Redistributing Agricultural Data by a Dasymetric Mapping Methodology

Maria de Belém Costa Freitas Martins, António Manuel de Sousa Xavier, and Rui Manuel de Sousa Fragoso

This paper examines the adaptation of dasymetric mapping methodologies to agricultural data, including their testing and transposition, in order to recover the underlying statistical surface (i.e., an approximation of the real distribution of data). A methodology based on the ideas of Gallego and Peedell (2001) and on the binary method is proposed. It has several steps: (i) the exclusion of target zones for which no observations exist (binary method), (ii) the application of an iterative process to define the most precise densities for data distribution, and (iii) the stratification/definition of sub-units with homogenous characteristics if the results of the previous step are not satisfactory, and the subsequent application of step two.

The methodology was applied in the Alentejo region of Portugal, using data from the 1999 Agricultural Census. Several counties are used as source zones. The aim was to generate a distribution of agro-forestry occupations as close as possible to reality. Two lines of analysis were followed: (i) application of the methodology simultaneously to all counties (definition of regional densities), and (ii) application of the methodology separately to the different sub-areas with similar characteristics (definition of sub-regional densities). For an easy application of the methodology, a computer tool was created, which allowed the easy optimization, validation, and exportation of the data into a Geographic Information System (GIS).

The results were validated using several error indicators at the county level, as well as in a sample of parishes. We show that the second variant of the methodology yielded more precise results, and is superior for the types of data available. This method yielded maps in which the distribution of the most relevant agro-forestry occupations is closest to reality.

Key Words: dasymetric mapping, agricultural data, spatial disaggregation, iterative process, Alentejo

Population density data are available to the European Commission at the level of the county or parish. However, this level of spatial resolution may be insufficient, in many cases, for planning or modeling purposes or to assess the impact of European Union policies. In some countries, as in France, where most counties have a rather small area (approximately 15 km^2 on average), the resolution may be sufficient, but it is clearly insufficient in other countries where the counties tend to be larger (Gallego and Peedell 2001).

Maria de Belém Costa Freitas Martins is Assistant Professor with Aggregate and António Manuel de Sousa Xavier is a Ph.D. student at the Universidade do Algarve in Faro, Portugal. Rui Manuel de Sousa Fragoso is Assistant Professor with Aggregate at the Universidade de Évora in Évora, Portugal. The three authors are also affiliated with the Center for Advanced Studies in Management and Economics at Universidade de Évora.

The authors gratefully acknowledge partial financial support from the Science and Technology Foundation (FCT) and the Operational Program for Science, Technology and Innovation (POCTI).

Therefore, maps produced have limited utility in detailed spatial analysis, especially where human populations are concentrated in a relatively small number of villages and cities (Bielecka 2005). Choropleth maps by administrative units give the impression that population is distributed homogeneously throughout each areal unit, even when portions of the region are, in actuality, uninhabited. Openshaw (1984) described these limitations as the modifiable areal unit problem (MAPU)—defined as a situation in which modifying the boundaries and scale of data aggregation significantly affects the result of spatial data analysis—and stated that it is often unclear whether the results of statistical data analysis indicate some reality about the individuals living in that region or are rather strictly a function of the particular areal unit used in the analysis.

As regards agricultural data, the situation is the same. Agricultural data is available at the NUTS

Agricultural and Resource Economics Review 41/3 (December 2012) 351–366

II, NUTS III, county, and parish levels, but no information is available regarding the true spatial distribution of the available agricultural variables. In Portugal, for instance, data for these levels of analysis are available only from official statistics and from studies on agricultural data disaggregation (Martins, Fragoso, and Xavier 2010, Fragoso, Martins, and Lucas 2008). However, agricultural policies, in many cases, correspond to territorial areas that do not conform to county or parish boundaries. Rather, agroecological conditions often dictate very different land uses within political boundaries.

A prominent method in areal interpolation is dasymetric mapping, defined generally as the use of an ancillary dataset to disaggregate coarse-resolution population data to a finer resolution (Eicher and Brewer 2001). Recent studies show that dasymetric mapping can yield estimates that are more precise for small area populations than other areal interpolation techniques as it uses not only statistical data but also ancillary data (Mrozinski and Cromley 1999, Mennis and Hultgren 2006). Dasymetric mapping as a procedure is applied to datasets for which the underlying statistical surface is unknown, but for which aggregated data already exists, though the zones of aggregation are not derived from the variation in the underlying statistical surface but are rather the result of some convenience of enumeration (Mennis and Hultgren 2006, p. 180). The process of dasymetric mapping is thus the transformation of data from the arbitrary zones of data aggregation to a dasymetric map in order to recover and depict the underlying statistical surface (Mennis and Hultgren 2006).

Although applied to human population in the beginning, these methodologies can be applied to different kinds of data (Mennis and Hultgren 2006), such as agricultural data. In what concerns planning and policy evaluation, this can be very useful for agriculture, since the statistics have a maximum resolution only for county and parish and not for homogeneous agricultural zones.

The use of geographic information systems (GIS) for modern mapping has renewed interest in dasymetric mapping (e.g., Eicher and Brewer 2001) and enhanced its possibilities. The lack of standardization in production methods is, however, an obstacle that prevents widespread use of dasymetric maps in GIS. Though areal interpolation research has not focused on map production, map production can advance the study of dasymetric mapping by introducing structured evaluation of methods that are related in their approaches (Eicher and Brewer 2001).

Therefore, this paper analyzes the transposition and adaptation of dasymetric mapping and tries to propose a combined methodology, using GIS, in order to solve the investigation problem for the Alentejo region of Portugal, a region for which previous studies on agricultural data disaggregation obtained results by administrative units, but not by land cover class (Martins, Fragoso, and Xavier 2010, Fragoso, Martins, and Lucas 2008). The aim is also to assess how far this methodology may serve as a complement to a dynamic model based on maximum entropy (Martins, Fragoso, and Xavier 2010, Fragoso, Martins, and Lucas 2008) that produces agro-forestry and livestock distribution results for county and parishes in NUTS II, Alentejo, Portugal.

The remainder of this paper is organized as follows. In the next section the mathematical formulation of the disaggregation problem is made. Then, in the following sections, several issues are addressed: the methodological framework, the empirical implementation, the results, and the model's validation. Finally, the last section stresses the main conclusions of this work.

The Problem's Formulation

In order to correctly formulate the dasymetric mapping problem, we consider that there is a source zone *i* and an ancillary zone *C* where *C* is associated with an ancillary class *k*. The purpose is to distribute the values of the target variable in the source zones *i* by the several classes *k*, assuming that in the different units *i* a class must have always the same density *D*. Thus we have

$$(1) \qquad S^r = \sum_{B_i \subset A_r} S^i \quad \text{and} \quad S^i = \sum_k C_k^i \times D_k^i \,.$$

The target zone *t* is defined as an area of overlap of *S* and *C*. The known aggregated variable is S^r and the disaggregated one in an administrative unit *i* is S^i, and C_k^i is the area of the ancillary class *k* in unit *i*. The A_r matrix represents the aggregated unit *r*, in which A_r, $r = 1...R$. So, one has to determine $D_k^i \ \forall\, k,i,$ *D* being the density of the variable *S* for the classes *k*, of which there is no available information at the disaggregated level in the sub-units B_i. The various procedures for its measurement will be fully developed in later sections of this manuscript.

Figure 1 depicts a graphical representation of the problem.

