical Geography* - *ECTQG 2011*, Athens, 2 - 5 de Setembro 2011.

*[13] Walford, N; Hayles, K; Thirty Years of Geographical (In)consistency in the British Population Census: Steps towards the Harmonisation of Small-Area Census Geography. Population, Space and Place, 10: 2011, 21pp.*

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Appendix 1: Lisbon local council


Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke