J. Quant. Econ. (2016) 14:179–198
DOI 10.1007/s40953-016-0030-y

ORIGINAL ARTICLE

Clustering of Territorial Areas: A Multi-Criteria Districting Problem

Rui Fragoso[1] · **Conceição Rego**[2] · **Vladimir Bushenkov**[3]

Published online: 12 February 2016

Abstract This paper aims to propose a framework for obtaining homogenous territorial clusters based on a max-*p*-regions optimisation problem, considering multiple criteria related to endogenous resources, economic profile and socio-cultural features of territories. This framework is developed in three steps. First, the dissimilarity criteria correlated with development at the territorial unit level are identified, using a multiple linear regression analysis. Then, a multi-criteria max-*p*-regions model is developed, in order to allocate each territorial unit (parishes) to a territorial agglomerate. Finally, the max-p-model is used to generate alternative efficient district maps according to the changes in the threshold of spatial attributes.

The authors acknowledge the comments and suggestions made by anonymous referees. The authors are pleased to acknowledge financial support from Fundação para a Ciência e a Tecnologia and (grant UID/ECO/04007/2013) and FEDER/COMPETE (POCI-01-0145-FEDER-007659).

✉ Rui Fragoso
rfragoso@uevora.pt

Conceição Rego
mcpr@uevora.pt

Vladimir Bushenkov
bushen@uevora.pt

1 Departamento de Gestão, Escola de Ciências Sociais, Centro de Estudos e Formação Avançada em Gestão e Economia e Instituto de Ciências Agrárias e Ambientais Mediterrânicas, Universidade de Évora, Largo dos Colegiais 2, 7000 Évora, Portugal

2 Departamento de Economia, Escola de Ciências Sociais, Centro de Estudos e Formação Avançada em Gestão e Economia, Universidade de Évora, Largo dos Colegiais 2, 7000 Évora, Portugal

3 Departamento de Matemática, Escola de Ciências e Tecnologia, Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Rua Romão Ramalho 59, 7000-671 Évora, Portugal

Keywords Cluster · Districting · Multi-criteria · Optimisation

JEL Classification C31 · R12

Introduction

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 2011), 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 et al. 2003a). According to Tavares-Pereira et al. (2007), 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 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. However, "the best solution" will probably be 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 et al. 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).

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). They 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 and 2011; 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 et al. 2003b; Muyldermans et al. 2002). Exact algorithms allow finding the optimal solution to an optimization problem and non-exact algorithms are generally based on heuristics processes which sometimes produce worse solutions.

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 space solution 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 to establish the number of regions that should be created.

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. 2012). 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 optimal number of districts in merging or partition processes. The comparison between elementary units of territory is another obvious problem in the literature (Cortona et al. 1999), since the agglomeration or partition process can be motivated by different objectives.

In order to address those issues, this paper aims to propose a holistic and systemic districting framework, which involves simultaneously multiple criteria for comparing elementary territorial units and an optimisation approach in which the number of new districts created is an endogenous variable. Thus, 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 variables used to compare different territorial units.

The paper also aims to show an application of the proposed framework. In this case a set of parishes in the Central Alentejo region (NUTS III level) in southern Portugal was chosen, since recently, policy-makers have discussed and approved a new territorial administrative organization of parishes and counties based on the aggregation of existing parishes.

The remainder of the paper is organized as follows. The proposed general framework and the context of application are presented in the next section. The attributes of multiple criteria are established in Sect. 3. A multi-criteria programming model based on a max-p-region formulation is presented in Sect. 4. Results are presented and discussed in Sect. 5, considering separately the determination of dissimilarity criteria and building efficient district maps. Finally, in Sect. 6, the main conclusions and suggestions are provided.

General Framework and Context of Application

In addition to endogenous determination of the optimal number of agglomeration p regions (districts), the general framework proposed in this paper considers multiple criteria related to the relevant variables that can be used to assess the dissimilarity between elementary territorial units. Thus, the proposed general framework is formed in three steps as shown in Fig. 1 and Table 1 illustrates how it can improve the existing approaches.