Methodological Framework

Previous Studies

The earliest reference to dasymetric mapping is the 1922 population map of European Russia by Russian cartographer Semenov Tian-Shansky. However, John Kirtland Wright (Wright 1936) popularized dasymetric mapping in the United States and is often incorrectly cited as its inventor, even though he noted the Russian origin of the term “dasymetric” (Mennis and Hultgren 2006).

Wright (1936) produced a population distribution map of Cape Cod using census data, where population counts were assigned to towns in a standard choropleth map, and saw that the choropleth map masked important details, not showing the true distribution of the data. To solve this problem he eliminated the “uninhabited” areas and recalculated the density for the town based on the reduced area. Then he divided each town into regions of land use and settlement and assigned densities to all but one land use class. The density of the last class could be calculated from the residual population of that town.

Langford, Maguire, and Unwin (1991) and Langford and Unwin (1994) described a dasymetric mapping procedure using land cover data derived from Landsat Thematic Mapper (TM) multispectral imagery to build a series of predictive models regressing population density on land use. These models were then used to redistribute ward population census data for Leicestershire in the United Kingdom to a 1 km resolution raster surface. However, they did not preserve the *pycnophylactic* property, defined by Tobler (1979) as when the summation of population data to the original set of areal units is preserved in the transformation to a new set of areal units.

Mennis (2003) proposed computing density values for each land cover class based on the density of all census block groups that lie entirely within that class, which made it difficult to find a reasonable number of source zones. While spatial resolution of the remotely sensed land cover data has improved, it is difficult to find even one block group that lies entirely within one land cover type.

Gallego and Peedell (2001) proposed using disaggregated population data with the help of the CORINE Land Cover (CLC) database, assuming that the ratio between the population densities of two land cover classes was the same for any commune in the European Community in the 1990s (15 countries). They obtained very good results and concluded that results can be improved if Europe is divided into regional groupings of communes with similar characteristics.

Eicher and Brewer (2001) presented three methods (binary, three-class, and limiting variable) for dasymetric mapping of population density. They also evaluated map accuracy using both statistical analyses and visual presentation of errors. A repeated-measure analysis of variance showed that the traditional limiting variable method had significantly reduced error relative to the four other methods.

Bielecka (2005) presented research results on the dasymetric mapping method, producing a dasymetric population density map of the northeastern part of Poland. She started from the binary dasymetric procedure as described in previous studies to eliminate unpopulated regions. She then used the methodology proposed by Gallego and Peedell (2001). The dasymetric mapping method is based on the assumption that the ratio between the population densities of two land cover categories is the same for any given commune. The CLC database was utilized as ancillary data because the land cover pattern best reflects population density and provides useful geo-referenced information for disaggregating census data. The results were very satisfactory, and it was shown that this methodology may present better results when examining highly stratified areas. The two dasymetric methods presented revealed the interregional variation in population density more realistically, in particular among urban and rural areas.

Mennis and Hultgren (2006) developed the method of Intelligent Dasymetric Mapping (IDM), which through sampling of source zones allows the redistribution to target density classes formed from the intersection between the source and ancillary zones. This technique is referred to as “intelligent” because each analyst can subjectively use its knowledge to define the data relationship, extract this relationship from the data using a

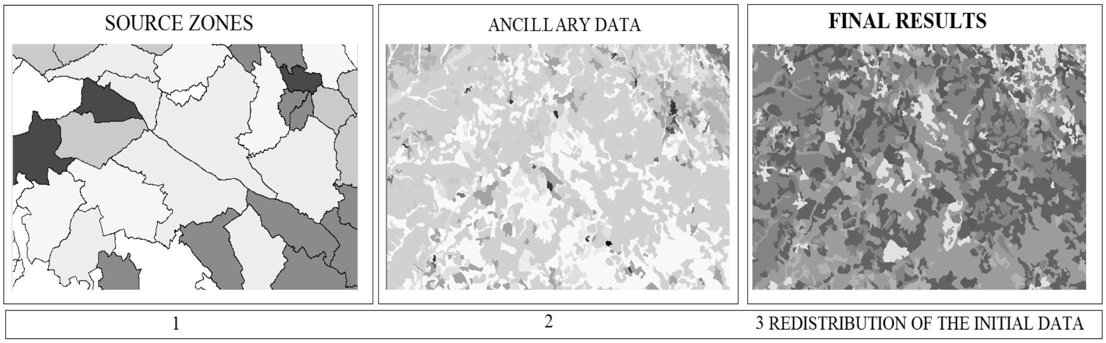

Figure 1. Graphical Representation of the Problem

novel empirical sampling technique, or combine the subjective and empirically based methods (Mennis and Hultgren 2006).

The analyst has three options for the sampling method employed. The "containment" method selects those source zones that are wholly contained within an individual ancillary class, leading to a limited number of valid source zones for calculating the densities of the target variable. The "centroid" method selects those source zones that have their centroids contained within an individual ancillary class, which may lead to a considerable number of source zones selected that may be not representative of that class in particular, leading to situations far from reality. The "percent cover" method allows the user to set a threshold percentage value and then selects those source zones whose area of occupation by a single ancillary class is equal to or exceeds that threshold. This method, however, may lead to erroneous results if the threshold percentages are set at low values.

Proposed Methodology

The methodology used is based mostly on the one proposed by Gallego and Peedell (2001), who used a modified version of areal weighting interpolation to map population density in Europe. They disaggregated population data with the help of the CLC database, assuming that the ratio between the population densities of two land cover classes is the same for any commune in Europe, and admitting that results can be improved if this area is divided into regional grouping communes with similar characteristics (Bielecka 2005).

Therefore, we assume the same premises assumed by the authors: that the density of a class is the same in all the source zones (administrative units), and that the data will be stratified if the initial results are not satisfactory. Also, population data—as well as other kind of data—can be disaggregated with the help of the CLC database, assuming that the ratio between the population density of two land cover classes is the same for any administrative unit. In order to use the methodology proposed by Gallego and Peedell (2001), we can initially assume that the coefficients are known. Therefore, we assume that inside each administrative unit i the density of a variable in a class k D_k^i is assumed to be proportional to given coefficients U_k for each land cover type k:

(2) $$D_k^i = U_k \times W_i,$$

where W_i is an adjustment factor to ensure that the total population in each administrative unit matches the total administrative data. It is also assumed that the total value of variable S for each administrative unit i is given by

(3) $$S^i = \sum_k C_k^i \times D_k^i$$

or

(4) $$S^i = \sum_k C_k^i \times U_k \times W_i,$$

where C_k^i is area of land cover type k in administrative unit i, D_k^i is the density of the studied variable for each land cover k in each administrative unit i, U_k are the coefficients for each land cover type, and W_i is the adjustment factor for each unit i.

D_k^i and W_i may also be calculated as follows:

$$W_i = \frac{S^i}{\sum_k C_k^i U_k}\ ,\ D_k^i = U_k \frac{S^i}{\sum_k C_k^i U_k}\ . \tag{5}$$

Disaggregation of the Agricultural Data

The proposed approach requires the use of the ideas previously discussed, as well as a different kind of method, as presented in Figure 2.

The first step in disaggregating the regional data is to use a binary dasymetric procedure, by excluding classes with zero densities from the project and defining equal densities to the others. The binary procedure is described by Langford and Unwin (1994) as well as by Eicher and Brewer (2001).

