CEFAGE-UE *Working Paper*

2014/02

Clustering of territorial areas: A multi-criteria districting problems

Rui Fragoso[1], Conceição Rego[2]. Vladimir Bushenkov[3]

[1] CEFAGE-UE, and ICAAM-UE, Universidade de Évora

[2] Department of Economics and CEFAGE-UE, Universidade de Évora

[3] CIMA-UE, Universidade de Évora

CLUSTERING OF TERRITORIAL AREAS: A MULTI-CRITERIA DISTRICTING PROBLEM[1]

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

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

JEL Code: C31, R12,

[1] The authors acknowledge the comments and suggestions made by anonymous referees.

CLUSTERING OF TERRITORIAL AREAS: A MULTI-CRITERIA DISTRICTING PROBLEM

1. Introduction

Each unit of territory is defined according to its endogenous resources, economic profile and socio-economic issues, which are criteria associated with their level of development and identifying 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 the homogeneity of their territorial units. According to Fischer (1980), a homogeneous region is a set of spatially contiguous areas showing a high degree of similarity regarding a set of attributes.

In Portugal, policy-makers have discussed and approved a new territorial administrative organization, of parishes and counties[2]. The main purposes of this reform are, inter alia, to promote territorial cohesion and local development, improve and develop local public services and promote economies of scale, efficiency and critical mass in parishes and municipalities.[3] Once this reform is implemented, there will be a reduction in the number of parishes.

This can be considered a multi-criteria districting 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 deal with several kinds of problems related to definition of electoral districts (Bozkaya et al., 2003 and 2009), regions for travelling salespeople teams (Zoltners and Sinha, 1983), areas in metropolitan internet networks to install 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 districting problems are frequent in the real world and involve multiple criteria, which are often incommensurable and conflicting.

The districting problem can be stated as dividing the territory into homogeneous clusters assessed by multiple criteria. The result is a set of homogeneous districts or

[2] Law nº 11-A/2013, 28th June, designated "Administrative Reorganization of Parishes".
[3] Article 2 of Law nº 22/2012, 30h May, designated "Approval of the legal regime of administrative reorganization of municipalities".

areas, composed of elementary units of territory. Each district is associated with a set of technical, economic, ecological, social and other constraints. According to the constraints considered and criteria used in the assessment process, different solutions or maps can be obtained. Thus, "the best solution" will be probably a compromise or a non-dominated solution in which improvement in one criterion leads to a worse result in at least one of the remaining criteria.

The land division problems that first led to using scientific methodologies concerned electoral districting, where the main purpose was to form political constituencies through impartial processes (Mehrotra, 1998). Vickrey (1961) presented one of the first studies on this topic, where the heuristic process used for constructing a zone is described. Hess et al. (1965) were the first to propose a mathematical programming model which states the districting problem as one of location/allocation. However, most studies have focused on salespeople. Generally, the main objective is to balance the workload among different zones (Easingwood, 1973; Hess and Samuels, 1971; Shanker et al., 1975; Zoltners and Sinha, 1983).

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

In this context, this paper aims to propose a simpler and systemic districting framework, which involves multiple criteria and exact algorithms to generate non-dominated 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 Central Alentejo region in southern Portugal, with the purpose of finding the parish structure that leads to the most 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. 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's political and administrative organization is a reflection of history but also population distribution, urban structure and economic and social activities.

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

The Portuguese political administrative system has its origins in the nineteenth century (Pereira, 1995). Currently, the administrative structure maintains features of the Napoleonic model of state organization, including strong centralism. Portuguese administrative and political structure is as follows: a central government and 308 local authorities (municipalities), reflecting the strong local tradition in Portugal (municipalities). At the local level there are two types of government: municipalities and parishes. Both central government and local authorities are directly elected by universal and secret ballot. Central government legislates and enforces policy for the whole country. Local government, of a more executive nature, carries out planning and landscape management functions, promotes economic development and manages territorial infrastructure, among other duties. According to what is proposed by Musgrave (1974), it is the central government's responsibility to ensure income distribution and economic stabilization, while local authorities provide local goods and services.

Thus, central and local government currently exercise public administration through 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 in the territory, in terms of basic administrative units. The amendment, approved by Law No. 22/2012, aims (article 2 of Law No.