Step 1 concerns identification of homogeneity criteria to assess dissimilarity between territorial units. This is based on causal relations that can be established between the independent and dependent variables that are most important in defining

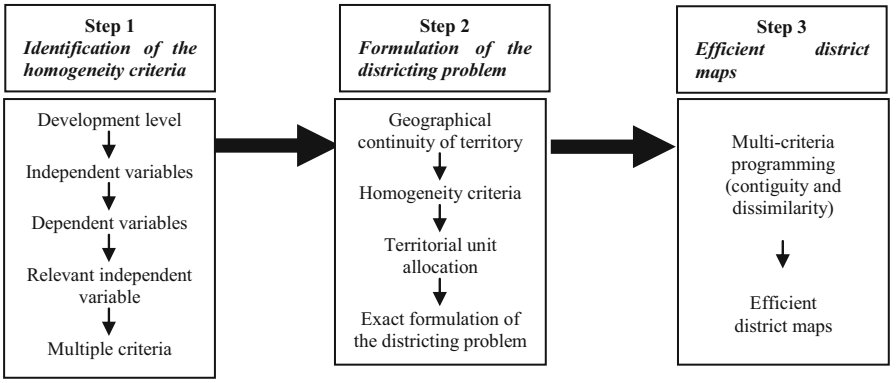

Fig. 1 General framework steps. Source: Own elaboration

Table 1 The positioning of the proposed framework compared to existing approaches

Authors	Contribution
Vickrey (1961)	Heuristic processes were used to construct a zone
Hess et al. (1965)	Proposed a mathematical programming model to make location/allocation of territorial units in a districting problem
Deckro (1979); Bourjolly et al. (1981); Bergey et al. (2003a); Bozkaya et al. (2003 and 2011)	Used multi-criteria approaches to construct zones
Ferligoj and Batagelj (1982); Gordon 1996; Wise et al. (1997) ;Hansen et al. (2003); Duque et al. (2012)	Focused on spatial continuity of territorial units and ways to measure their homogeneity using exact and non-exact algorithms
The proposed framework	Focused on spatial continuity of territorial units, uses a multiple criteria index to assess homogeneity and an exact algorithm to explore an efficient spatial solution

Source: Own elaboration

the level of development and socio-economic profile of territorial units. However, the choice of these variables depends on data availability.

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.

According to Duque et al. (2012), exact formulation of the districting problem in the scope of the max-p-regions approach can be stated as follows.

Let $A = \{A_1, A_2, \ldots, A_n\}$ be a set of elementary territorial units which can be described by the attributes $y \in Y = \{1, 2, \ldots, m\}$ 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 $d_{ij} \equiv d(A_i, A_j), d_{ij} \geq 0, d_{ij} = d_{ji}$ and $d_{ii} = 0$ for $i, j = 1, 2, \ldots, n$. Let $W = (V, E)$ be the connected contiguity graph associated with A, so that the vertices $v_i \in V$ correspond to territorial units $A_i \in A$ and edges $\{v_i, v_j\} \in E$ if, and only if, territorial units A_i and A_j share a common border. The partition of territorial units $A = \{A_1, A_2, \ldots, A_n\}$ into p districts R_k, $k = 1, 2, \ldots, p, 1 \leq p \leq n$, can be represented by $P_p = \{R_1, R_2, \ldots, R_p\}$ so that:

$$
\begin{aligned}
&|R_k| > 0, \quad R_k \cap R_{k'} = \emptyset \quad \text{for} \quad k, k' = 1, 2, \ldots, p \wedge k \neq k';\\
&\bigcup_{k=1}^{p} R_k = A; \sum\nolimits_{A_i \in R_k} l_i \geq Thr \quad \text{for} \quad k = 1, 2, \ldots, p,\\
&0 \leq Thr \leq \sum\nolimits_{A_i \in A} l_i,
\end{aligned}
$$

where *Thr* is the given threshold. Let Π denote the set of all feasible partitions of A. As evaluation criteria for a feasible partition $P_p \in I$ we will use the heterogeneity $h(R_k)$ of district k with $R_k \in P_p$ and the total heterogeneity $H(P_p)$of partition $P_p \in I$ calculated as:

$$
h(R_k) = \sum\nolimits_{ij: A_i, A_j \in R_k, i \leq j} d_{ij}, \quad \text{and} \quad H(P_p) = \sum\nolimits_{k=1}^{p} h(R_k).
$$

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

$$
\begin{aligned}
&\text{Determine } P_p^* \in \Pi \quad \text{so that} \quad \left|P_p^*\right| = \max\left(\left|P_p\right| : P_p \in \Pi\right), \quad \text{and}\\
&\nexists P_p \in \Pi : \left|P_p\right| = \left|P_p^*\right| \wedge H(P_p) < H\left(P_p^*\right).
\end{aligned}
$$

As stated before, the context of application of this framework is a set of parishes in the Central Alentejo region (NUTS III level) in southern Portugal, for which a new law of territorial administrative organization was recently approved based on the aggregation of existing parishes.