The second step is to apply the iterative process proposed by Gallego and Peedell (2001). To solve the investigation problem and because of some limitations of most of the existing land use cartography, the best way to assess the disaggregation of the counties' target variables is to compare the results with reality. And so we (i) disaggregate regional data with CLC using a set of coefficients, (ii) reaggregate the attributed population on a county basis, (iii) compare value per county with the known variable and compute a disagreement indicator, and (iv) modify the coefficients to reduce the disagreement (Gallego and Peedell 2001).

The procedure is described mathematically in detail, as follows. The administrative value of a variable S for a region r is given by

$$S^r = \sum_k C_k^r \times U_k \times W_r\ . \tag{6}$$

The density for each land cover k in region r is

$$D_k^r = U_k \frac{S^r}{\sum_k C_k^r U_k}\ , \tag{7}$$

where C_k^r is the area of land cover type c in region r, D_k^r is the density of population we attribute to land cover type k in region r, and W_r is an adjustment factor to ensure that the total population in each region coincides with the known total.

The population estimated for each county i in region r is

$$S_e^i = \sum_k C_k^i \times D_k^r\ . \tag{8}$$

For each region a set of coefficients weighting population to the land cover category has to be computed in an iterative way until the difference indicator becomes stable. The difference indicator δ^r for the region r was computed as the sum of the absolute values of the differences between the population attributed to each county and the known value (of the statistical data):

$$\delta^r = \sum_{i \in r} |\ S_e^i - S^i\ |\ . \tag{9}$$

The ratio between the attributed population and the known population is computed as follows:

$$\gamma^i = \frac{S_e^i}{S^i}\ . \tag{10}$$

In order to adjust the correlation weighting coefficients within a region, a correlation P_k^r between the ratio of the population attributed to each administrative unit and the known population in that administrative unit and the ratio between the area of land cover and the total area of the region has to be calculated. So we have

$$P_k^r = \frac{\sum_{i=1}^{I} (\gamma^i - \overline{\gamma}^i)(yc_k^i - \overline{y}c_k^i)}{\sqrt{\sum_{i=1}^{I} (\gamma^i - \overline{\gamma}^i)^2} \times \sqrt{\sum_{i=1}^{I} (yc_k^i - \overline{y}c_k^i)^2}}\ , \tag{11}$$

where γ^i is the ratio between the attributed population and the known population, and

$$yc_k^i = \frac{C_{ki}}{C^i}$$

is the weighted value of the land cover class k regarding the total value in administrative unit i.

This correlation was thereafter used to compute the new value of the U_k according to equation (12):

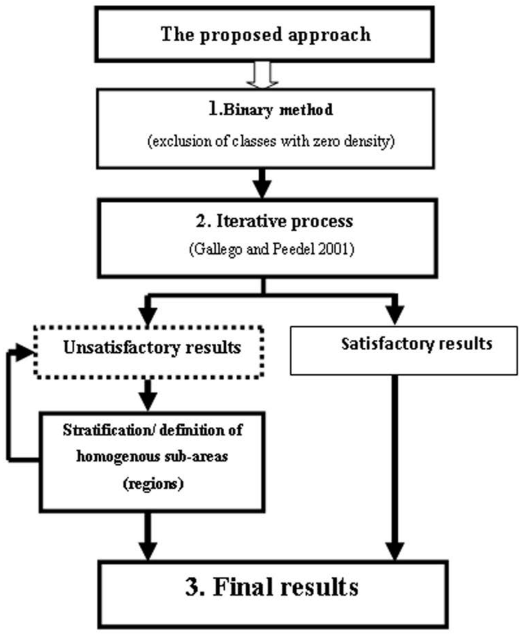

Figure 2. Systemic Representation of the Proposed Methodology

$$U'^{r}_{k} = U_k \left(1 - \frac{p^r_k \times \delta^r}{2 \times S^r} \right). \quad (12)$$

These coefficients are altered in an iterative way, allowing improvement of the methodology's results.

Empirical Implementation

Application Area and the Ancillary Data Used

The methodology proposed was applied to the counties of the Alentejo agrarian region in the south of Portugal, since they constitute an area in which agricultural activity is important (DRAPAl 2007) and where there are several rural areas with problems (Carvalho and Godinho 2004).

The region's total number of administrative units is 299 parishes and 47 counties (the latter will be used for validation). The data were collected from the 1999 Agricultural Census for the Alentejo region at county, parish, and regional levels. The following land uses were considered: cereal (CC), fallow (FF), temporary forages and pastures (PF), permanent crops (PC), permanent pastures (PP), shrubs and forest without pastures (SF), and other occupations (OU).

The ancillary data were from the CLC 2000 database. This is a geographic land cover/land use database encompassing most of the countries of Europe, and it has been used—as, for instance, in Bielecka (2005)—as ancillary data, since the pattern of soil distribution reflects the density of the population variables.

Aiming to maximize the information available and to allow the application of the methodology, it was necessary to simplify the number of categories in the CLC 2000 database. Therefore, the following categories were considered (Figure 3): annual crops, forest, agro-forest systems, heterogeneous agricultural areas, permanent crops, pastures and shrubs, other spaces, and water courses.

For a correct application of the procedures we discussed earlier, we defined zero densities for some areas that we knew had no determined land use. However, we avoided assigning zero densities if possible in order to assess the precision of the iterative process, and to assess whether it can, with a minimum of information available to the user, obtain good results.

In applying the proposed model, we followed two empirical lines (Figure 4). In the first one, we derived regional densities for the Alentejo region. Therefore only one set of coefficients was derived. In the second one, a stratification/ subdivision of the administrative regional unit to the different sub-areas with similar characteristics was implemented. These areas correspond to the four sub-regions of Alentejo: Alto Alentejo, Alentejo Central, Baixo Alentejo, and Alentejo Litoral.

These are not "real" homogenous areas, but administrative sub-regions that have different characteristics; in order to more easily apply the methodology, we did not follow other, more rigorous divisions. Therefore, we envisage that a future study must use some of the divisions previously created by the project MONTADOS or the Gabinete de Planeamento de Políticas Agro-alimentares (GPAA).

The Computer Tool

A recursive use of the methodology implied a continuous treatment of the data using programs such as Microsoft Excel. So, in order to allow the