22/2012), among other aspects, to promote territorial cohesion and local development, promote the improvement and development of local public services, promote economies of scale, efficiency and critical mass locally and restructure, by joining a significant number of parishes across the country, with a special focus on urban areas. Administrative reorganization of the state, reflected in a reduced number of parishes, should lead (article 8 of Law No. 22/2012), among other aspects, to the county being preferentially considered as a pole of attraction for its parishes, and parishes having a higher rate of economic and social development, a greater number of people and a greater concentration of community facilities.

In essence, this amendment reflects 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 and easier transport 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 divided into pieces, or based on the concept of agglomeration, in which the territory is composed of a set of elementary units (Cortona et al., 1999). Districting problems can involve only one criterion, such as equal voting potential 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). Criteria can be used according to a fixed hierarchy reflecting the decision-maker’s 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 territorial units was chosen. The agglomeration of territorial units into homogeneous districts was studied by many authors who focused on spatial continuity of territory units, ways to measure territorial homogeneity and strategies to explore the solution space efficiently and 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 studies was 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-dominated solutions, in which *n* territorial 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 pre-defined spatial attribute, such as district population, district land area or other district feature.

In this approach, the number of districts is modelled as an endogenous variable, and in contrast to other existing approaches, the spatial contiguity constraint is satisfied without imposing constraints on the shape of districts, such as 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 territorial units (shape of districts). According to Duque et al. (2012), the max-*p*-regions formulation is presented below.

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

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

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

Considering as an 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:

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

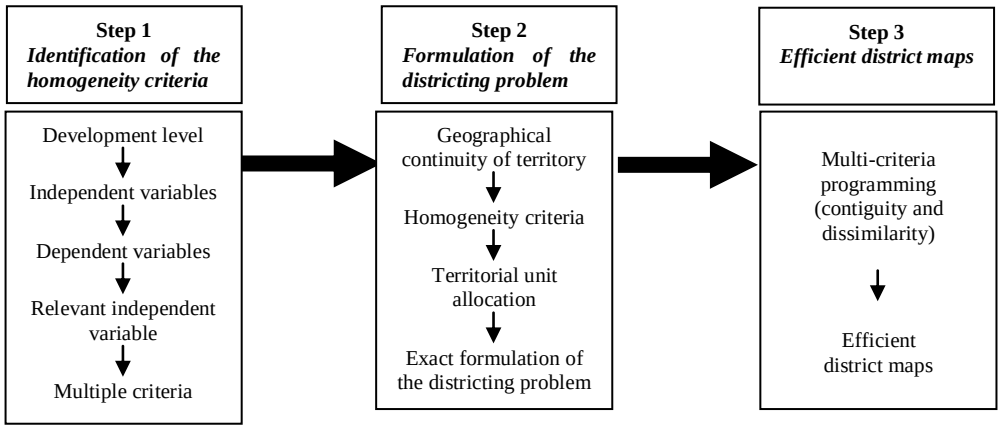
Thus, the max-p-region can be formulated as:

Determine $P_p^* \in \prod$ so 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 to the max-*p*-regions approach of Duque et al. (2012), the approach proposed in this paper considers multiple criteria related to the socio-economic profile of the elementary territorial units to assessing their dissimilarity. Thus, the proposed general framework is formed in three steps as shown in Figure 1.

Figure 1. General framework steps



Step 1 concerns identification of homogeneity criteria to assess dissimilarity between territorial units. This is based on casual relations that can be established between the independent and dependent variables that are most important in defining the level of development and socio-economic profile of territorial units. In step 2, there is exact formulation of the districting problem based on the max-*p*-regions approach. The model developed maximizes the number of districts composed of contiguous territorial units and simultaneously minimizes their dissimilarity. Finally, in step 3, the general framework comprises the simulation of different imposed thresholds of pre-defined spatial 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, whose formulation included the minimization of district population deviation from the average, compactness, socio-economic homogeneity, similarity to the existing plan and maintaining 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. 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, the following factors were considered: local population size (ln); distance from the coast (Km); distance from a major city (Km) and the interaction term between a place's latitude and its height above sea level.

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

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

Casual relationships between dependent and independent variables with high levels of statistical significance allow identification of the variables that best explain the development index and hence those that could also be chosen as attributes of homogeneity criteria to be used later in the max-*p*-regions model to assess the dissimilarity between territorial units.

