Development of a QGIS Plugin to Dasymetric Mapping

Nelson Mileu
Universidade de Lisboa

Margarida Queirós
Universidade de Lisboa

Follow this and additional works at: https://scholarworks.umass.edu/foss4g
Part of the Geographic Information Sciences Commons

Recommended Citation
DOI: https://doi.org/10.7275/3628-0a51
Available at: https://scholarworks.umass.edu/foss4g/vol18/iss1/9

This Paper is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
Development of a QGIS Plugin to Dasymetric Mapping

Nelson Mileu\textsuperscript{a,\ast}, Margarida Queirós\textsuperscript{a}

\textsuperscript{a}Institute of Geography and Spatial Planning, Universidade de Lisboa, 1600-276 Lisbon, Portugal.

\textbf{Abstract:} Demographic data is usually represented by a choropleth map, where the statistical data is aggregated to areal units. This type of representation has several limitations associated with spatial analysis and distribution. A common alternative to display statistical data in meaningful spatial zones is the dasymetric mapping technique. Though dasymetric mapping has existed for many decades, the open source GIS tools to explore dasymetric mapping methods are scarce. In this paper, a Geographical Information System (GIS) open source application was developed in QGIS software that applies the dasymetric mapping method to the Portuguese Guimarães municipality 2011 Census block-group populations and uses Corine Land Cover Data Set to redistribute the block-group populations into a 25-m grid. The application employs a simple centroid sampling approach (supported by the Addresses theme obtained in the INSPIRE ATOM Download Service from the Portuguese National Statistics Institute, INE) to acquire information on the population densities for different land use classes, and it uses the ratio of class densities to redistribute population to sub-source zone areas. Several tools available from QGIS and Geographic Resources Analysis Support System (GRASS) GIS were employed to generate a resident population dasymetric map. The application development was supported in Python 2.7 language and PyQt4 API. The developed QGIS plugin is an innovative tool that allows the population mapping to any case study that has statistical data and the Corine Land Cover layer. To test the results obtained from the tool, census block populations were compared with the dasymetric map. The results indicate that dasymetric mapping produce more accurate population distributions than the choropleth approach.

\textbf{Keywords:} Dasymetric mapping; census data; corine land cover; QGIS; Python; plugin

*Corresponding author

\textit{Email address:} nmileu@campus.ul.pt

\textit{URL:} http://ceg.ulisboa.pt/investigacao/investigadores/

Submitted to FOSS4G-Europe 2018 Conference Proceedings, Portugal. May 10, 2018
1. Introduction

Demographic data is usually represented by a choropleth map, where the statistical data is aggregated to areal units. This type of representation has several limitations associated with spatial analysis and distribution (Mennis, 2003). A common alternative to display statistical data in meaningful spatial zones is the dasymetric mapping technique. Though dasymetric mapping has existed for many decades, the open source GIS tools to explore dasymetric mapping outputs are scarce.

This paper describes Dasymetric QGIS plugin development for generating a raster representation of demographic data using a dasymetric mapping technique that incorporates the Corine Land Cover layer. For testing the QGIS plugin that automates the dasymetric mapping process, the Portuguese Guimarães municipality was used as case study. Based on the 2011 Census subsections and the Corine Land Cover Data Set the population was redistributed into a 25-m grid. The application employs a simple centroid sampling approach (supported by the Addresses theme obtained in the INSPIRE ATOM Download Service from the Portuguese National Statistics Institute (INE)) to acquire information on the population densities for different land use classes, and it uses the ratio of class densities to redistribute population to sub-source zone areas.

The remainder of this paper is organized as follows: the literature review is described in Section 2. The dasymetric mapping method and the plugin development are described in the methodology Section. The case study description and data used are given in Section 4. In section 5 the results and a brief discussion are the focus. Finally, a summary of the findings and an outline of future research directions are provided in Section 6.

2. Literature Review

In GIS literature several methods for estimating population have been described. A review on these methods, made by Wu et al. (2005), groups the methods into two categories: areal interpolation methods and statistical modeling methods (Figure 1). As described by Wu et al. (2005) the areal interpolation method is focused on zone transformation involving the conversion of the data from one dataset of spatial units to another. The statistical modelling methods make use of different socioeconomic variables for population estimation, using census data in the model training process.

![Figure 1: Population estimation methods. Adapted from: Wu et al. (2005)](image)

In this paper only the areal interpolation approach with ancillary information is analyzed, since it is the theoretical framework for the QGIS plugin development. The areal interpolation with ancillary information supports the population estimation on existing relations to other information like land use or transportation network. In this category the
dasymetric method is one of the most used by researchers (Eicher and Brewer, 2001; Mennis and Hultgren, 2006; França et al., 2014) using several approaches and different sources of ancillary information.

The first references about the use of this method are attributed to the Russian cartographer Tian-Shansky who developed the multi-sheet population density map of European Russia (Mennis, 2003; Bielecka, 2007). However, it was Wright (1936) that popularized dasymetric mapping presenting population density based upon the division of a given administrative unit into smaller areas complying with different types of geographical environments (Mennis, 2003).