In Europe, and particularly in the south, regions present great diversity. In fact, in several European countries, regions with high levels of agglomeration of population and economic activity coexist with other less favoured ones, where economic activities and population are scarce. Portugal is a country showing great diversity of territorial occupation. Inland, low density areas predominate. On the coast, there are two metropolitan areas (Lisbon and Porto), where levels of population density are higher.

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 state centralism also reflected in a vertical hierarchical model of territorial governance. Thus, Portugal's administrative territory is organised in 308 counties, local authorities (municipalities) depending directly on Central Government. However, these counties are divided into parishes which are smaller elementary territorial units.

Recently, policy-makers have discussed and approved a new territorial administrative organization of parishes and counties[1] based on the aggregation of existing parishes. One of the main purposes of this reform is to promote efficiency and critical mass in counties based on aggregation and geographical proximity of parishes. Once this reform is implemented, there will be a reduction in the number of parishes from 4,400 to 3,091.

The Attributes of Multiple Criteria

Edmonton's municipal electoral districts in Canada are defined based on a set of socio-economic criteria which include 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, a set of variables that can have an effect on the local population growth rate was considered. Thus, causal relationships considering population growth rate between 2001 and 2011 as the dependent variable and several socioeconomic variables as independent variables were established through multiple linear regression analysis, using the least square model. Table 2 presents the variables used in this regression analysis.

Population growth rate was chosen as the dependent variable, since Portnov and Schwartz (2009) consider that spatial clustering features can help to explain population growth in urban areas.

To calculate this variable, data from the 2001 and 2011 Population Census were used and the choice of socioeconomic independent variables also took into account the available data at the parish level from the Population Census of 2011. The period between 2001 and 2011 was used to calculate the variable of population growth rate because this corresponds to the last two population censuses in Portugal and hence data can better show the current territorial dynamics.

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

Table 2 Variables used in the multiple linear regression analysis

Type of variable	Variables
Dependent variable	Population growth rate between 2001 and 2011
Independent variables (2011)	
Territorial variables	Population density (1000 residents per Km^2)
	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 (%)
	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)

A Population Census is carried out nationally every ten years by the Portuguese Agency of Statistics and these are the only official statistics available at the geographically disaggregated level of parishes.

Our study derives from the data collected in the 2001 and 2011 population censuses for the parishes of a group of counties in Central Alentejo, around the municipality of Évora in southern Portugal. The units of data collection (parishes) are listed in an Annex.

Causal relationships between dependent and independent variables with high levels of statistical significance allow identification of the variables that best explain the population growth rate 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 Multi-Criteria Max-*p*-Regions Model

After having identified the relevant attributes that can be used as multiple criteria to assess the dissimilarity between elementary territorial units, we will formulate the

districting problem inspired in the max-p-regions model of Duque et al. (2012). This is a multi-criteria model that generates non-dominated solutions in which elementary territorial units are aggregated into the maximum number of districts. Each new district created satisfies an imposed minimum threshold value. This threshold is an exogenous spatial attribute that is pre-defined, such as district population, district land area or other spatial district feature.

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). In contrast to other existing approaches, spatial contiguity is satisfied without imposing constraints on the shape of districts, such as maximum compactness.

In order to write the model the following notation is used:

- Index sets:
  i- elementary territorial units, $i \in I = \{1, \ldots, n\}$;
  k - potential agglomeration districts, $k = \{1, \ldots, n\}$;
  c- contiguity order, $c = \{1, \ldots, q\}$, with $q = (n-1)$;
  y -attributes that describe territorial units i, $y = \{1, \ldots, m\}$;
  N_i- set of territorial units j that share a border or are adjacent to territorial units i, with $i, j \in I$ and $i \neq j$.
- Parameters:
  l_i- spatially extensive attribute value of territorial unit i;
  Thr- minimum value of attribute l at the districting scale;
  $d_{i,j}^{y}$- dissimilarity of territorial units i and j, $i, j \in I$ according to attribute y;
  h– scaling factor, with $h = 1 + \log\lfloor(\sum_i \sum_{(j|i<j)} d_{i,j}^{y})\rfloor$.
- Decision variables:
$$x_i^{k,c} = \begin{cases} 1, \text{if territorial unit } i \text{ is allocated to district } k \text{ in order } c \\ 0, \text{otherwise} \end{cases}$$
$$t_{i,j} = \begin{cases} 1, \text{if territorial units } i,\ j \text{ belong to the same district } k \\ 0, \text{otherwise} \end{cases}$$
- Objective functions:

$$f_1(x) = \sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k,0}.10^h$$

$$f_2(t) = \sum_{i} \sum_{j|i<j} \sum_{y} d_{i,j}^{y}.t_{i,j}$$

Thus, formulation of the max-p-regions model can be written as follows:

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

Subject to

$$\sum_{i=1}^{n} x_i^{k,0} \leq 1 \quad \forall k = 1, \ldots n \tag{2}$$

$$\sum_{k=1}^{n} \sum_{c=0}^{q} x_i^{k,c} = 1 \quad \forall i = 1, \ldots n \tag{3}$$

$$x_i^{k,c} \leq \sum_{j \in N_i} x_j^{k,(c-1)} \quad \forall k = 1, \ldots n; \quad \forall i = 1, \ldots n; \quad \forall c = 1, \ldots q \tag{4}$$

$$\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{k,c} l_i \geq Thr \sum_{i=1}^{n} x_i^{k,0} \quad \forall k = 1, \ldots n \tag{5}$$

$$t_{i,j} \geq \sum_{c=0}^{q} x_i^{k,c} + \sum_{c=0}^{q} x_j^{k,c} - 1 \quad \forall i, j = 1, \ldots n, \quad \text{with} \quad i < j; \quad \forall k = 1, \ldots n \tag{6}$$

This is a mixed integer programming (MIP) model formulated as a multi-criteria program, where decision variables ($x_i^{k,c}$ and $t_{i,j}$) are treated as binary variables, $\forall i, j = 1, \ldots n, \quad \forall c = 0, \ldots q$. The model maximizes the number p of potential districts k formed by adjacent territorial units i, while minimizing the dissimilarity between territorial units i and j.

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 $(k, 0)$.

The objective function is a minimization function given by the sum of objectives $f_1(x)$ and $f_2(t)$, which instead of being weighted are merged in a single value.

The objective $f_1(x)$ controls the number of p regions created and is obtained by adding the number of elementary territorial units, that is, "root" territorial units $x_i^{k,0}$. In order to consider a hierarchy where the number of p regions comes before the dissimilarity goal, the first term of objective $f_1(x)$is multiplied by scaling factor h.

The dissimilarity goal depends on the binary variable value $t_{i,j}$ and the parameter $d_{i,j}^{y}$ of the dissimilarity relationships between territorial units according to attributes y. The parameter $d_{i,j}^{y}$ 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 $d_{i,j}^{y}$ 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 the objective function and decision variables is subject to constraints (2)–(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 territorial unit i should correspond exactly to one district k and 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 p created by the model is very sensitive to the value of the pre-defined threshold (Thr).

Finally, constraint (6) determines the pair of adjacent territorial units i and j that should be considered for calculating the total dissimilarity ($f_2(t)$).

Results

The results are presented in two phases. The first regards determination of the dissimilarity criteria and the second the building of efficient district maps obtained from the multi-criteria max-p-regions model.

In the former, basic statistics of the data set used are presented. Then, the results of the multiple linear regression established between population growth rate and the set of independent variables presented in Table 2 are analysed and discussed. The objective is to find the most relevant multiple criteria to be used for assessing the dissimilarity between territorial units.

In the latter, several solutions of the max-p-regions model are explored based on two types of simulation which consider distinct spatial thresholds and different levels of these thresholds.

Determination of Dissimilarity Criteria

Table 3 summarizes the averages, standard deviation and minimum and maximum values of the data set used. The average value of population growth rate between 2001 and 2011 is negative (−9.54 %). The standard deviation is 12.23 % and the minimum and maximum values are −33.93 % and 27.32 %. These basic statistics show that these variable values are widely scattered. Among the explanatory variables, only population density and percentage of total population present a more scattered set of values.

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's 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.

Table 4 presents the results of multiple linear regression analysis regarding the average value of model coefficients, the respective standard deviations and the t-student statistics.

The variables of *Population's average age*, *Percentage of population with secondary school education, Active population rate* and *Employed population in tertiary social activities* are the most significant in explaining the dependent variable of population growth rate between 2001 and 2011 ($p < 0.05$).

