Clustering of territorial areas: A multi-criteria districting problem

Rui Fragoso, CEAFGE-UÉ and ICAAM-UÉ, rfragoso@uevora.pt

Conceição Rego, CEFAGE-UÉ, mcpr@uevora.pt

Vladimir Bushenkov, CIMA-UÉ, bushen@uevora.pt

Abstract

Endogenous resources, economic profile and socio-economic issues are the criteria that define the development level and the identity features of a territorial unit. The territorial units that organize the country, in political and administrative terms – parishes and counties –, have a hierarquical structure, which initially reflected the organization of productive activities as well as the tradition State organization. The success of development policies addressed to territorial agglomerates depends on its homogeneity and of their territorial units. Facing to this the clustering of territorial areas can be stated as a districting multi-criteria problem. Thus, this paper aims to propose a framework for obtaining homogenous territorial clusters based on a Pareto frontier that includes multi-criteria related to the territorial endogenous resources, economic profile and socio-cultural features. This framework is developed in two phases. First, the criteria correlated with the development at the territory unit level are determined through statistical and econometric methods. Then, a multi-criteria approach is developed to allocate each territory unit to an agglomerate of territory according to the Pareto frontier established. The framework is applied to the context of a set of parishes and counties of the Alentejo Central region, southern Portugal. Results are presented and discussed in the scope of a regional strategy of development.

Key-words: Alentejo; Cluster; Districting; Multi-criteria.

CLUSTERING OF TERRITORIAL AREAS: A MULTI-CRITERIA DISTRICTING PROBLEM

1. Introduction

Each unit of territory is defined according with its endogenous resources, economic profile and socio-economic issues which are criteria associated with their development level and identity features. The territorial units that organize the country in political and administrative terms – parishes and counties – have a hierarchical structure which reflects the organization of productive and social activities. The success of development policies addressed to territorial agglomerates depends on its homogeneity and of their territorial units. According to Fischer (1980), a homogeneous region is a set of spatially contiguous areas, which show a high degree of similarity regarding a set of attributes.

In Portugal, the policy makers has discussed, and approved, a new territorial administrative organization, of parishes and counties[1]. Some of the main purposes of this reform, are, inter alia, promote territorial cohesion and local development; improve and develop local public services as well as promote economies of scale, efficiency and critical mass in the parishes and municipalities.[2] In the end of this reform, there will the reduction of the number of parishes.

This is can be stated as a districting multi-criteria problem where elementary units of territory are aggregated into clusters or larger districts and a district map or partition is produced (Tavares-Pereira, 2007). The districting approach has been widely used to treat several kind of problems related with the definition of electoral districts (Bozkaya et al., 2003 and 2009), working zones for a travel salesperson team (Zoltners and Sinha, 1983), areas in metropolitan internet networks for installing hubs (Park et al, 2000), areas of manufactured and consumer goods (Flischmann and Paraschis, 1988), school districting (Ferland and Guénette, 1990) and electric power zones (Bergey at al., 2003a). According to Tavares-Pereira (2004) these kind of districting problems are frequent in

[1] Law nº 11-A/2013, 28th June, designated "Administrative Reorganization of Parishes Territory".

[2] Article 2 of the law nº 22/2012, 30h May, designated "Approves the legal regime of administrative reorganization of the municipalities territory".

real world and involve multiple criteria, which are often incommensurable and conflicting.

The districting problem can be stated as the partition of the territory into homogeneous clusters assessed by multiple criteria. The result is a set of homogeneous districts or areas, which are composed by elementary units of territory. Each district is associated to a set of constraints, such as technical, economic, ecological, social and others. According to constraints considered and criteria used in the assessment process, different solution or maps can be obtained. Thus, "the best solution" will be probably a compromise or a non-dominate solution in which the improvement on a criteria leads to a worse result on at least one of the remaining criteria.

The territory partition problems that led first to using scientific methodologies were the electoral districting problems where the main purpose was the construction of political districts generated by impartial processes (Mehrotra, 1998). Vickrey (1961) presented one the first works made about this topic, where is described the heuristic process used for constructing a zone. Hess et al. (1965), were the first to propose a mathematical programming model which states the districting problem as a location/allocation problem. However, the problem of salespersons is that about more works have been made. Generally the main objective is balancing the workload among different zones (Easingwood, 1973; Hess and Samuels, 1971; Shanker et al., 1975; Zoltners and Sinha, 1983).

In the four last decades many new developments and applications about aggregating areas into homogeneous areas through districting problems have been made and new challenges have emerged (Duque et. al, 2011). One is the need to have simpler and systematic frameworks that allow aggregate areas into homogeneous regions by allocating elementary units of territory to districts and determining the number of districts and of units of territory in the sequence of merging or partition processes.

Facing to that, this paper aims to propose simpler and systemic districting framework, which involves multiple criteria and exact algorithms to generate the non-dominate efficient solutions. A multi-criteria programming model for allocating elementary units of territory to districts is developed and the attributes of the multiple decision criteria are found considering the level of correlation between the different socio-economic variables involved. The model is applied to a set of parishes in the

Alentejo Central region, southern Portugal, with the purpose to find the structure of parishes that lead to a more efficient development process.

The remainder of the paper is organized as follows. The proposed general framework is presented in the next section. The attributes of multiple criteria are established in Section 3. The formulation of a multi-criteria programming model is presented in Section 4. Results are presented and discussed in Section 5. Finally, in Section 6, the main conclusions and suggestions are provided.