More recently, and accompanying the development of geographic information systems several works have been developed using different types of dasymetric methods and tools. For Eicher and Brewer (2001), the most common dasymetric mapping method is the binary method, in which ancillary data classes are regarded as either populated or unpopulated. But other more sophisticated methods were developed. For example, Holloway et al. (1997) used multiple datasets to detect and remove uninhabited lands from the area of analysis. Four types of area were ruled out and to redistribute the census population to the ancillary feature classes, a predetermined percentage was assigned to each class.

Eicher and Brewer (2001) offer a review and evaluation of a number of dasymetric mapping techniques, including the use of raster-based approaches to areal interpolation. The authors describe the “grid three-class” method, which they demonstrate by using raster land-use data (with three classifications) to redistribute county-level population data to subcounty units. They assigned a predetermined percentage of a county population to a given land-use area (70% of the population of a county to urban, 20% to agricultural/woodland, and 10% to forested land uses).

Mennis (2003) used a three-tier raster classification of urban land cover derived from the Landsat Thematic Mapper as ancillary data. Within the remotely sensed land-cover data, urban features were put into three classes of high density, low density, and nonurban, with no distinction of wooded areas, agriculture, or slope.

Mennis and Hultgren (2006) describe the intelligent dasymetric mapping approach that takes as input count data mapped to a set of source zones and a categorical ancillary data set, and redistributes the data to a set of target zones formed from the intersection of the source and ancillary zones. The IDM method was programmed as a Visual Basic for Applications (VBA) script within the ArcGIS software.

In recent years, several articles have been published with reference to the method proposed by Mennis (2003), applied in several countries and using different ancillary data (Sleeter, 2007, Freire, 2007, França et al., 2014). Regarding the development of software tools for the calculation of dasymetric maps, also several scripts in different programming languages have been developed (Sleeter and Gould, 2007). Following the original Mennis (2003) method of dasymetric mapping, the United States Geological Survey (USGS) implemented a tool in ArcGIS platform. This tool automate the dasymetric mapping process allowing to map and analyze any area of interest provided the appropriate data sources.

3. Methodology

To implement the dasymetric method in QGIS a modified equation based on Holloway et al. (1997) was used to calculate the population for each land-cover cell (pixel).

\[ P = \frac{(R_n A_n) \times N}{E} \]  

(1)
Where, \( P \) is the population of a cell,
\( R_n \) is the relative density of the mapping unit population with land-cover type A
\( A_n \) is the area of mapping unit
\( E \) is the expected population of enumeration unit calculated using relative densities
\( N \) is the actual population of enumeration unit

In order to automate the dasymetric mapping process a QGIS plugin was developed\(^1\). To develop the plugin, the QGIS and Geographic Resources Analysis Support System (GRASS) GIS were used. The application development was supported by Python 2.7 language and PyQt4 API.

All the steps used to automate the dasymetric mapping process are described in Figure 2. Since the statistical enumeration areas layers are in vector format, the first step is the conversion to raster for ID and population fields, using the GRASS GIS v.to.rast.attribute tool.

To quantify the population densities for each land-cover class within Guimarães municipality (Portugal), the addresses layer was used. The land use classes were merged and reconfigured into classes of high density residential, low density residential, industrial/commercial, road and rail networks, other urban areas, agriculture area, forest

---

\(^1\) https://github.com/nmileu/QGIS_Dasymetric_plugin
areas; and water bodies/wetlands, using the GRASS GIS r.recode tool. Assuming a linear relation between the number of dwellings and population, the densities were obtained intersecting the addresses layer and corine land cover layer, and calculating the percentages per land use class (Figure 3). The main difference between these and other approaches lies in the use of all addresses to obtain the densities, instead of using a sampling process.

![Diagram](Figure 3: Densities calculation.)

The relative density raster layer was obtained using the GRASS GIS r.reclass tool and the previously calculated density values (Table 1).

<table>
<thead>
<tr>
<th>Land use code</th>
<th>Land use class</th>
<th>Relative density (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>high density residential</td>
<td>22, 2</td>
</tr>
<tr>
<td>112</td>
<td>low density residential</td>
<td>40,2</td>
</tr>
<tr>
<td>121</td>
<td>industrial/commercial</td>
<td>6,1</td>
</tr>
<tr>
<td>122</td>
<td>road and rail networks</td>
<td>0,5</td>
</tr>
<tr>
<td>199</td>
<td>other urban areas</td>
<td>0,1</td>
</tr>
<tr>
<td>299</td>
<td>agriculture area</td>
<td>27,5</td>
</tr>
<tr>
<td>399</td>
<td>forest areas</td>
<td>3,4</td>
</tr>
<tr>
<td>499&amp;599</td>
<td>water bodies/wetlands</td>
<td>0,00</td>
</tr>
</tbody>
</table>

The value of E (the expected population of enumeration unit) is obtained by the proportions of land-cover types in each statistical enumeration area. To calculate the proportions a Python function was developed. This function tabulates the ID raster layer and the CLC recoded raster layer, resulting in the weights per land use class for all statistical enumeration areas. After calculating the value of E, the tabulated table in joined with vector statistical enumeration areas and then, converted to raster, using the GRASS GIS v.to.rast.attribute tool. After obtaining all input layers, the population per cell is calculated using the initial equation (1) in QgsRasterCalculator.