These results are what would be expected. The literature on economic growth shows that aging 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

Table 3 Basic statistics ($N = 67$)

Variables	Average	Standard deviation	Minimum	Maximum
Population growth rate between 2001 and 2011 (%)	−9.54	12.23	−33.93	27.32
Population density (1000 residents per Km^2)	328.71	1196.16	1.20	7420.90
Distance to major centre (Km)	35.44	17.96	0.00	61.00
Percentage of total area (%)	1.49	1.20	0.00	5.53
Percentage of total population (%)	1.49	2.07	0.03	9.84
Population's average age (years old)	47.32	3.73	38.66	55.02
Total dependence index (%)	69.43	12.73	39.10	100.00
Age dependence index (%)	50.12	15.01	17.10	83.80
Potential sustainability index (%)	2.18	0.92	0.00	5.80
Per. of pop. with secondary school education (%)	11.21	3.20	5.93	18.99
Percentage of population with higher education (%)	6.85	4.86	1.22	20.94
Illiteracy rate (%)	12.61	4.68	2.85	26.02
School drop-out rate (%)	0.63	1.03	0.00	4.46
Active population rate (%)	44.09	4.22	35.14	53.89
Employed population rate (%)	89.02	5.44	62.69	96.15
Employed population rate in primary activities (%)	17.58	10.30	2.05	46.67
Employed pop. rate in secondary activities (%)	21.76	6.24	6.67	38.01
Employed pop. rate in tertiary social activities (%)	29.66	8.42	11.90	48.17
Empl. pop. rate in tertiary economic activities (%)	31.01	6.14	17.45	45.00
Unemployment rate (%)	10.99	5.44	3.85	37.31

Source: Own elaboration; data from INE (2001, 2011)

population growth and higher economic growth. Typically, this relationship is reflected in 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,

Table 4 Results of multiple linear regression analysis ($R^2 = 0, 761$)

Coefficients	Number of cases	Average	Std. error	*t-student*
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

both the value and the signal of the coefficient are as expected, despite the significance level being 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 between 2001 and 2011, they were prepared to obtain the dissimilarity criteria and apply the max-p-regions model. To calculate the dissimilarity criteria, the four relevant variables (*Population's average age*, *Percentage of population with secondary school education, Active population rate* and *Employed population in tertiary social activities)* 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.

Efficient District Maps

In order to find an efficient structure of parishes, a max-p-regions model was developed for each of the counties considered in the sample and, in addition to the baseline

scenario, two different simulations were made. 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 new parish (district) 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 parameterisation of the minimum threshold value defined in each county.

The max-p-regions model comprises the aggregation of small areas into a maximum number p of homogeneous regions, so that the value of a spatially extensive attribute is greater than or equal to a minimum threshold. The spatially extensive attributes of population size (simulation 1) and area (simulation 2) are often taken into account in regional policy planning and they are also mentioned as examples by Duque et al. (2012).

Tables 5 and 6 show the max-p-regions 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 representing 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 any 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 the minimum threshold of population size increases to a value corresponding to 40 % of the county average per parish, the number of parishes falls to 55. 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 18 % in the total number of parishes. In the counties of Montemor-o-Novo, Arraiolos, Reguengos de Monsaraz and Redondo, that reduction could reach 20, 29, 40 and 50 %, respectively. 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 49, representing an average decrease of 27 % 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 37. This is 45 % of the number of parishes in the baseline scenario. In these two scenarios all counties are affected by a reduction in the number of parishes, the counties of Montemor-o-Novo, Reguengos de Monsaraz and Portel having the greatest reductions.

In the case of the county of Évora, it was not possible to consider as the minimum threshold 40, 70 and 100 % of average population size due to a wide discrepancy between urban and rural parishes. In order to overcome this problem, a minimum threshold value of 500, 700 and 900 people was used in scenarios 1, 2 and 3, respectively. The results also show a reduction in the number of parishes from 19 in the baseline scenario to 16 in scenarios 1 and 2 and to 14 in scenario 3.

Table 5 Max-*p*-regions model results for the spatial attribute of population size (simulation 1)