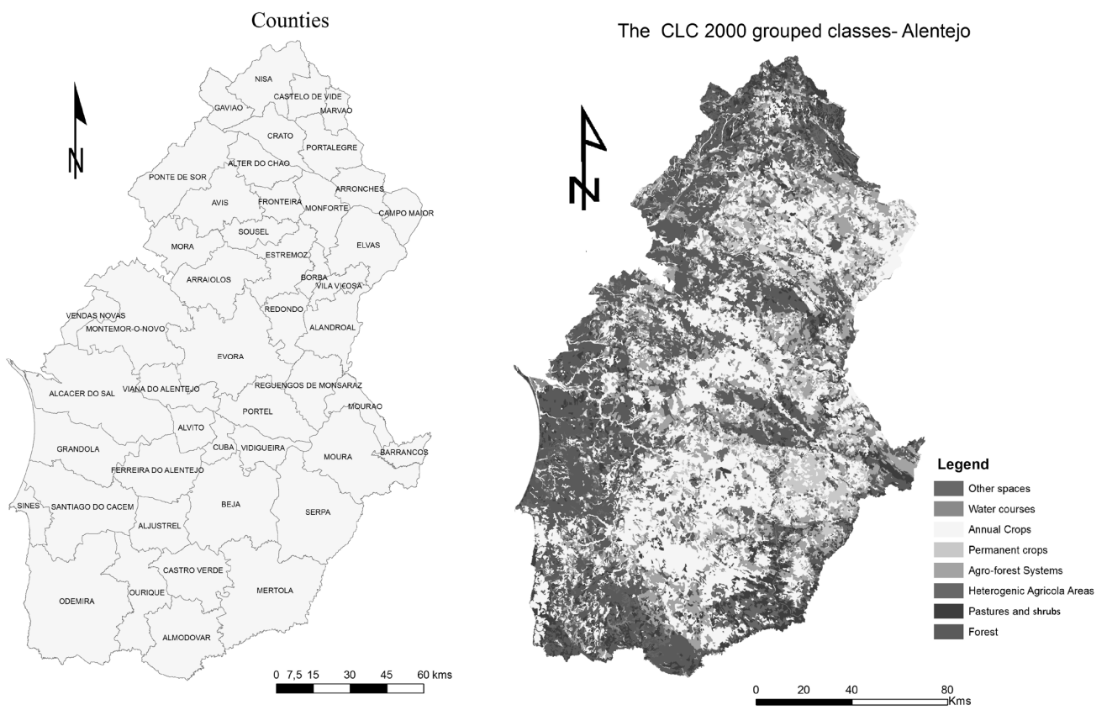

Figure 3. Source Zones and Ancillary Data

Note: The ancillary data are from the CORINE Land Cover 2000 database.

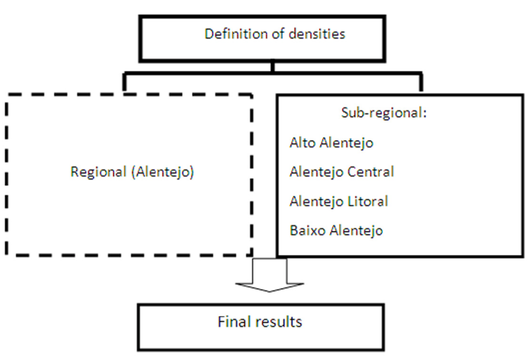

Figure 4. Framework for Applying the Model

easy application of the proposed methodology, we created a computer tool that allows the importation, treatment, optimization, and exportation of data into a GIS system. This simple tool uses Microsoft Excel's data calculation capabilities and provides a set of different pre-programmed actions, in a simplified environment. Note that it works in several Microsoft Excel spreadsheets and that it is only a combination of macro commands (the pre-programmed instructions discussed previously).

The overall structure of the tool is illustrated in Figure 5, which shows the interface and major functions of the tool created.

The use of the tool is quite easy. Figure 6 shows the start page and the optimization page aspect of the application in Microsoft Excel (note that the program's environment is omitted). There, we have several options, including the importation of data. After inserting the data, we just have to select the number of iterations intended to obtain satisfactory results. Excel then calculates and reports several measures of the validity of the results. If the process doesn't produce satisfactory results, a stratification of the initial administrative areas, which is not made inside the application, will be needed.

The algorithm presented here allows in its application the ability to export the geospatially manipulated data into a format which can be read by most Geographical Information Systems software

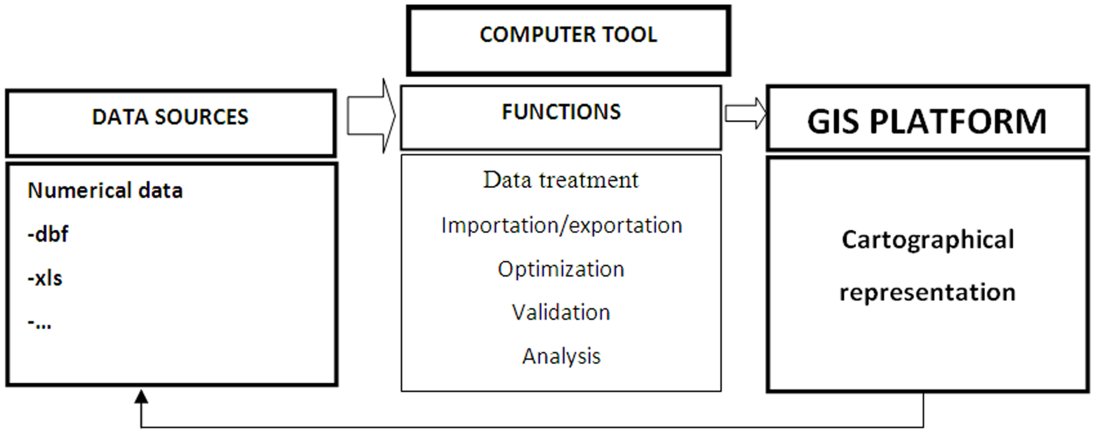

Figure 5. Overall Structure of the Computer Tool Created

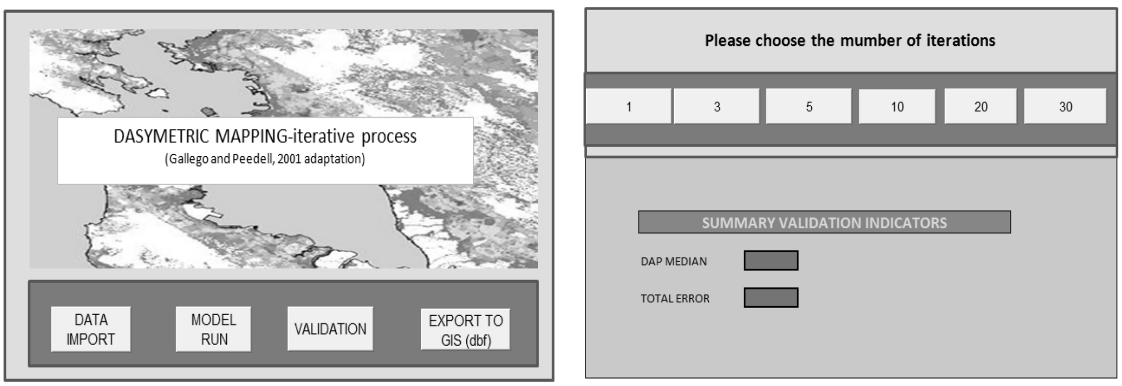

Figure 6. The Start Page and the Optimization Window

programs, including ArcMap,[1] since there are several pre-defined data treatment operations for that purpose. If we choose that option, a macro command will create another Microsoft Excel file with that data, ready to be imported to GIS after saving it. Then we may construct a map in ArcGIS 9.3, as shown in Figure 7.

All algorithms that are intended to generate approximations of reality allow for unanticipated phenomena to systematically enter and bias the results in an unknown direction. Algorithms designed to approximate spatial phenomena are no exception. This cautionary statement does not diminish the utility of the algorithm in the context of available and competitive technology. Rather, caution must always be exercised in the interpretation of the geospatially manipulated data. We recommend that experts in geospatial analysis, in geospatial technology, and knowledgeable of the specific region evaluate the results with the aim toward adding assurance that the whole spatial pattern makes sense, and that sub-areas of the spatial patterns do not have glaring errors.

Results

This model was applied in the Alentejo region, allowing the establishment of regional densities for each land cover, and so an approximation to a real density of the studied variable was created. Sub-regional densities for each of the sub-regions selected were also obtained. Table 1 represents a sample of those sub-regional densities calculated.

[1] See http://www.esri.com/.