2. General framework

2.1. Political and administrative territorial organization

The country political and administrative organization reflects historical reasons but also the distribution of the population, the urban structure and the economic and social activities across space.

The theories developed in the field of economic geography (eg, Cantillon in 1730 or later Christaller in 1933; these studies, among others, are described in Simões Lopes, 1984), clearly show the association between the hierarchical structure of urban places and labour needs of productive activities. The management structure of the State is not disconnected from the way the population and productive activities occupy the territory.

The administrative Portuguese political system has its origins in the nineteenth century (Pereira, 1995). Currently, the administrative structure maintains features of Napoleonic model of organization of the state, including the strong centralism. The administrative and political Portuguese structure is organized as follows: a central government and 308 local governments (municipalities), reflecting the strong Portuguese local tradition (municipalities). At the local level there are two types of government: the municipalities and parishes. The central government and the local authorities are both directly elected by universal and secret ballot. Central government functions fit to legislate and enforce the policy for the whole country. Local government, with a more executive nature, prosecute, among others, planning and landscape management functions, promotion of economic development and territorial infrastructures management. According to the proposed by Musgrave (1974) to the central government must fit the features of the income distribution and economic stabilization, while local authorities will fit the activities of providing local goods and services.

Thus, the central government and local government currently exercise public administration with one government, 308 municipalities and 4,400 parishes. Shortly, and in accordance with Law No. 11-A/2013, the number of parishes will decrease to around 3091, thus reducing the state's presence throughout the territory, in terms of their basic administrative units. The amendment, approved by Law No. 22/2012, aims (article 2 of Law No. 22/2012), among others, to promote territorial cohesion and local development, promote the improvement and development of local public services, promote the attainment economies of scale, efficiency and critical mass of local and restructure, by aggregation, of a significant number of parishes across the country, with special focus on urban areas. The administrative reorganization of the state, reflected in the reduction in the number of parishes, should be conducted (article 8 of Law No. 22/2012), among others, that the county will be preferentially considered as a pole of attraction of the parishes adjoining, as well as the parishes that have a higher rate of economic and social development, a greater number of people and a greater concentration of community facilities.

In essence, this amendment intends to reflect the need to reorganize the state's presence in the territory, taking into account new forms of organization of economic activities, population distribution between urban and rural areas as well as the ease of transportation and communication between different places.

2.2. Districting problems

The techniques used in districting problems can be based on the concept of division, in which the territory is considered as a whole and is devised into pieces, or based on the concept of agglomeration, in which the territory is composed by a set of elementary units (Cortona et al., 1999). The districting problems can involve only one criterion, such as voting potential equality or workload equality (Grafinkel and Nemhauser, 1970; Hess et al., 1965; Hojati, 1996), or multiple conflicting criteria (Bergey et al. 2003a; Bourjolly et al., 1981; Bozkaya et al., 2003; Deckro, 1979). The criteria can be used according to a fixed hierarchy reflecting the decision-maker preferences or integrated in a mixed objective function. The type of approach can be classified in exact and non-exact algorithms (Mehrotra, 1992; Bergey, 2003b; Muyldermans et al., 2002).

In this case, a technique based on the agglomeration of elementary territory units was chosen. The agglomeration of territory units into homogeneous districts was treated by a host of authors that have focused on spatial continuity of territory units, ways to measure territorial homogeneity and strategies to explore the solution space efficiently and to check its feasibility (Byfuglien and Nordgärd, 1973; Lefkovitch, 1980; Ferligoj and Batagelj, 1982; Legendre, 1987; Murtagh, 1992; Maravalle and Simeone, 1995; Gordon, 1996; Wise et al., 1997; Hansen et al., 2003; and Duque et al., 2012). One of the main challenges of these works was the definition of the number of regions that should be created.

In order to address those issues in a simple and systematic way, this paper considers an analytical framework based on the max-*p*-regions problem of Duque et al. (2011). The max-*p*-regions problem is a formulation that involves exact algorithms to generate non-dominate solutions, in which n territory units are aggregated into a maximum number of districts ensuring that each district satisfies an imposed minimum threshold value (th). This threshold value is a predefined spatially attribute, such as district population, district surface or other district feature.

In this approach the number of districts is modelled as an endogenous variable and in opposite to other existing approaches, the spatial contiguity constraint is satisfied without imposing constraints on the shape of districts, such as the maximum compactness. Thus the max-*p*-regions model is a suitable tool to be used in applied analysis without subjectivity in the definition of both scale (number of districts) and aggregation of elementary territory units (shape of districts). According to Duque et al. (2012) the max-p-regions formulation is presented below.

Let be $A=\{A_1, A_2,...,A_n\}$ a set of elementary territory units which can be described by the attributes $y \in Y =\{1,2,...,m\}$ with $m \geq 1$ and l_i is a spatially extensive attribute of the territory unit A. In this context it is also necessary to consider the dissimilarities between territory units $d_{ij} \equiv d(A_i,A_j)$ and the continuity graph $W=(V,E)$ associated to A, such that the vertices $v_i \in V$ correspond to territory units $A_i \in A$ and edges $\{v_i,v_j\} \in E$ if and only if territory units A_i and A_j share the same border. The partition of territory units A into p districts R can be denoted by $P_p=\{R_1,R_2,...R_p\}$ such that:

$$R > 0 \wedge R_k \cap R_{k'} = 0 \quad for\ k, k' = 1,2, \ldots, p \wedge k \neq k'$$

$$\bigcup_{k=1}^{p} R_k = A \ \wedge \sum_{A_i \in R_k} l_i = threshold$$

Considering as evaluation criterion for a feasible partition $P_p \epsilon \Pi$ the heterogeneity of district k with $R_k \epsilon P_p$ and the total heterogeneity of partition $P_p \epsilon \Pi$ are respectively:

$$h(R_k) = \sum_i \sum_j d_{ij} \quad \text{and} \quad H(P_p) = \sum_{k=1}^{p} h(R_k)$$

