Generalized Dasymetric Mapping Algorithm for Accessing Land-Use Change

Antonio Manuel Rodrigues and Jose Antonio Tenedorio
CICS.NOVA Interdisciplinary Centre of Social Sciences Faculdade de Cincias Sociais e Humanas, Universidade Nova de Lisboa, Lisbon, Portugal
amrodrigues@fcsh.unl.pt
http://www.cics.nova.fcsh.unl.pt/

Abstract. The use of multivariate micro-data, data aggregated for small-areas, allows detailed analysis of the physical and social structures of regional landscapes. Such exercises are in many cases univariate and static in nature; this happens when geometries are not coincident between datasets. Common solutions to such inconsistencies involve the use of areal interpolation techniques to build coherent information sets; when ancillary information is available, dasymetric mapping using control units may then be used. Techniques vary on the type and quality of the ancillary (or control) information. The purpose of the present article is to present a generalized tool to tackle common practical analytical problems and which produces geometrically coherent datasets. It is generalised because: (1) it is flexible, allowing distinct parametrization depending on the data; (2) it is based on Open Source tools anchored on robust database management systems (DBMS) technologies. Its aim is to provide the regular GIS user with a tool to tackle a common problem of geometric mismatch.

Keywords: Dasymetric mapping · Computational statistics · FOSS · Database Management Systems

1 Introduction

Multivariate or temporal analysis of geographical data aggregated for specific areas, polygons, is highly dependent on the geometrical coherence between these spatial units. When the shape of these units from distinct datasets does not coincide, areal interpolation and Dasymetric mapping techniques are important [2,3,7]. An areal interpolation exercise may be described as the re-allocation of numeric data according to some geometrical schema; original geometries may be called source spatial data and the end-geometries as target spatial data; the final datasets are spatially coherent because data describing all phenomena studied are aggregated according to the same set of areas (spatial units).

The characteristics of spatial data aggregated for a finite set of spatial units (regions) imply a number of problems which should be tackled, otherwise statistical results may be flawed. When using univariate statistical techniques for
single variable exploratory work, the simple fact that events are aggregated into areas originate lost of information usually related to what is commonly known as ecological fallacy \cite{1}. In multivariate or time-series analysis, when geometrical schema used to collect data for distinct phenomena are not the same, there is the need to use areal interpolation techniques, to which Dasymetric mapping belong to, in order to guarantee geometrical correspondence.

Although dasymetric mapping techniques are not new, there is no tool which is both general in terms of scope, free in terms of access and robust in terms of technology. The present article represents the end point in a series of research efforts where Dasymetric techniques were used with distinct types of control datasets \cite{4,5,7,9}. The aim was to produce a tool, based only on Open Source technologies which can be used in a large variety of settings. The algorithm used is implemented in SQL using functions from PostGIS, a library which introduces spatial and geographic objects for PostgreSQL (http://postgis.net/ and http://www.postgresql.org/).

Three distinct analytical scenarios were considered in the final design, with an increasing degree of complexity: first, when the goal is to re-allocate count data from source to target geometries and no ancillary information is available - plain areal interpolation. Second, when ancillary data is binary in nature, making the distinction between zones where events can and cannot occur (ex. buildings footprint which identify where individuals reside as opposed to all other areas) - areal interpolation with binary ancillary data. Third, when ancillary data is stratified; in these cases, a weighting scheme is designed to account for distinct probabilities that events occur in specific areas (ex. slope may influence the probability of observing certain species of plants) - areal interpolation with stratified ancillary data.

Practical applications are crucial to test implementation. In the final section of the article, three distinct examples are considered for all three scenarios and three distinct study areas. All cases chosen reflect one common problem, which is the non-coincidence of geometries between census micro-datasets from different years. These datasets are rich in detail and provide the best sources of demographic data, with impressive detail (small areas - census tracts) covering large areas (nation-states).

It is important to be aware of the degree of uncertainty of any Dasymetric mapping exercise. This is a function of degree of dis/agreement in shape (conformity) and size (equivalence) between source and target geometrical datasets. In the former case, Modifiable Areal Unit Problems (MAUP) increase the greater the differences in shape between geometrical schemas \cite{7}; in other words, the lesser is the degree of conformity. In cases when target geometries are smaller than source (ie. when data is pulverized), the degree of equivalence is smaller.

2 Methodology

Most phenomena, physical or human, have a spatial attribute (ie. they can be referenced to some theoretical model of, often, the globe). Hence, it is recurrent