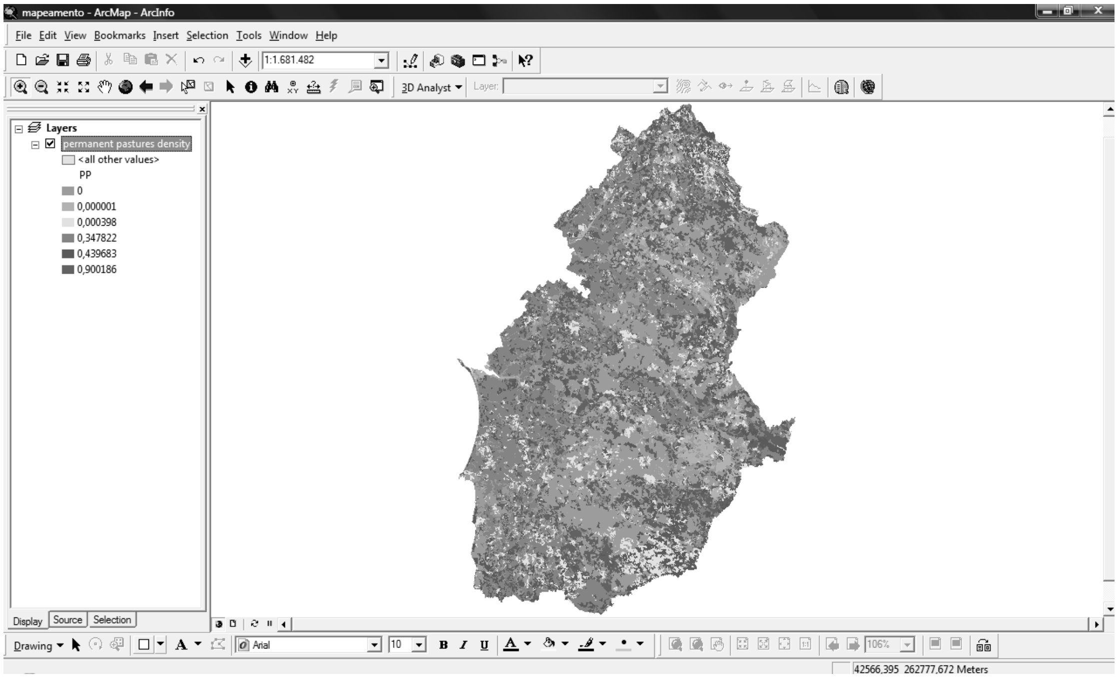

Figure 7. An Example of a Data Representation in ArcGIS

Table 1. Density Calculated for Each CLC 2000 Class, Using the Alto Alentejo Region's Counties

	CC	FF	PF	PC	PP	SF	OU
Annual crops	0.268051	0.244811	0.000000	0.009948	0.001924	0.000000	0.064669
Permanent crops	0.000001	0.000000	0.000000	0.734371	0.000000	0.000000	0.147200
Heterogeneous agricultural areas	0.000000	0.000000	0.028151	0.247437	0.080288	0.441495	0.000000
Forest	0.000000	0.000000	0.072692	0.000000	0.133718	0.149719	0.019241
Water courses	0.000002	0.000000	0.000000	0.000000	0.000000	0.000000	0.000002
Other spaces	0.000000	0.000000	0.000000	0.000348	0.000000	0.000000	0.587586
Pastures and shrubs	0.000000	0.000000	0.086346	0.000275	0.984666	0.001573	0.000000
Agro-forest systems	0.112720	0.478853	0.127148	0.000961	1.065190	0.000000	0.000009
Annual crops	0.268051	0.244811	0.000000	0.009948	0.001924	0.000000	0.064669

Notes: Column heading abbreviations are cereal (CC), fallow (FF), temporary forages and pastures (PF), permanent crops (PC), permanent pastures (PP), shrubs and forest without pastures (SF), and other occupations (OU).

Source: Model results.

Based on the defined densities at regional or sub-regional level, maps of the different variables' distribution were constructed. Figure 8 represents an example of a map of the distribution of densities for each CLC 2000 class regarding density and total area of permanent pastures. The

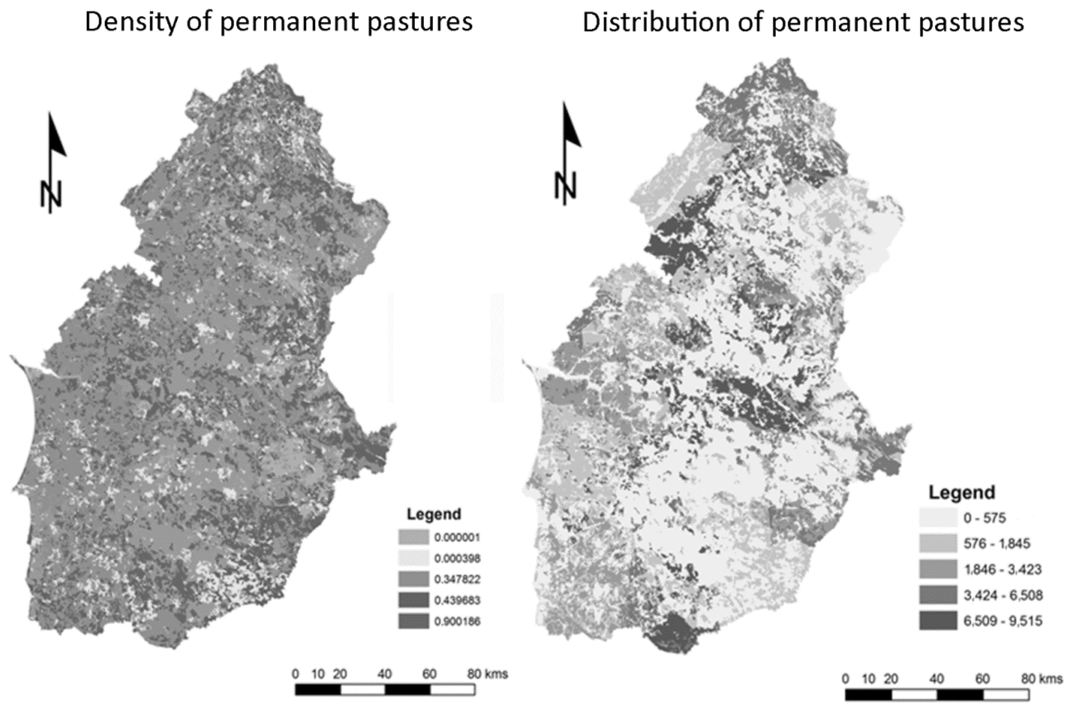

Figure 8. Final Results for the Permanent Pastures Variable

Source: Model results.

total number of created maps was 14 for the total of the variables' value over the distribution and 14 for the densities of all the variables.

Validation of the Results

Validation of the results was done using the data discussed earlier for each county, but also for some counties' parishes, in order to assess the limits of the results.

For measuring the deviations of each aggregated variable for the counties, one may use several deviation measures. In fact, a variety of methods for assessing the quality of results of GIS analyses can be used; they include, among others, root mean square (RMS) error, coefficients of variation, mean percent error, and relative error (Bielecka 2005).

Early studies, such as Eicher and Brewer (2001) and Mennis and Hultgren (2006), compared the estimated and observed data and constructed error maps. They used RMS as a numerical measure of error analysis because it can be easily applied to count data. Other authors, who presented an iterative disaggregation process (Gallego and Peedell 2001, Bielecka 2005), used only the percent error and the total error, since they are a part of the disaggregation process. The relative error gives us the weight of the estimated data in relation to the real data. Values close to one will indicate a good precision of the disaggregation process. Their formulation is exposed in equations (9) and (10).