The QGIS plugin was developed using PyQt4. The QGIS plugin interface gets all the inputs, including the vector layers, the relative densities, the output raster resolution and path (Figure 4).
4. Study Area and Dataset Used

The present research was developed in the municipality of Guimarães located in the north of Portugal, near 41°26′42″N and 8°17′27″W coordinates (Figure 5). Guimarães, is a middle size town with 158,124 inhabitants, distributed by 240.955 Km² and 48 administrative parishes. The choice of Guimarães for the study area was motivated by the fact that it is a municipality with diverse patterns of human occupation, including urban areas with high population densities and isolated rural occupations with low population densities, as well as the presence of demographic voids.
The dasymetric mapping process used the following layers in: 1) Land use; 2) Statistical subsections and; 3) Addresses.

Regarding land use characterization there are various available resources for information. One of these resources is the Corine Land Cover (Coordination of Information on the Environment Land Cover, CLC) database. It is a standardized data set produced by the European Environment Agency (EEA) and its member countries, in the form of detailed descriptions of land occupation and features, at an original scale of 1:100,000, using 44 Classes of the 3-level Corine nomenclature, with a definition of 25 ha, and for several years. The choice of this dataset to test the dasymetric mapping plugin is due to the fact that they are current data, standardized and available to any European country in shapefile format.

The population data is based on the 2011 Portuguese population and housing census geographic reference (BGRI) layer, produced by the Portuguese Statistics Institute (INE). This layer contains the administrative delimitation, the statistical sections and subsections in shapefile format. To test the dasymetric mapping the 2671 statistical subsections of Guimarães municipality were used. The statistical subsection is the territorial unit which identifies the smallest homogenous area, representing a block in urban areas, a locality or part of a locality in rural areas or to residual areas which may or may not contain dwellings.

The addresses layer (67182 points) was used to calculate the relative weights of land use, and was obtained from the INSPIRE ATOM Download Service (provided by the INE). The National Base of Addresses is a 2D vectorial database that contains addresses of Continental Portugal by municipality. This data set is structured according to the INSPIRE Directive, Annex I Addresses theme, and was obtained from the National Dwellings File (FNA) from INE.

5. Results and Discussion
The population dasymetric map of Guimarães was obtained from the QGIS plugin process with a 25m resolution (Figure 6). The map is based on the population by subsection and was disaggregated with the recoded Corine Land Cover classes. The population dasymetric map results in a superior visual enhancement of the data, since inhabited cells, like water courses, present zero population.

![Figure 6: Dasymetric map of 2011 population in Guimarães municipality.](image)

To evaluate the results obtained from the QGIS Dasymetric plugin, a value comparable to the statistical subsection was obtained. This value results from the sum of the dasymetric cells for each statistical subsection. Figure 7 represents an histogram of the absolute differences between the two variables, calculated by subtracting the population values for each statistical subsection and the sum of the dasymetric population cell values. The histogram shows an approximately normal distribution, where the majority of the differences are near zero. This distribution reveals that dasymetric map preserves an accurate statistical subsection level summation of population data.
Figure 7: Histogram showing difference between subsections populations and dasymetric map values.

Figure 8, shows the spatial error distribution. The urban statistical subsections reveal positive errors, indicating a population overestimation. On the other hand, the majority of the rural statistical subsections reveal negative errors, indicating a population underestimation.

6. Conclusion

This research shows that the developed QGIS plugin, implements the dasymetric mapping process with success. The plugin is an innovative tool, since it automates the
The dasymetric mapping process in QGIS and establishes a new approach to obtain the relative weights for land use classes based on the addresses open data layer. The tool allows the dasymetric mapping for any case study in Europe that has a statistical data layer and the Corine Land Cover layer. The obtained results are in accordance to the cartographic scale and selected resolution. However the Corine Land Cover cartographic detail is a limitation in the cartographic results, since the generalization is a characteristic of this data set. It is possible to use any other land use layer, but the GRASS GIS recode files need to be adapted.

Although the automated process does not involve any technical involvement or GIS expertise outside the plugin interface the overall task of dasymetric mapping is time consuming. This issue will be tested in new versions passing the algorithm to another thread. Other improvements to the plugin will stand on the integration of a validation tab where different statistical tests (e.g. ANOVA) will be available for the user.

In this test phase the plugin will be hosted on GitHub platform being available to the community. After the tests, it is expected to add the plugin to the QGIS plugins repository.

Acknowledgement This work was supported by national funds from FCT - Portuguese Foundation for Science and Technology (UID/GEO/00295/2013).

References