The analysis started with multiple linear regression analysis between population growth rate (2001-2011) and the set of explanatory variables presented in Table 1. These variables, for 2011, were collected from the Census data, available on the National Office of Statistics (INE) website. Explanatory variables can be grouped in the following four types: territorial variables; population structure; population qualification; and economic indicators, which measure the intensity of economic activity.

Table 1: Variables used in the multiple linear regression analysis

Type of variable	Variables
Independent variable	Population growth rate between 2001 and 2011
Independent variables (2011)	
Territorial variables	Population density (%)
	Distance to major centre (Km)
	Percentage of total area (%)
Population structure	Percentage of total population (%)
	Population's average age (years old)
	Total dependence index (%)
	Age dependence index (%)
	Potential sustainability index (%)
Population qualification	Percentage of population with secondary school education in (%)
	Percentage of population with higher education (%)
	Illiteracy rate (%)
	School drop-out 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: Own elaboration; data from INE (2001, 2011)

As noted earlier, this study intends to identify homogeneous territorial units in terms of development. In this case, we will not make a 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 there is no statistical information available on the variable that typically measures the evolution of income (GDP per capita). The dimensions selected to find the differences in terms of the territory's development are taken from what is suggested in the literature on economic growth and development. For example, according to the concept of human development proposed by UNDP (the United Nations development programme), the indicators used by this

organization are divided into 3 categories / dimensions (Diniz, 2010): long life expectancy and health, knowledge and a decent standard of living. These dimensions are discussed here through the variables relating to Population Structure, Population Qualification and Economic indicators respectively. Furthermore, we integrate the study variables characterizing the territory (territorial variables), which may contribute to differentiating the levels of development of the various territorial units under review.

After checking the hypothesis of linear regression, namely linearity, normality and co-linearity, an analysis of 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 territorial units, the correlation between explanatory variables and the *t* student statistic were calculated, and hence the level of statistical significance of coefficients was evaluated. Thus, the variables with the lowest level of significance were deleted from the model. The regression explicative power was assessed using R square and adjusted R square. This procedure is an interesting advantage of this framework, since it allows us to choose multiple criteria based on the variables most closely related to the socio-economic profile of each territorial unit.

Our study derives from the data collected for the parishes of a group of counties in Central Alentejo, around the municipality of Évora. The units of data collection (parishes) are listed in Table 1-A (in annex).

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

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

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

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

where, x is the vector with n decision variables; X is a 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 optimization, meaning in this case the purpose of

maximizing all criteria simultaneously; *G* is a vectorial function composed of *m* criteria; and *Z* is the feasible region in the space criterion.

A vector $\bar{z} \in Z$ is non-dominated, if, and only if, there is no other vector $z \in Z$, so 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 of elementary territorial 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* territorial units;

$d^y_{i,j}$ is the parameter of dissimilarity relationships between territorial units *i* and *j*, with $i,j \in I$ according to attribute *y*;

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

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

In this formulation, the decision variables follow binary endogenous variables, so that:

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

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

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

This is a mixed integer programming (MIP) model formulated as a multi-criteria program, where the objective function is to (1) maximize the number *p* of potential districts *k* formed by adjacent territorial units *i*, while minimizing the dissimilarity between territorial units *i* and *j*. The two criteria considered in the optimization are not weighted, as is usual. 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 before the goal of minimizing dissimilarity between territorial 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 territorial units according to attributes *y*. The parameter $d^y_{i,j}$ is the difference between the normalized values of attributes *y* in territorial units *i* and *j*. As the dissimilarity goal is a single criterion, the values of parameter $d^y_{i,j}$ have to be aggregated by adding all y attributes into a single value for each pair of *i* and *j* territorial units.

The objective function will improve until a big enough value of *p* is attained, so 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 others with higher dissimilarity. However, the value of objective function and decision variables is subject to constraints (2) to (6).

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

In constraint (5) the value of the spatially extensive attribute is calculated for each district *k* and has to be equal to or greater than a minimum threshold, which is an exogenous parameter. This constraint plays an important role, since the number of

districts created by the model is very sensitive to the value of the pre-defined threshold (*th*).

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

5. **Results**

Identification of the variables explaining the population growth rate (2001-2011) was carried out through the multiple linear regression analysis with a set of explanatory variables (Table 1). Table 2 show that the variables of *Population's average age*, *Percentage of population with secondary school education* and *Active population rate* are the most significant in explaining the dependent variable ($p<0.05$).