The comparison between estimated and real data was based on the analysis of the following deviation indicators (Martins, Fragoso, and Xavier 2010, Fragoso, Martins, and Lucas 2008):

(13)

$$\mathrm{PAD}_s^i = \left|\frac{Y_s^i - \widehat{Y}_s^i}{Y_s^i}\right| \times 100, \quad \mathrm{WPAD}_s^i = y_s^i \left|\frac{y_s^i - \widehat{y}_s^i}{y_s^i}\right|,$$

$$\text{and } \mathrm{WPAD}^i = \sum_{s=1}^{s} \mathrm{WPAD}_s^i ,$$

and, at the aggregated level,

(14) $$\mathrm{WPAD} = \sum_{i=1}^{I} \frac{s^i}{S} \times \mathrm{WPAD}^i .$$

The prescription absolute deviation (PAD_s) is the absolute percentage variation of the estimated ($\widehat{Y}_s$) occupation relating the observed values (Y_s) (or instead the total resulting values) applied to each unit *i*. The weighted prescription absolute deviation ($WPAD_s^i$) is the deviation in each land use category in unit *i* weighted by its true importance or probability of occupation; $WPAD^i$ corresponds to the sum of the $WPAD_s^i$ values, giving the idea of the real total deviation for the values of the unit *i*. The WPAD corresponds to the weighted sum of the $WPAD^i$ by the weight or importance of each unit *i* regarding the total value.

The analysis of the resulting WPAD for the model's variant that used regional densities was 73 percent, which is a very unsatisfactory result. On the other hand, the $WPAD^i$ median was also very high, since its value was 57.6 percent. Better results were obtained using the sub-regional densities. The WPAD was 26.7 percent and the $WPAD^i$ median was 29.5 percent, which means that the total error for half of the Alentejo region's counties is very low.

Figure 9 represents the distribution of the $WPAD^i$ for each of the Alentejo region's counties. The analysis shows a clear improvement of the results using sub-regional densities. However, there are some counties for which the total error becomes unsatisfactory: Barrancos, Alandroal, and Alcácer do Sal.

To analyze the factors that contribute to the total error, we also constructed maps of each county's contribution to the WPAD (Figure 10). In the first map of Figure 10, the counties yielding a higher WPAD are Odemira, Castro Verde, and Mértola Counties (all situated in the sub-region Baixo Alentejo).

The second map in Figure 10 corresponds to the model's variant that used sub-regional densities, and reveals that the areas that yielded a higher WPAD are Alcácer do Sal, Montemor-o-novo, Évora, Alandroal, Beja, and Almodovar.

To understand which land uses had better results, we made an analysis of the PADs for each occupation (Table 2). The algorithm produced results that were considered satisfactory for most land use categories. The analysis revealed very satisfactory results; the worst results were for shrubs and forest without pastures (SF) land use.

In order to assess the precision of the disaggregation process for the Alentejo region's most important land use, we constructed a map for the permanent pastures (Figure 11). Its analysis shows that there are several counties with very satisfactory results in the Alto Alentejo and Alentejo Litoral areas. These satisfactory results are enhanced by the fact that the area used for permanent pastures is very high, and so it is not limited by the CLC 2000 minimum mapping unit.

Finally, in order to assess in more detail the precision of the disaggregation process, we selected the parishes of three sample counties from areas with different biophysical characteristics located in different areas of Alentejo. The counties were Ourique, Vidigueira, and Castelo de Vide. Figure 12 shows the location of these three counties.

The results obtained (Table 3) reveal a WPAD of 42 percent for the Castelo de Vide parishes, 156 percent for the Ourique parishes, and 40 percent for the Vidigueira parishes. However, very satisfactory results were obtained in some occupations. For instance, the permanent pastures in Castelo de Vide (which is the most important land use in this county) always have a PAD of 17 percent, the cereals in Vidigueira always have values of under 34 percent, and permanent crops and permanent pastures have very satisfactory values in almost all of the parishes of the counties.

The validity of the results is conditional on the precision of CLC 2000. In order to assess this situation, and due to the fact that the CLC has a minimum mapping unit of 25 hectares, the aggregation of some of these parishes into bigger units was made, leading to a substantial increase of the results' precision for this level of analysis. This is especially valid in parishes with similar biophysical conditions. For instance, in Ourique, one of the sampled areas with more errors, the simple aggregation of two parishes (Santana da Serra and Ourique) leads to a $WPAD^i$ of 7.4 percent. Figure 13 shows the PAD for the spatial unit resulting from the union of these parishes.

Concluding Remarks

This paper demonstrates the use of satellite-derived ancillary land cover data to map densities of several variables using conventional choropleth maps as source zones. We proved that the combination of the method proposed by Gallego and Peedell (2001) and the binary method may provide better results than the use of only one of

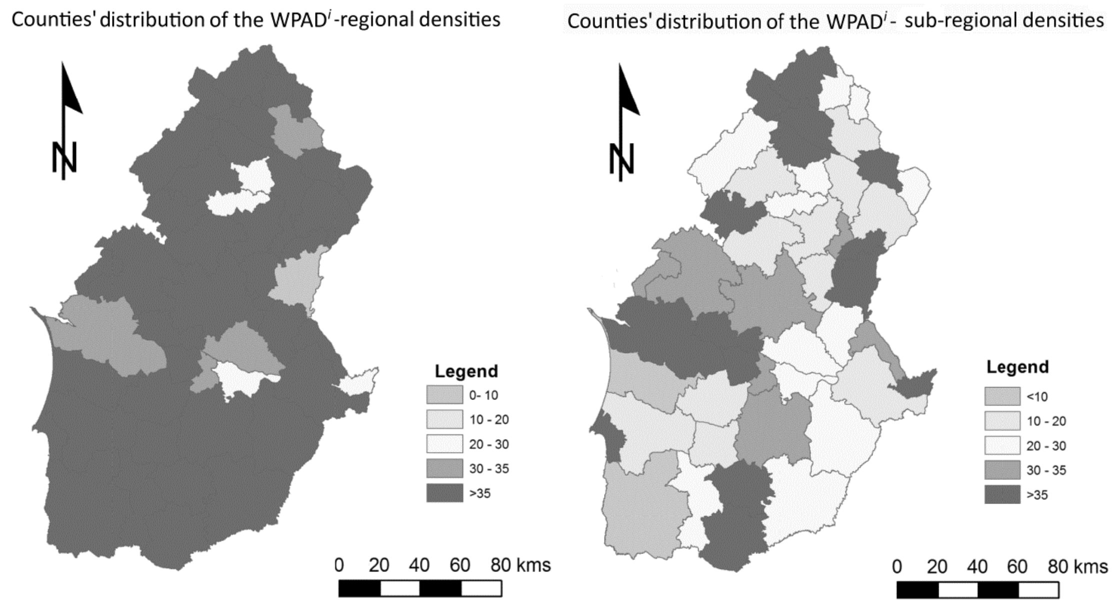

Figure 9. WPAD^i Using Regional and Sub-Regional Densities

Source: Model results.