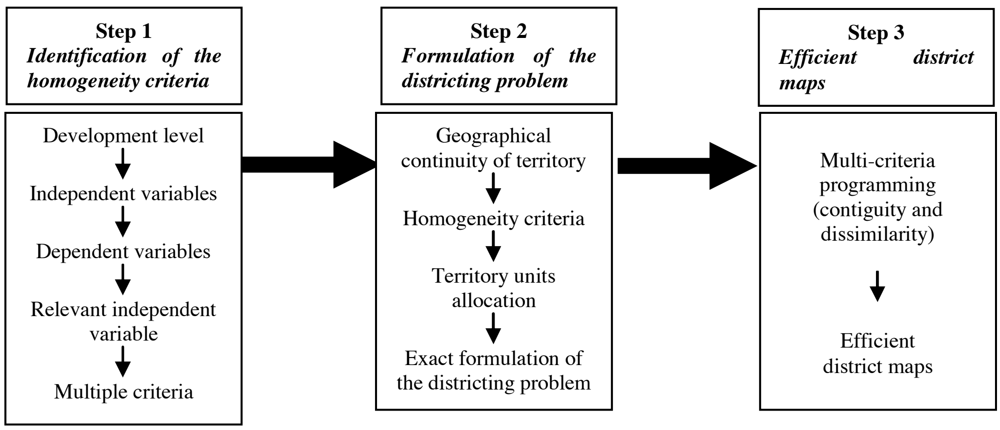
Thus, the max-p-region can be formulated as:

Determine $P_p^* \in \prod$ such that $|P_p^*| = max(|P_p| = P_p \in \prod \quad)$, and

$\nexists P_p \in \prod: |P_p| = |P_p^*| \wedge H(P_p) < H(P_p^*)$

Compared with the max-*p*-regions approach of Duque et al. (2012), the approach proposed in this paper considers multiple criteria related with socio-economic profile of the elementary territory units for assessing its dissimilarity. Thus, the proposed general framework is preformed in three steps as it is shown in Figure 1.

Figure 1. General framework steps



Step 1 respects the identification of homogeneity criteria for assessing dissimilarity between territory units. This is based on casual relations that can be established between independent and dependent variables that are the most important in the definition of the development level and socio-economic profile of territory units. In the step 2 an exact formulation of the districting problem is performed based on the max-*p*-regions approach. The model developed maximizes number of districts composed by contiguous territory units and simultaneously minimizes its dissimilarity. Finally in the step 3, the general framework comprises the simulation of different imposed thresholds of predefined spatially attributes and hence the construction of efficient district maps.

3. The attributes of multiple criteria

Edmonton's municipal electoral districts in Canada are defined based on a set of socio-economic criteria which includes population equality among districts, future growth, community league boundaries, compactness, communities of interest, least number of changes and contiguity (Bozkaya et al., 2011). In their original work, Bozkaya et al. (2003) modelled the districting criteria into a weighted objective function, which formulation included the minimization of district population deviation from average, compactness, socio-economic homogeneity, similarity to the existing plan and maintenance of the communities of interest.

In order to test the research hypothesis that the spatial clustering of urban localities helps to explain their population growth, Portnov and Schwartz (2009) used data on Europe's settlements. Multiple regression analysis, using both least square and spatial lag models, was applied to assess the effect of several factors on the annual population growth of urban localities. The annual population growth was treated as the absolute rate of population growth per 1000 residents and in a standardized way, as the difference between the local population growth rate and that of the whole country. As explanatory variables of annual population growth, it was considered the following factors: local population size (ln); distance to the sea shore (Km); distance to the major city (Km) and the interaction term between a place's latitude and its elevation above the sea level.

In our case to establish the multiple criteria for assessing the dissimilarity between territory units, it was considered the effect of several factors on the local development, having simultaneously into account the available data at parish level. For this geographic disaggregation level the only available data are from the Census of Population of 2011 from the National Office of Statistics (INE, 2011).

Thus as a proxy of the development index of elementary territory units (parishes) was considered the population growth rate between from 2001 to 2011, i.e., the population grow rate between the last two census and to chose the multiple criteria of dissimilarity was determined casual relationships with several factors through multiple linear regression analysis, using the least square model.

Casual relationships between dependent and independent variables with high statistic significance levels allow identifying the variables that better explain the development index and hence those that could be also chosen as attributes of homogeneity criteria that will be used latter in the max-*p*-regions model to assess the dissimilarity between territory units.

The analysis started with the multiple linear regression analysis between population growth rate (2001-2011) and the set of explanatory variables presented in Table 1. Explanatory variables can be grouped into the following four types: territorial variables; population structure; population qualification; and economic indicators, which measure the intensity of the economic activity.