Counties	Baseline				Scenario 1			
	Threshold (populat.)	Objective function	Dissimil. criteria	No. of parishes	Threshold (populat.)	Objective function	Dissimil. criteria	No. of parishes
Arraiolos	220	−1.73E+06	0.0	7	420	−1.24E+06	8.06	5
Estremoz	32	−3.40E+06	0.0	13	100	−2.79E+06	0.934	11
Évora	300	−1.27E+07	0.0	19	500	−9.49E+06	0.658	16
Montemor-o-Novo	500	−2.46E+06	0.0	10	697	−1.97E+06	0.813	8
Portel	300	−9.25E+05	0.0	8	321	−9.25E+05	0.0	8
Redondo	1200	−9.00E+03	0.0	2	1406	−4.16E+03	0.354	1
Reguengos de Monsaraz	688	−1.18E+05	0.0	5	866	−7.10E+04	3.611	3
Viana do Alentejo	890	−2.88E+03	0.0	3	766	−2.88E+03	0.0	3
Total	–	−2.13E+07	0.0	67	–	−1.65E+07	14.430	55
Arraiolos	736	−9.90E+05	0.62	4	1052	−7.42E+05	16.94	3
Estremoz	200	−2.78E+06	0.419	11	400	−2.53E+05	1.128	10
Évora	700	−9.48E+06	6.903	16	900	−8.03E+06	7.061	14
Montemor-o-Novo	1220	−1.97E+6	0.813	8	1744	−1.97E+06	0.813	2
Portel	562	−5.70E+05	1.075	5	804	−4.63E+05	4.93	3
Redondo	2461	−4.16E+03	0.354	1	3516	−4.16E+03	0.354	1
Reguengos de Monsaraz	1516	−4.73E+04	10.668	2	2166	−4.73E+04	10.668	2
Viana do Alentejo	1340	−1.92E+03	2.062	2	1914	−1.92E+03	2.062	2
Total	–	−1.58E+07	22.914	49		−1.15E+07	43.956	37

Source: Multi-criteria optimization model results

Table 6 Max-*p*-regions model results for the spatial attribute of area (simulation 2)

Counties	Baseline				Scenario 1			
	Threshold (surface Km^2)	Objective function	Dissimil. criteria	No. of parishes	Threshold (surface Km^2)	Objective function	Dissimil criteria	No. of parishes
Arraiolos	3700	−1.73E+06	0.0	7	3900	−1.49E+06	7.898	6
Estremoz	55	−3.29E+06	0.0	13	1500	−3.04E+06	0.560	12
Évora	20	−1.13E+07	0.0	19	500	−9.49E+06	2.409	16
Montemor-o-Novo	5500	−2.47E+06	0.0	10	4900	−2.47E+06	0.0	10
Portel	3000	−9.25E+05	0.0	8	3700	−8.09E+05	0.309	7
Redondo	6000	−9.03E+03	0.0	2	7300	−4.16E+03	0.354	1
Reguengos de Monsaraz	5300	−1.18E+05	0.0	5	3700	−1.18E+05	0.0	5
Viana do Alentejo	3000	−2.88E+03	0.0	3	5200	−1.92E+03	2.062	2
Total	–	−1.98E+07	0.0	67	–	−1.74E+07	13.592	59
Arraiolos	6800	−1.24E+06	0.325	5	9700	−7.45E+05	16.938	3
Estremoz	2700	−2.27E+06	6.351	9	3900	−2.27E+06	6.490	9
Évora	1000	−9.49E+06	2.409	16	1500	−8.90E+06	3.158	15
Montemor-o-Novo	8600	−2.22E+06	0.199	8	12300	−1.73E+06	7.538	7
Portel	5200	−6.94E+05	0.764	6	7000	−6.94E+05	0.764	6
Redondo	12900	−4.16E+03	0.354	1	18400	−4.16E+03	0.354	1
Reguengos de Monsaraz	6400	−9.47E+04	0.457	4	10170	−7.10E+04	0.789	3
Viana do Alentejo	9100	−1.92E+03	2.062	2	13100	−1.92E+03	2.062	2
Total	–	−1.60E+07	12,921	51	–	−1.44E+07	38,093	46