Table 2. Results of multiple linear regression analysis (R^2=0,761)

Coefficients	Number of cases	Coefficients		*t-student*
		Average	Std. Error	
Constant	67	31.732	78.132	0.406
Population density	67	-0.001	0.001	-1.015
Distance to major centre	67	0.115	0.081	1.423
Percentage of total area	67	0.900	0.895	1.005
Percentage of total population	67	-0.085	0.761	-0.112
Population's average age	67	-3.413	1.482	-2.304**
Total dependence index	67	0.597	0.569	1.048
Age dependence index	67	0.111	0.753	0.147
Potential sustainability index	67	-2.036	1.846	-1.103
Percentage of population with secondary school education	67	-1.163	0.568	-2.049**
Percentage of population with higher education	67	0.697	0.511	1.364
Illiteracy rate	67	-0.630	0.392	-1.609
School drop-out rate	67	-2.338	1.312	-1.782*
Active population rate	67	1.406	0.474	2.969***
Employed population rate in secondary activities	67	0.227	0.198	1.145
Employed population rate in tertiary social activities	67	0.402	0.217	1.855**
Employed population rate in tertiary economic activities	67	0.271	0.188	1.443
Unemployment rate	67	0.302	0.190	1.591

** Indicates 0.05 significance level; * 0.1 significance level; and *** 0.01 significance level

Source: Multiple linear regression analysis model

These results are what would be expected. The literature on economic growth shows that aged populations have lower levels of growth since individuals with 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 levels of qualification in the population correspond

to greater population growth and higher economic growth. Typically, this relationship is reflected by the ratio of the population with higher education.

In this case, we find that the *Percentage of population with secondary school education* is negatively associated with population growth, i.e., intermediate qualification levels are seen not to be fundamental for population growth. The value and sign of the variable of *Employed population rate in tertiary social activities* show the relevance of non-tradable local services. These services supporting the population, produced by the tertiary sector, generally contribute to increasing employment and improving the quality of life in local communities. In the case of the *School drop-out rate* variable, both the value and the signal of the coefficient are as expected, despite the significant level be only 10%. This estimation has a reasonable explanatory capacity (R^2 = 0,761).

After having identified the most relevant variables explaining the population growth rate in the period 2001-2011, we prepare them to obtain the dissimilarity criteria and apply the max-*p*-model. To calculate the dissimilarity criteria, the relevant variables were normalized dividing their value in each parish by the average value of the respective county and then the normalized values were summed in a composite index. The difference between the indexes of two parishes gives their dissimilarity value.

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

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 in the four scenarios considered in simulations 1 and 2 respectively.

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

In simulation 1, the total number of parishes in the baseline scenario is 67. When we increase the minimum threshold of population size to a value corresponding to 40% of the county average per parish, the number of parishes falls to 40. Thus, if we stipulate that the minimum population of each parish is at least 40% of the present county average per parish, then we can expect a reduction of 40% in the total number of parishes. In the counties of Reguengos de Monsaraz, Évora and Redondo, that reduction could reach 50%, and in Arraiolos, Estremoz and Montemor-o-Novo, it is over 30%. In the counties of Portel and Viana do Alentejo, the number of parishes remains the same as the baseline scenario.

In scenario 2, where the minimum population size in each parish should be at least 70% of the county average per parish, the total number of parishes is 20, representing an average decrease of 70% in relation to the baseline scenario. In the last scenario, the minimum population size by parish corresponds to the county average per parish, which leads to the number of parishes in the sample falling to only 17. In these two scenarios all counties are affected by a reduction in the number of parishes, the counties of Estremoz, Évora, and Reguengos de Monsaraz having the greatest reductions.

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 scenarios 1, 2 and 3 diminishes by 34%, 60% and 64% on average, respectively. The greatest reductions occur in the counties of Montemor-o-Novo and Reguengos de Monsaraz and reach 80% of the baseline scenario value. With respect to the dissimilarity criteria, their

values follow the same pattern as the objective function. This could lead to concluding that a reduction in the number of parishes can achieve more homogenous elementary territorial units, as well as smaller differences between the various parishes in the same county and hence a more efficient territorial partition.