Table 1: Variables used in the multiple linear regression analysis

Type of variable	Variables
Territorial variables	Population density (%) Distance to major centre (Km) Percentage of total surface (%)
Population structure	Percentage of total population (%) Population average age (years old) Total dependence index (%) Age dependence index (%) Potential sustainability index (%)
Population qualification	Percentage of population with high school (%) Percentage of population with higher education (%) Illiteracy rate (%) School abandon rate (%)
Economic indicators	Active population rate (%) Employed population rate (%) Employed population rate in primary activities (%) Employed population rate in secondary activities (%) Employed population rate in tertiary social activities (%) Employed population rate in tertiary economic activities (%) Unemployment rate (%)

Source: Personal elaboration

As noted earlier, this study is intended to identify homogeneous territorial units in terms of development. In this case we will not make distinction between development and economic growth. Our dependent/interest variable (rate of population growth between 2001 and 2011) is used as a proxy for economic growth of the territory under study. It should be noted that for these territorial levels no available statistical information about the variable that typically measures the evolution of income (per capita GDP). The dimensions selected to find the differences in terms of the development of the territory go back to what is suggested in the literature on economic growth and development. According, for example, with the concept of human development proposed by UNDP (the United Nations program for development), the indicators used by this organization are divided into 3 categories / dimensions (Diniz, 2010: 50): a long life health, knowledge and a decent standard of living. These dimensions are here discussed, respectively, through the variables relating to Population Structure, Population Qualification and Economic indicators. Furthermore, we integrate the study variables characterizing the territory (territorial variables) that may contribute to differentiate the levels of development of the various territorial units under review.

After one has to be checked the hypothesis of linear regression, namely, linearity, normality and co-linearity, an analysis to estimated coefficients and respective values of standard deviation was performed. In order to reduce the number of explanatory variables and hence the multiple criteria to be used to assess the dissimilarity between territory units, the correlation between explanatory variables was calculated and the *t* student statistic and hence the statistical significance level of coefficients were evaluated. Thus, the variables with lowest significance level were deleted from the model and a new regression was established between the population growth rate and the new set of explanatory variables. The explicative power of both regressions was assessed using R, R square and adjusted R square. This procedure is an interesting advantage of this framework, once it allows choosing multiple criteria based on the variables that are the most related with socio-economic profile of each territory unit.

The study we develop derives from the data collected for the parishes of a set of counties of Alentejo Central, around the municipality of Évora.

4. The multi-criteria max-*p*-regions model

The general multi-criteria program can be written as follows:

$$max\{f_1(x) = z_1\}$$
$$max\{f_2(x) = z_2\}$$
$$\vdots$$
$$max\{f_m(x) = z_m\}$$
$$s.t. \quad x \in X$$

or, $\quad "max" \quad Z=\{G(x) = z \in \mathbb{R}' | x \in X|\}$

where, x is the vector with n decision variables; X is feasible region of the decision space; m is the number of criteria; f is a real function defined $\mathbb{R}^M$; z is a criteria function value; "*max*" is the sense of the optimization meaning in this case the purpose is to maximize all criteria simultaneously; G is a vectorial function composed by m criteria; and Z is the feasible region in the criterion space.

A vector $\bar{z} \in Z$ is non-dominated, if and only if, it does not exist another vector $z \in Z$, such that $z \geq \bar{z}$ and $z \neq \bar{z}$. The set $Z^{nd} \subseteq Z$ of all non-dominated criteria vectors is called the Pareto frontier. Then a solution $\bar{x} \in X$ is efficient or Pareto optimal if the corresponding set of criteria $Z = G(\bar{x})$ is non-dominated.

The exact formulation of the max-*p*-regions model can be written based on the general formulation of the multi-criteria program as follows:

(1) $Min\{-f_1(x) + f_2(t)\}$, with

$f_1(x) = \sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k,0} . 10^h$ and $h = 1 + log\left(\sum_i \sum_j d_{i,j}^y\right)$

$$f_2(t) = \sum_i \sum_j \sum_y d_{i,j}^y . t_{i,j}$$

Subject to,

(2) $\sum_{i=1}^{n} x_i^{k,0} \leq 1 \qquad \forall k = 1, \dots n$

(3) $\sum_{k=1}^{n} \sum_{c=0}^{q} x_i^{k,c} \leq 1 \qquad \forall i = 1, \dots n$

(4) $x_i^{k,c} \leq \sum_j x_j^{k,(c-1)} \qquad \forall k = 1, \dots n;\ \forall i = 1, \dots n;\ \forall c = 1, \dots q$

(5) $\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{k,c} . l_i \geq th. \sum_{i=1}^{n} x_j^{k,0} \qquad \forall k = 1, \dots n$

(6) $t_{i,j} \geq \sum_{c=0}^{q} x_i^{k,c} + \sum_{c=0}^{q} x_i^{k,(c-1)} \quad \forall i,j = 1, \dots n;\ \forall k = 1, \dots n$

where, *i* and *I* are the index and of elementary territory units, *I={1,...,n};*

k is the index of potential districts, *k={1,...,n}*;

c is the index of contiguity order, *c={1,...,q}*, with *q=(n-1);*

y is the index of the attributes that describes *i* territory units;

$d^y_{i,j}$ is the parameter of dissimilarity relationships between territory units *i* and *j*,

with $i,j \in I$ under attribute *y*;

l_i is the parameter of spatially extensive attribute value of territory unit *i*;

th is the parameter of the minimum value of the attribute 1 at the districting scale;

In this formulation the decision variables are following binary endogenous variables such that:

$$x_i^{k,c} = \begin{cases} \text{1, if territory unit i is allocated to district } k \text{ in order } c \\ \text{0, otherwise} \end{cases}$$

$$t_{i,j} = \begin{cases} \text{1, if territory units i, j belong to the same district k} \\ \text{0, otherwise} \end{cases}$$

In this formulation the optimal *p* number of districts *k* is unknown and when a district is created, it starts by its "root" elementary territory unit, which is assigned with order zero in district *k* ($x_i^{k,0}$). This model ensures that territory units *i* are assigned to district *k* according to the territory units adjacent to the "root" territory unit *k0*.