Source: Multi-criteria optimization model results

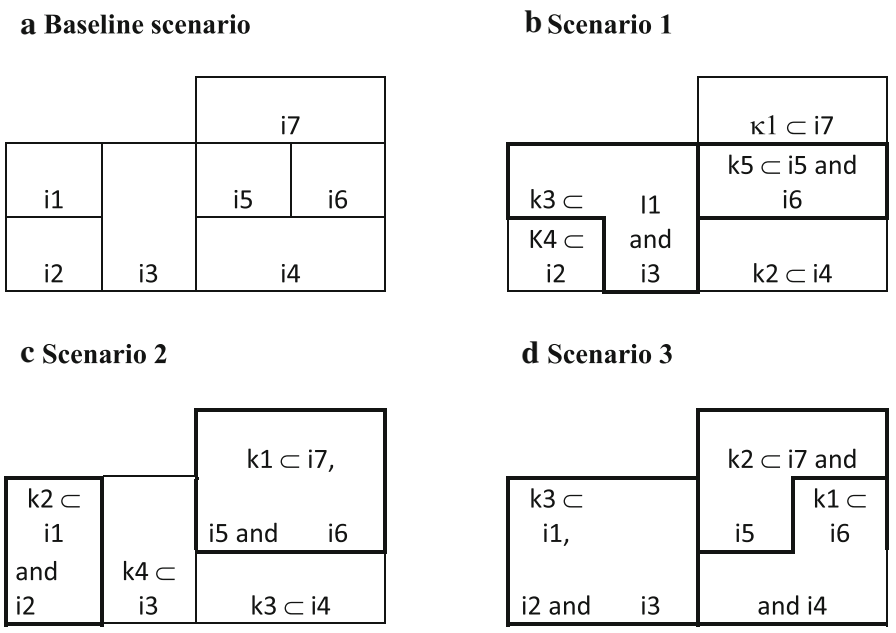

Fig. 2 District map of Arraiolos in scenarios of simulation 1. **a** Baseline scenario. **b** Scenario 1. **c** Scenario 2. **d** Scenario 3

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 drops.

The value of the objective function in scenarios 1, 2 and 3 diminishes by 23, 26 and 46 % on average, respectively. The greatest reductions occur in the counties of Redondo and Reguengos de Monsaraz and reach more than 50 % of the baseline scenario value. With respect to the dissimilarity criteria, their values are zero in the baseline scenario and increase as the number of parishes per county diminishes, the total value being 14.430 in scenario 1, 22.914 in scenario 2 and 43.956 in scenario 3.

Estimation of this model was similarly performed using the variable of area as spatially extensive attribute (Table 6, 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 most significant in the counties of Arraiolos and Redondo.

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 than the disparity between areas, and secondly, that some parishes have very small populations.

Figure 2 presents the specific example of the efficient district map obtained for the county of Arraiolos, considering population size as the pre-defined spatially extensive attribute.

In this figure the initial parishes are represented by i and the new districts by k. The bold lines represent the composition of the new parishes resulting from the imposition of a minimum threshold value for the spatially extensive attribute. The parishes are represented by squares that are aggregated according to their adjacent borders.

In the baseline scenario, the county of Arraiolos comprises 7 parishes (see Annex). The minimum threshold value imposed for population size of new parishes (districts) is 420 people in scenario 1, 736 people in scenario 2 and 1052 people in scenario 3. In scenario 1, the number of parishes in the county of Arraiolos is reduced from 7 to 5 and the parishes of Sabugueiro (i1) and Arraiolos (i3), as well as the parishes of São Gregório (i5) and Santa Justa (i6) are merged in two new parishes, k3 and k5.

In the scenarios 2 and 3 the number of parishes falls to 4 and 3. The former scenario comprises two new parishes—the districts k1 and k2—resulting from aggregation of the parishes of São Gregório (i5) and Santa Justa (i6) and of the parishes of Sabugueiro (i1) and Gafanhoeira (i2), respectively. As was stated before, the latter scenario considers as the minimum threshold of population size 1052 people. In these conditions 3 new parishes (districts) emerge with the following composition: k1 including the parishes of Igreginha (i4) and Santa Justa (i6); k2 including the parishes of São Gregório (i5) and Vimieiro (i7); and k3 including the parishes of Sabugueiro (i1), Gafanhoeira (i2) and Arraiolos (i3).

Conclusion

This paper proposes a framework for obtaining homogenous territorial clusters based on a max-p-regions optimisation model that includes multi-criteria related to endogenous territorial resources, economic profile and socio-cultural features. This framework is developed in three steps. First, the criteria correlated with population growth rate at the territorial unit level are determined through a multiple linear regression analysis. Then, a multi-criteria max-p-regions model is developed, in order to allocate each territorial unit to an agglomerate of territory. Finally, efficient district maps are drawn for different simulations of spatial attributes and their thresholds.

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 the most important variables in explaining the differences in population growth rate 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 population growth rate. On the other hand, the greater the active population or the rate of employment in tertiary social activities, the greater the population growth rate.

The second part of the analysis started by applying the max-p-regions model to the current situation in terms of administrative organization of parishes. Then several simulations were made considering as spatially extensive attribute population size and area per parish. The model's results are shown to be coherent with the current administrative situation. As expected, increases in the spatially extensive attribute threshold result in a lower number of parishes. The simulations show also that the number of parishes may be lower if the spatially extensive attribute is pop-

ulation size instead of area. 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.