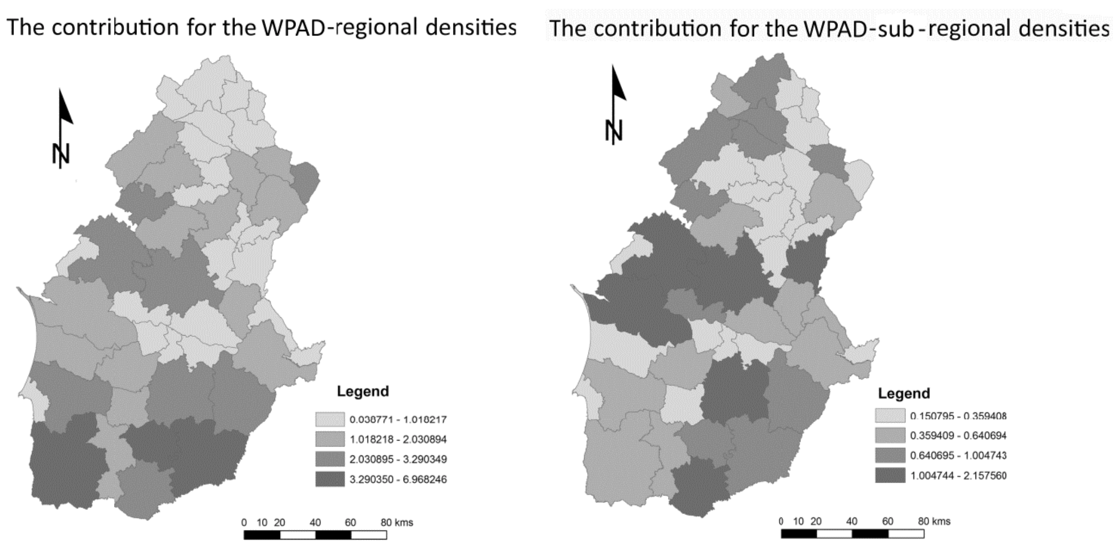

Figure 10. Counties' Contribution to the WPAD

Source: Model results.

Table 2. Synthesis Indicators of the Validation of the Result

	Sub-Regional Densities				Regional Densities			
	Median	Average	Max.	Min.	Median	Average	Max.	Min.
CC	25.3	60.6	482.9	0.1	30.7	66.5	663.2	1.9
FF	31.5	79.5	935.4	2.0	199.7	262.9	1,168.4	37.2
PF	16.4	25.1	148.7	0.5	66.0	59.5	158.9	0.8
PC	11.4	17.3	255.3	0.3	14.5	19.6	61.8	0.4
PP	22.1	38.6	232.5	1.5	24.7	41.4	254.8	0.6
SF	37.7	54.2	405.1	1.5	38.3	63.2	384.0	1.1
OU	26.4	38.9	241.0	0.1	31.9	46.4	227.8	0.2

Notes: Row abbreviations are cereal (CC), fallow (FF), temporary forages and pastures (PF), permanent crops (PC), permanent pastures (PP), shrubs and forest without pastures (SF), and other occupations (OU).

Source: Model results.

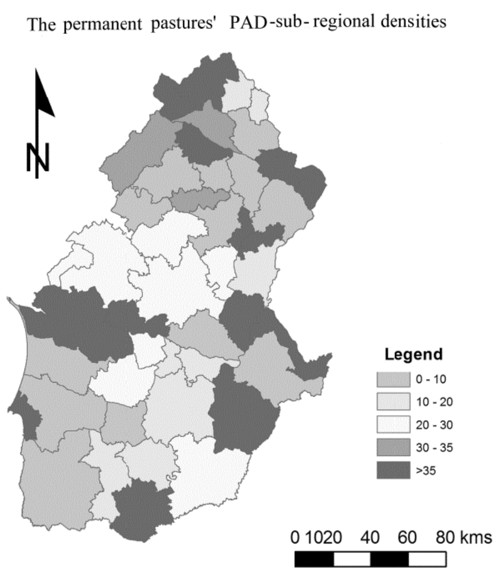

Figure 11. Permanent Pastures' Prescription Absolute Deviation Based on Sub-Regional Densities

Source: Model results.

these. This allows the maximization of benefits and the exclusion of several disadvantages of each of these methods used in an isolated way.

The method developed may easily solve problems regarding the lack of data that could not be completely overcome by previous studies (e.g., Martins, Fragoso, and Xavier 2010, Fragoso, Martins, and Lucas 2008, Howitt and Reynaud 2002), since it allows a good approximation of the real distribution of some farms' data, and can also be used to enhance and add additional information to those models that allow disaggregation only by administrative or statistical units. Therefore, this method may be used in the sustainable management of different rural areas as it will be able to provide a more detailed information background.

However, there are still some criticisms that may be made of this model. Néry et al. (2007) suggest that the iterative algorithm proposed by Gallego and Peedell (2001) does not correctly adjust distributions to meet the pycnophylatic property. This occurs because global residuals are used to calculate distributions rather than zonal values—leading to discrepancies between the predicted and observed totals within zone.

New methods of improving this methodology are therefore needed. It is a fact that this methodology proves to be useful, but there are still several questions that need to be answered, namely the possibility of the method's application to true homogenous areas and the possibility of using an alternative algorithm that distributes the residual values in a different way.

Therefore, in the future, we intend to use and test other dasymetric mapping methodologies and

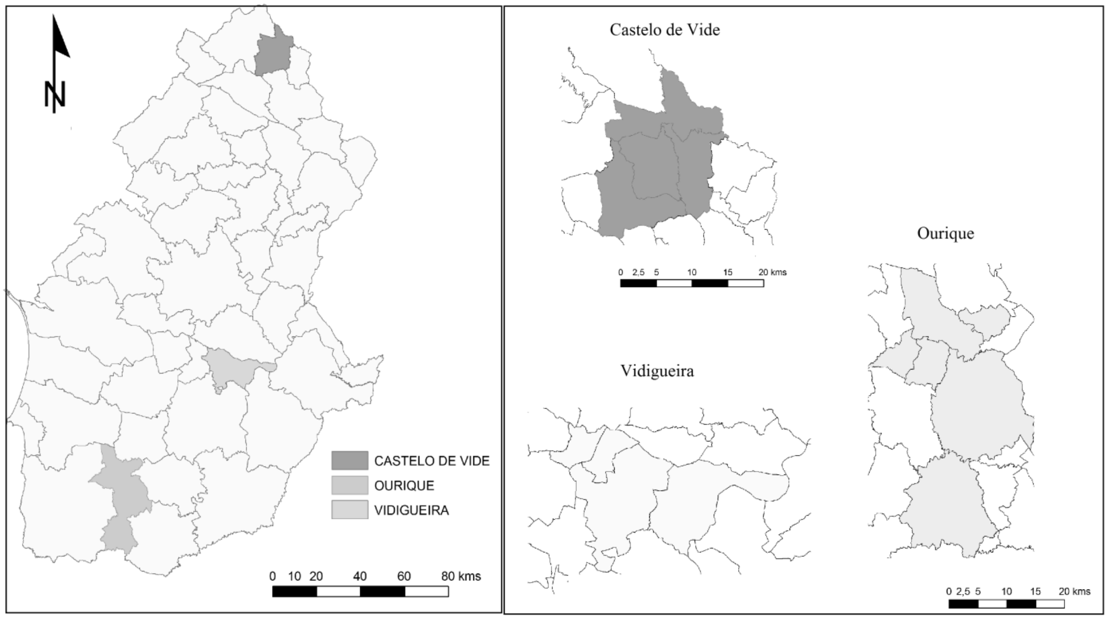

Figure 12. Location of the Sampled Parishes

Table 3. Synthesis Indicators of the Validation of the Results for the Sampled Parishes

	Vidigueira		Ourique		Castelo de Vide	
	Median	Average	Median	Average	Median	Average
CC	32	27	33	52	268	397
FF	185	239	115	139	154	171
PF	52	134	145	1,675	38	50
PC	11	12	24	90	33	27
PP	32	52	48	44	21	19
SF	914	4,097	296	610	579	558
OU	59	81	82	168	182	155

Notes: Row abbreviations are cereal (CC), fallow (FF), temporary forages and pastures (PF), permanent crops (PC), permanent pastures (PP), shrubs and forest without pastures (SF), and other occupations (OU).