This is a mixed integer programming (MIP) model formulated as a multi-criteria program, which the objective function (1) maximize the *p* number of potential districts *k* comprise by adjacent territory units *i*, while minimize the dissimilarity between territory units *i* and *j*. The two criteria considered in the optimization are not weighted, as usually. Instead, they are merged in a single value and the first term is multiplied by the scaling factor *h*, in order to achieve a hierarchy in which the number *p* of districts *k* comes first that the goal of minimize dissimilarity between territory units.

The dissimilarity goal depends on the binary variable value $t_{i,j}$ and the parameter $d^y_{i,j}$ of the dissimilarity relationships between territory units under attributes *y*. The parameter $d^y_{i,j}$ is the difference between the normalized values of attributes *y* in territory units *i* and *j*. As the dissimilarity goal is a single criteria the values of parameter $d^y_{i,j}$ have to be aggregated by addition of all y attributes into a single value for each pair of *i* and *j* territory units.

The objective function will improve until a big enough value of *p* is attained such that, this solution will be preferred to any other with a small value of *p*. For the same value of *p*, solutions with lower dissimilarity will be preferred over other with higher dissimilarity. Although the value of objective function and decision variables is subject to the set constraints (2) to (6).

Constraint (2) ensures that each district *k* should not have more than one “root” territory unit, which is assigned with a order of zero (*c=0*). Constraint (3) imposes that to each elementary territory unit *i* should correspond at least one district *k* respecting the contiguity order *c*. According with constraint (4) any territory unit *i* is allocated to a district *k* at order *c*, if an adjacent territory unit *j* of *i* is also allocated to the same district *k* at order *c-1*.

In constraint (5) the value of spatially extensive attribute is calculate for each district *k* and has to be greater or equal to a minimum threshold, which is an exogenous parameter. This constraint plays an important role, once the number of districts created by the model is very sensitive to the value of the predefined threshold (*th*).

Finally, constraint (6), which allows to determine the pairwise of adjacent territory units *i* and *j* that should be considered for calculating the total dissimilarity at the objective function in the term of $f_2(t)$.

5. **Results**

The identification of variables that explain the population growth rate (2001-2011) started with the multiple linear regression analysis with a set of explanatory variables (cf. table 1). The results of model 1 (table 2) show us that the variables *Population average age*, *Percentage of population with high school* and *Active population rate* are the most significant to explain the dependent variable ($p<0,05$). These results are in according with the expected, despite the p-value is not highly robust (best critical value is 0.005). The literature relating to the economic growth shows that aged populations have lower levels of growth since individuals of childbearing potential are relatively fewer. The proportion of active population is positively related to population growth. This is a fundamental relationship in the field of economic growth. Usually, higher qualification levels of the population correspond to larger population growth as well as larger economic growth. Typically, this relationship is reflected by the ratio of the population with higher education. In this case, we find that *Percentage of population with high school* is negatively associated with population growth, ie, intermediate qualification levels does not reveal fundamental for population growth. This estimation has a reasonable explanatory capacity (R^2 = 0,761).

Following the results obtained, were reduced the number of variables corresponding to the attributes which will be used to assess the dissimilarity. The estimation of model 2 (cf. table 2) was based on the variables identified as most significant in the model 1, including all the variables with p-value < 0,1.

Table 2. Results of multiple linear regression analysis under model 1 and model 2

Variables	Model 1			Model 2		
	Coefficients		p value	Coefficients		p value
	Average	Std. Error		Average	Std. Error	
Constant	31.732	78.132	0.686	45.547	38.387	0.240
Population density	-0.001	0.001	0.315	-	-	-
Distance to major centre	0.115	0.081	0.161	-	-	-
Percentage of total surface	0.900	0.895	0.320	-	-	-
Percentage of total population	-0.085	0.761	0.912	-	-	-
Population average age	-3.413	1.482	0.026	-1.852	0.479	0.000
Total dependence index	0.597	0.569	0.300	-	-	-
Age dependence index	0.111	0.753	0.883	-	-	-
Potential sustainability index	-2.036	1.846	0.275	-	-	-
Percentage of population with high school	-1.163	0.568	0.046	-1.013	0.488	0.042
Percentage of population with higher education	0.697	0.511	0.179	-	-	-
Illiteracy rate	-0.630	0.392	0.114	-	-	-
School abandon rate	-2.338	1.312	0.081	-1.976	1.208	0.107
Active population rate	1.406	0.474	0.005	0.670	0.418	0.114
Employed population rate in secondary activities	0.227	0.198	0.258	-	-	-
Employed population rate in tertiary social activities	0.402	0.217	0.070	0.526	0.185	0.006
Employed population rate in tertiary economic activities	0.271	0.188	0.155	-	-	-
Unemployment rate	0.302	0.190	0.118	-	-	-

Source: Multiple linear regression analysis model

The results obtained with model 2 show us that the *Population average age* and *Employed population rate in tertiary social activities* are the variables that better explain the behaviour of the rate of population growth. The value and sign of the variable *Employed population rate in tertiary social activities* show us the relevance of non-tradable local services. These support services, to the population, contributes, produced by the third sector, in general, to increasing employment and improving the quality of life in local communities. In the case of variable School abandon rate either the value or the signal of the coefficient are the expected despite the p-value is not highly robust. In fact, as stated before, the relationship identified as positive relates economic growth as well as the increase in population, with improvement of human capital. Consequently, the dropout decreases the overall level of human capital, as well as the behaviour of economic growth.