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

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

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

Counties	Baseline				Scenario 1				Scenario 2				Scenario 3			
	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 spatial attribute of area (simulation 2)

Counties	Baseline				Scenario 1				Scenario 2				Scenario 3			
	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

6. Conclusion

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

The results of multiple linear regression analysis show us the most important variables in explaining the differences in development in the area considered. We conclude, as expected, that the more elderly the population or the higher the school drop-out rate, the lower the area's development. On the other hand, the greater the active population or the rate of employment in tertiary social activities, the greater is the development. In the 2nd part of the analysis, we started from the current situation in terms of administrative organization of parishes. The results of the *Max-p-model* show that tests to increase the homogeneity between parishes, using the variables of *population size* and *area*, it is possible to reduce the disparity between parishes, reducing the number of units. The simulations show that the number of parishes may be lower if the variable of analysis is population size. This result takes into account the wide disparity of the population in current parishes, as well as the small number of inhabitants in most places.

As stated before, this work is a first draft to understand how the current economic and social characteristics of parishes are in accordance with respective administrative frontiers. The simulations showed how the number of parishes can be reduced taking into account the population and size of the territory, based on some variables. In the future we intend to analyse some counties in particular, 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 relational 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). Contiguity-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.

Annex

Table 1-A: Counties and Parishes of case study

Counties	Parishes	Parishes 2013 (according to Law 11-A/2013)
Arraiolos	Arraiolos Igrejinha Santa Justa S. Gregório Gafanhoeira (S. Pedro) Vimieiro Sabugueiro	Arraiolos Igrejinha Vimieiro Parish Union of Gafanhoeira (S. Pedro) and Sabugueiro Parish Union of Santa Justa and S. Gregório
Estremoz	Arcos Glória Estremoz (Santa Maria) Evoramonte (Santa Maria) Santa Vitória do Ameixial Estremoz (Santo André) Santo Estevão S. Bento do Ameixial S. Bento de Ana Loura S. Bento do Cortiço S. Domingos de Ana Loura S. Lourenço de Mamporcão Veiros	Arcos Glória Evoramonte (Santa Maria) Parish Union of Estremoz (Santa Maria and Santo André) Parish Union of S. Bento do Cortiço e Santo Estevão Parish Union of S. Lourenço de Mamporcão and S. Bento de Ana Loura Parish Union of do Ameixial (Santa Vitória and São Lourenço) S. Domingos de Ana Loura Veiros
Évora	Nª Srª da Boa Fé Nª Srª da Graça Divor Nª Srª Machede Nª Srª da Tourega Évora (Santo Antão) S. Bento do Mato Évora (S. Mamede) S. Manços S. Miguel de Machede S. Vicente do Pigeiro Torre de Coelheiros S. Sebastião da Giesteira Canaviais Nª srª de Guadalupe Bacelo Horta das Figueiras Malagueira Sé e S. Pedro Srª da Saúde Horta das Figueiras Malagueira Sé e S. Pedro Srª da Saúde	Parish Union of Bacelo and Srª da Saúde Parish Union of Évora (S. Mamede, Sé S. Pedro and Santo Antão) Nª Srª da Graça Divor Nª Srª Machede S. Bento do Mato S. Miguel de Machede Torre de Coelheiros Parish Union of Malagueira and Horta das Figueiras Parish Union of Tourega and Guadalupe Parish Union of S. Manços and S. Vicente do Pigeiro Parish Union of S. Sebastião and Boa Fé Canaviais
Montemor o Novo	Cabrela Lavre Nª Srª do Bispo Nª Srª da Vila Santiago do Escoural S. Cristovão Ciborro Cortiçadas de Lavre Silveiras Foros de Vale Figueira	Cabrela Santiago do Escoural S. Cristovão Ciborro Foros de Vale Figueira Parish Union of Cortiçadas de Lavre and Lavre Parish Union of Nª Srª da Vila, Nª Srª do Bispo and Silveiras

Portel	Alqueva Amieira Monte Trigo Oriola Portel Santana S. Bartolomeu do Outeiro Vera Cruz	Parish Union of Amieira and Alqueva Parish Union of Oriola and S. Bartolomeu do Outeiro Monte Trigo Portel Santana Vera Cruz
Redondo	Montoito Redondo	Montoito Redondo
Reguengos de Monsaraz	Reguengos de Monsaraz Campo Corval Monsaraz Campinho	Reguengos de Monsaraz Corval Monsaraz Parish Union of Campo and Campinho
Viana do Alentejo	Alcáçovas Viana do Alentejo Aguiar	Alcáçovas Viana do Alentejo Aguiar