Source: Model results.

also to improve the existing one with new developments and better integration in the GIS environment, allowing a better and faster result for the end user. The current computer tool is under development and, in spite of its simplicity, we aim to improve it further by identifying the possible weaknesses that it may have.

The results will also be enhanced in the near future by the possibility of using more accurate data, namely the 2009 Agricultural Census, and the Soil Occupation Maps (the COS 2007 cartography), which will soon be available. The results will also be more precise with the inclusion of other information priors such as newly available

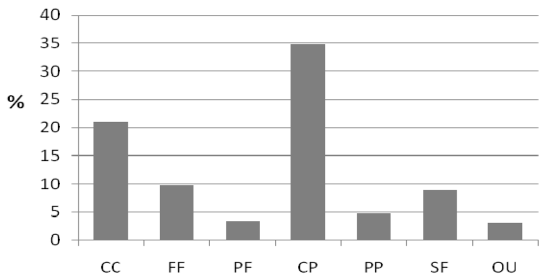

Figure 13. Prescription Absolute Deviation (PAD) for the Aggregated Parishes Unit (Ourique and Santana da Serra)

Source: Model results.

cartographic information that complements experts' opinions. This will allow us to obtain very satisfactory results at the parish level and to overcome the limitations detected in the sampled parishes.

We recognize that further improvements can and should be made to this algorithmic procedure. Instead of being the final statement on the topic, we are asking a question relevant to the body of knowledge, and presenting a procedure that adds to the body of knowledge with benefits to academics and practitioners. We have identified what to us are obvious next-step research questions whose answers will increase the utility of the procedure. The issue we are identifying is large, the benefits equally large, so we encourage others to suggest methods for improving what has been presented here.

References

Bielecka, E. 2005. "Dasymetric Population Density Map of Poland." *Proceedings of the 22nd International Cartographic Conference* (in Coruña, Spain).

Direcção Regional de Agricultura e Pescas do Alentejo (DRAPAl). 2007. "Programa de Desenvolvimento Rural-Alentejo" (Rural Development Program—Alentejo). DRAPA1, Évora, Portugal.

Carvalho, M., and M. Godinho. 2004. "A Nova Reforma da Política Agrícola Comum e Suas Consequências num Sistema Agrícola Mediterrâneo de Portugal" (The New Common Agricultural Policy Reform and Its Consequences on a Mediterranean Agricultural System in Portugal). Available at http://www.sober.org.br/palestra/2/1099.pdf (accessed January 15, 2009).

Eicher, C., and C. Brewer. 2001. "Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation." *Cartography and Geographic Information Science* 28(2): 125–138.

Fragoso, R., M. Martins, and R. Lucas. 2008. "A Minimum Crossed Entropy Model to Generate Information to the Disaggregated Level." In *Proceedings of the 4th IASME/WSEAS International Conference on Energy, Environment, and Sustainable Development*. WSEAS Press, Faro, Portugal.

Gallego, F.J., and S. Peedell 2001. "Using CORINE Land Cover to Map Population Density: Towards Agri-Environmental Indicators." Topic Report No. 6/2001, European Environment Agency, Copenhagen.

Howitt, R., and A. Reynaud. 2002. "Spatial Disaggregation of Production Data by Maximum Entropy." Paper presented at the Tenth Congress of the European Association of Agricultural Economists, Saragoça, Spain, August 28–31.

Instituto Nacional de Estatística (INE). 1989. "Recenseamento Geral da Agricultura de 1989." INE, Lisbon, Portugal.

____. 2001. "Recenseamento Geral da Agricultura de 1999." INE, Lisbon, Portugal.

____. 2002. "Recenseamento Geral da População—Censos 2001." INE, Lisbon, Portugal.

____. Various years. "Estatísticas Agrícolas." INE, Lisbon, Portugal.

____. Various years. "Anuários Estatísticos da Região do Alentejo." INE, Lisbon, Portugal.

____. Various years. "Inquérito à estrutura das explorações agrícolas." INE, Lisbon, Portugal.

Langford, M., D.J. Maguire, and D.J. Unwin. 1991. "The Areal Interpolation Problem: Estimating Population Using Remote Sensing in a GIS Framework." In *Handling Geographic Information*. Essex, UK: Longman Scientific and Technical.

Langford, M., and D. Unwin. 1994. "Generating and Mapping Population Density Surface Within a Geographical Information System." *Cartographic Journal* 31: 21–26.

Martins, M.B., R. Fragoso, and A. Xavier. 2010. "Recovery of Incomplete Agricultural Land Uses and Livestock Numbers by Entropy." Paper presented at the annual meetings of the European Association of Agricultural Economists in Berlin, Germany (April 15–16).

McClean, C.J. 2007. "The Spatial Disaggregation of GB and European Agricultural Land Use Statistics." Paper presented at the GISRUK (GIS Research UK) meetings in Maynooth, Ireland (April 11–13). Available at http://ncg.nuim.ie/gisruk/materials/proceedings/PDF/7B4.pdf (accessed February 20, 2011).

Mennis, J. 2002. "Using Geographic Information Systems to Create and Analyze Statistical Surface of Population and Risk for Environmental Justice Analysis." *Social Science Quarterly* 83(1): 281–297.

____. 2003. "Generating Surface Models of Population Using Dasymetric Mapping." *The Professional Geographer* 55(1): 31–42.

Mennis, J., and T. Hultgren. 2006. "Intelligent Dasymetric Mapping and Its Application to Areal Interpolation." *Cartography and Geographic Information Science* 33(3): 179–194.

Mrozinski, R., and R. Cromley. 1999. "Singly—and Doubly—Constrained Methods of Areal Interpolation for Vector-Based GIS." *Transactions in GIS* 3: 285–301.

Néry, F., P. Monterroso, A. Santos, and J. Matos. 2007. "Interpolação Zonal de Estatísticas Socio-económicas." Paper presented at the Conferência Internacional de Cartografia e Geodésia, Lidel, Lisbon (April 19–20). Available at www.igeo.pt/instituto/cegig/got/3_Docs/papers_Main_PT.html (accessed January 15, 2009).

Openshaw, S. 1984. "The Modifiable Areal Unit Problem." *Concepts and Techniques in Modern Geography* 28: 38–41.

Painho, M., and M. Caetano. 2006. *Cartografia de Ocupação do Solo de Portugal Continental 1985–2000 Corine Land Cover 2000*. Amadoram Agência Portuguesa do Ambiente.

Tobler, W.R. 1979. "Smooth Pycnophylactic Interpolation for Geographical Regions." *Journal of the American Statistical Association* 74 (367): 519–530.

Wright, J.K. 1936. "A Method of Mapping Densities of Population with Cape Cod as an Example." *Geographical Review* 26(1): 103–110.