After having identified the most relevant variables that explain the population growth rate in the period 2001-2011, we prepare them to obtain the dissimilarity criteria and apply the max-*p*-model. For calculating dissimilarity criteria, the relevant variables were normalized dividing its value in each parish by the average value of the respective county and then the normalized values were summed in a composite index. Through the difference between the indexes of two parishes one obtains its dissimilarity value.

In order to find a more efficient structure of parishes, a max-*p*-model was developed for each one of the counties considered in the sample and two different simulations under a baseline scenario and three alternative scenarios were made. The two simulations are based on two different types of spatially extensive attributes and their scenarios corresponding to different levels of the minimum threshold. In simulation 1 the spatially extensive attribute considered is the population size in each parish and in simulation 2 is surface in Km^2. For both simulations were considered a baseline scenario, which regards the actual situation and three alternative scenarios corresponding to the parameterization of the minimum threshold value define in each county. For that parameterization was considered in the scenarios 1, 2 and 3 40%, 70% and 100% of the respective spatially extensive attribute value, respectively.

Tables 3 and 4 show the max-*p*-model results for the value of the objective function, total value of dissimilarity criteria and the number of parishes according to the respective minimum threshold used under the four scenarios considered in simulations 1 and 2, respectively.

For both simulations the max-p-model results in the baseline scenario represent the actual situation observed in the sample studied, which is an indication that the model could be well calibrated using in the specific empirical context of this study. In this scenario the minimum threshold considered for the spatially extensive attribute is below of the minimum value of either parish, which makes that model solution in this situation is determined only by the trade-offs between the two goals of the objective function.

In simulation 1 the total number of parishes in the baseline scenario is 67. Where we increase the minimum threshold of the population size to a value corresponding to 40% of the county average per parish, the number of parishes diminishes to 40. Thus, if we impose that minimum population of each parish is at least 40% of the actual county average per parish, then we should expect a reduction of 40% on the total number of parishes. In the counties of Reguengos de Monsaraz, Évora and Redondo that reduction could attain 50%, and in Arraiolos, Estremoz and Montemor-o-Novo is higher than 30%. In the counties of Portel and Viana do Alentejo the number of parishes remains the same of baseline scenario.

In scenario 2, for which minimum population size in each parish should be at least 70% the county average value per parish, total number of parishes is 20, representing an average decrease of 70% relatively to the baseline scenario. In the last scenario the minimum population size by parish corresponding to the county average value per parish, which leads that the number of parishes in the sample falls to only 17. In these two scenarios all counties are affected by the diminution on the number of parishes, being the counties of Estremoz, Évora, and Reguengos de Monsaraz, those where reductions are the greatest.

Another interesting result is the evolution pattern of the values of the objective function and dissimilarity criteria as the minimum threshold of population size increases and the number of parishes falls. The value of the objective function in the scenarios 1, 2 and 3 diminishes 34%, 60% and 64% in average, respectively. The greatest reductions occur in the counties of Montemor-o-Novo and Reguengos de Monsaraz and reach to 80% of the baseline scenario value. With respect to the dissimilarity criteria, their values follows the same pattern that the objective function. This could lead to conclude that diminishing the number of parishes one can achieve to more homogenous elementary territory units, as well as well as smaller differences between the various parishes in the same county and hence to a more efficient territorial partition.

. Similarly was made the estimation of this model from the variable surface size (Table 4, simulation 2). The results also show that as we increase the degree of homogeneity in the variable under study – surface size - the number of parishes in each county decreases. These changes are more significant in the counties of Arraiolos, Évora and Viana do Alentejo (in this case, the simulation carried out gives to only one parish).

Table 3. Max-*p*-model results for the spatially attribute of population size (simulation 1)

	Baseline				Scenario 1				Scenario 2				Scenario 3			
Counties	Threshold (Populat)	Objective Function	Dissimil . criteria	Nº of Parishes	Threshold (Populat.)	Objective Function	Dissimil . criteria	Nº of Parishes	Threshold (Populat.)	Objective Function	Dissimil . criteria	Nº of Parishes	Threshold (Populat.)	Objective Function	Dissimil .criteria	Nº of Parishes
Arraiolos	220	-1.06E+06	65.4	7	420	-6.07E+05	65.3	4	736	-4.53E+05	64.9	3	1052	-3.04E+05	62.3	2
Estremoz	32	-3.40E+06	83.9	13	440	-2.15E+06	79.2	8	770	-1.07E+06	72.8	4	1100	-8.06E+05	71.7	3
Évora	320	-8.20E+06	103.5	19	1192	-3.65E+06	53.5	7	2085	-2.18E+06	40.5	4	2979	-2.18E+06	40.5	4
Montemor-o-Novo	500	-2.50E+06	80.8	10	697	-1.70E+06	79.7	7	1220	-7.40E+05	75.7	3	1744	-4.90E+05	74.3	2
Portel	300	-6.40E+05	49.8	8	321	-6.40E+05	49.8	8	562	-1.60E+05	47.4	2	804	-1.60E+05	47.4	2
Redondo	1200	-8,30E+03	0.71	2	1406	-3.80E+03	0.71	1	2461	-3.80E+03	0.71	1	3516	-3.80E+03	0.71	1
Reguengos de Monsaraz	688	-8.70E+04	25.6	5	866	-4.49E+04	23.8	2	1516	-1.70E+04	7.6	1	2166	-1.70E+04	7.6	1
Viana do Alentejo	890	-2.80E+03	7.3	3	766	-2.80E+03	7.3	3	1340	-2.80E+03	7.3	2	1914	-2.80E+03	7.3	2
Total				67				40				20				17

Source: Multi-criteria optimization model results

Table 4. Max-*p*-model results for the spatially attribute of surface size (simulation 2)

	Baseline				Scenario 1				Scenario 2				Scenario 3			
Counties	Threshold (Surface Km^2)	Objective Function	Dissimil. criteria	Nº of Parishes	Threshold (Surface Km^2)	Objective Function	Dissimil criteria	Nº of Parishes	Threshold (Surface Km2)	Objective Function	Dissimil criteria	Nº of Parishes	Threshold (Surface Km2)	Objective Function	Dissimil .criteria	Nº of Parishes
Arraiolos	3700	-1.06E+06	65.5	7	3900	-9.10E+05	65.5	6	6800	-6.07E+05	65.1	4	9700	-3.04E+05	62.3	2
Estremoz	55	-3.49E+06	83.9	13	1500	-3.22E+06	83.9	12	2700	-1.88E+06	77.8	7	3900	-1.61E+06	75.9	6
Évora	20	-8.30E+06	103.5	19	2700	-5.67E+06	87.2	13	4800	-3.92E+06	77.5	9	6800	-3.05E+06	68.8	7
Montemor-o-Novo	5500	-2.47E+06	80.8	10	4900	-2.47E+06	80.8	10	8600	-1.97E+06	80.4	8	12300	-9.90E+05	37.1	4
Portel	3700	-5.66E+05	49.8	8	3000	-6.47E+05	49.8	8	5200	-3.23E+05	48.9	4	7500	-3.23E+05	48.9	4
Redondo	6000	-8.34E+03	0.71	2	7300	-3.81E+03	0.71	1	12900	-3.81E+03	0.71	1	18400	-3.81E+03	0.71	1
Reguengos de Monsaraz	5300	-8.74E+05	25.6	5	3700	-8.74E+05	25.6	5	6400	-6.99E+05	25.6	4	10170	-3.50E+05	14.8	2
Viana do Alentejo	3000	-2.88E+03	7.3	3	5200	-1.90E+03	7.3	2	9100	-1.90E+03	7.3	2	13100	-9.60E+02	3.1	1
Total				67				57				39				27

Source: Multi-criteria optimization model results

When comparing the results obtained from the simulations 1 and 2, we can conclude that the reduction in the number of parishes is higher from the variable population size. This means, on one hand, that initially the disparity between the number of inhabitants in each parish is larger and, on the other hand, the fact that some municipalities have very small populations.

6. Conclusion

This paper discuss a framework for obtaining homogenous territorial clusters based on a Pareto frontier that includes multi-criteria related to the territorial endogenous resources, economic profile and socio-cultural features. This framework is developed in two phases. First, the criteria correlated with the development at the territory unit level are determined through statistical and econometric methods. Then, a multi-criteria approach is developed to allocate each territory unit to an agglomerate of territory according to the Pareto frontier established. The framework is applied to the context of a set of 67 parishes of 8 counties of the Alentejo Central region, southern Portugal.

The results of multiple linear regression analysis show us the most important variables in the explanation of the differences on the development in the area considered. We conclude, as expected, that the more elderly population or dropout rate, the smaller the area's development; the greater active population or the Employed population rate in tertiary social activities, greater development. In the 2nd part of the analysis we started from the initial situation in terms of administrative organization of parishes. The results of *Max-p-model* shows that tests to increase the homogeneity between the parishes, from the variables *population size* and *surface size*, it is possible to reduce the disparity between the parishes, reducing the number of entities. The simulations show that the number of parishes may be lower if the analysis variable is the population size. This result takes into account the wide disparity inhabitants in parishes now existing, as well as the small number of inhabitants who live in most places.

As we have stated before, this work is a first draft to understand how the current economic and social characteristics of the parishes are in accordance with the respective administrative frontiers. The simulations showed how the number of parishes may be reduced taking into account the population and size of territory, based on some variables. In the future we intend to analyse, in particular, the case of some counties, to discuss the simulated maps of parishes. In addition, we intend to compare the results obtained in these simulations with what is defined in the new law of administrative organization of Portuguese parishes.

References

Bergey, P. K., Ragsdale, C.T., & Hoskote, M. (2003a). A decision support system for the electrical power district problem. *Decision Support Systems*, 36: 1-17.

Bergey, P.K., Ragsdale, C. T., & Hostoke (2003b). A simulated annealing genetic algorithm for the electrical power districting problem. *Annals of Operations Research*, 121: 33-55.

Bourjolly, J.M, Laporte, G., & Rousseau, J.M. (1981). Découpage electoral automatisé : application á l'Île de Montréal. *INFOR : Information Systems and Operational Research*, 19 : 113-124.

Bozkaya, B., Erkut, E., Hiaght, D., & Laporte, G. (2003). Designing New Electoral Districts for the City of Edmonton. *Interfaces*, 41(6): 534-547.

Bozkaya, B., Erkut, E., & Laporte, G. (2003). A tabu search heuristic and adaptative memory procedure for political districting. *European Journal of Operational Research*, 144:12-26.

Byfuglien, J., & Nordgärd, A. (1973). Region-Building: A comparison of methods. *Norwegian Journal of Geography*, 27: 127-151.

Cortona, P.G., Manzi, C., Pennisi, A., Ricca, F., & Simeone, B. (1999). *Evaluation and optimization of electoral systems*. SIAM monographs on discrete mathematics and applications. Philadelphia: SIAM.

Deckro, R.F. (1979). Multiple objective districting: a general heuristic approach using multiple criteria. *Operational Research Quarterly*, 28:953-961.

Diniz, F. (2010). *Crescimento e Desenvolvimento Económico: modelos e agentes do processo*, Edições Sílabo, Lisboa.

Duque, J.C., Anselin, L., & Rey, S.J. (2012). The max-*p*-regions problem. *Journal of Regional Science*, 52(3): 397-419.

Easingwood, C. (1973). A heuristic approach to selecting sales regions and territories. Operational Research Quarterly, 24(4):527-534.

Ferland, J.A., & Guénette, G. (1990). Decision support system for the school districting problem. *Operations Research*, 38: 15-21.

Ferligoj, A. and Batagelj, V. (1982). Clustering with relacional constraint. *Psychometrika*, 47(4): 413-426.

Fischer, M.M. (1980). Regional Taxonomy. A comparison of some hierarchic and non-hierarchic strategies. *Regional Science and Urban Economics*, 10: 503-537.

Flischmann, B., & Paraschis, J.N. (1988). Solving a large scale districting problem: a case report. *Computers & Operational Research*, 15(6): 521-533.

Grafinkel, R.S., & and Nemhauser, G.L. (1970). Optimal political districting by implicit enumeration techniques. *Management Science*, 16(8): 495-508.

Gordon, A.D. (1996). A survey of constrained classification. *Computational Statistics & Data Analysis*, 21(1): 17-29.

Hansen, P., Jaumard, B., Meyer, C., Simeone, B., & Doring, V. (2003). Maximum Split Cluster Under Connectivity Constraints. *Journal of Classification*, 20: 143-180.

Hess, S.W., & Samuels, S.A. (1971). Experiences with a sales districting model: criteria and implementation. Management Science, 18(4): 41-54.

Hess. S.W., Siegfeldt, J.B., Whelan, J.N., & Zitlau, P.A. (1965). Nonpartisan political redistricting by computer. Operations Research, 13(6): 998-1006.

Hojati, M. (1996). Optimal political districting. Computers & Operation Research, 23(12): 1147-1161.

INE (Instituto Nacional de Estatística) (2011). *Censos 2011*. Instituto Nacional de Estatística: Portugal.

Lefkovitch, L. (1980). Conditional Clustering. *Biometrics*, 36(1): 45-58.

Legendre, P. (1987). Constrained Clustering. In Legendre, P., & Legendre, L. (eds.). *Developments in Numerical Ecology*. NATO ASI Series, 14. Berlin: Springer-Verlag: 289-307.

Maravalle, M., & Simeone, B. (1995). A Spanning Tree Heuristic for Regional Clustering. *Communications in Statistics-Theory and Methods*, 24(3): 625-639.

Mehrotra, A., Johnson, E.L., & Nemhauser, G.L. (1998). An optimization based heuristic for political districting. *Management Science*, 44(8): 1100-1114.

Mehrotra, A. (1992). *Constrained graph*. PhD thesis, Georgia Institut of Technology.

Muyldermans, L., Cattrysse, D., Oudheusden, D.V. & Lotan, T. (2002). Districting for salt spreading operations. *European Journal of Operational Research*, 139: 521-532.

Murtagh, F. (1992). Contiguit-Constrained Clustering for Image Analysis. *Pattern Recognition Letters*, 13: 677-683.

Musgrave, R. (1974). *Teoria das Finanças Públicas: Um Estudo de Economia Governamental*, Atlas.

Pereira, A. (1995). Regionalism in Portugal, in *The European Union and the regions*, (eds) Barry Jones and Michael keating, Claredon Press Oxford, pp: 269-280

Park, K., Lee, K., Park, S., & Lee, H. (2000). Telecommunications node clustering with the node compatibility and network survivability requirements. *Management Science*, 46(3): 363-374.

Portnov, B.A., & Schwartz, M. (2009). Urban clusters as growth foci. Journal of Regional Science, 49(2): 287-310.

Tavares-Pereira, F., Figueira, J., Mousseau, V., & Roy, B. (2004). *Multiple criteria districting problems, models, algorithms, and applications: the public transportation Paris region pricing system*. Research report nº. 21/2004, INESC – Coimbra, Portugal

Tavares-Pereira, F, Figueira, J., Mousseau, V., & Roy, B. (2007). The public transportation network pricing system of the Paris region. *Annuals of Operations Research*, 154: 69-92.

Shanker, R.J., Turner, R.E., & Zoltners, A.A. (1975). Sales territory design: an integrated approach. *Management Science*, 22(3): 309-320.

Simões Lopes (1984). *Desenvolvimento Regional: Problemática, Teoria e Modelos*, 2ª Edição, Fundação Calouste Gulbenkian, Lisboa

Vickrey, W. (1961). On the preventions of gerrymandering. *Political Science Quarterly*, 76(1): 195-110.

Wise, S.M., Haining, R. P., & Ma. J. (1997). Regionalisation Tools for Exploratory Spatial Analysis of Health Data. In: Fischer M., & Getis, A. (eds.). *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling and Computational Intelligence*. New York: Springer: 83-10

Zoltners, A.A., & Sinha, P. (1983). Sales territory alignment: a review and model. Management Science, 29(3): 1237-1256.