The framework proposed in this paper, involving simultaneously multiple criteria attributes to assess dissimilarity between territorial units and an optimisation approach based on the max-p-regions model, is shown to be very useful in dealing with problems of clustering territorial areas. The good results obtained with its application to a set of parishes in southern Portugal encourage us to make further improvements to the framework, namely in terms of the multi-criteria optimisation model concerning the development of a more explicit Pareto frontier.

Annex: Counties and Parishes of the Case Study

Counties	Parishes	Area (Km^2)	Resident population	Counties	Parishes	Area (Km^2)	Resident population
Arraiolos	Arraiolos	146,08	3386	Portel	Alqueva	78,85	329
	Igrejinha	84,52	932		Amieira	98,29	362
	Santa Justa	42,92	225		Monte Trigo	107,11	1240
	S. Gregório	74,27	341		Oriola	36,25	400
	Gafanhoeira (S. Pedro)	46,12	494		Portel	156,44	2661
	Vimieiro	252,56	1589		Santana	41,95	542
	Sabugueiro	37,28	396		S. Bartolomeu do Outeiro	37,50	436
					Vera Cruz	44,62	458
Total		683,75	7363			601,01	6428
Estremoz	Arcos	23,89	1152	Évora	Nª Srª daBoa Fé	32,38	322
	Glória	72,75	532		Nª Srª da GraçaDivor	84,14	486
	Estremoz (Santa Maria)	63,30	6284		Nª Srª Machede	185,19	1123
	Evoramonte (Santa Maria)	99,38	569		Nª Srª daTouega	196,17	686
	Santa Vitória do Ameixial	55,51	342		Évora (SantoAntão)	0,27	1323
	Estremoz (Santo André)	0,60	2378		S. Bento do Mato	66,55	1151
	Santo Estevão	33,58	74		Évora (S. Mamede)	0,23	1724
	S. Bento do Ameixial	41,97	335		S. Manços	108,35	938
	S. Bento de Ana Loura	26,52	32		S. Miguel de Machede	81,52	794
	S. Bento do Cortiço	23,38	699		S. Vicente do Pigeiro	84,88	364
	S. Domingos de Ana Loura	16,30	341		Torre de Coelheiros	226,24	715
	S. Lourenço de Mamporcão	16,88	524		S. Sebastião da Giesteira	43	760
	Veiros	39,72	1036		Canaviais	19,41	3442

Counties	Parishes	Area (Km^2)	Resident population	Counties	Parishes	Area (Km^2)	Resident population
					N^a sr^a de Guadalupe	67,17	465
					Bacelo	10,30	9309
					Horta das Figueiras	45,38	10006
					Malagueira	19,05	12373
					Sé e S. Pedro	0,63	1691
					Sr^a da Saúde	36,21	8924
Total		513,78	14219			1307,07	56596
Montemor o Novo	Cabrela	192,26	649	Reguengos de Monsaraz	Reguengos de Monsaraz	101,72	7261
	Lavre	114,37	740		Campo	123,93	688
	N^a Sr^a do Bispo	121,83	4931		Corval	96,41	1389
	N^a Sr^a da Vila	187,03	6070		Monsaraz	88,29	782
	Santiago do Escoural	138,70	1335		Campinho	53,64	708
	S. Cristovão	145,92	540				
	Ciborro	55,49	714				
	Cortiçadas de Lavre	99,33	821				
	Silveiras	110,62	567				
	Foros de Vale Figueira	67,40	1070				
Total		1232,95	17437			463,99	10828
Redondo	Montoito	61,71	1298	Viana do Alentejo	Alcáçovas	268	2111
	Redondo	307,80	5733		Viana do Alentejo	94,70	2742
					Aguiar	30,97	890
Total		369,51	7031			393,67	5743

Source: INE (2011)

References

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

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

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

Bozkaya, B., E. Erkut, D. Hiaght, and G. Laporte. 2011. Designing new electoral districts for the city of edmonton. *Interfaces* 41(6): 534–547.

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

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

Cortona, P.G., C. Manzi, A. Pennisi, F. Ricca, and B. Simeone. 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.

Duque, J.C., L. Anselin, and S.J. Rey. 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., and G. Guénette. 1990. Decision support system for the school districting problem. *Operations Research* 38: 15–21.
Ferligoj, A., and V. Batagelj. 1982. Clustering with relational constraint. *Psychometrika* 47(4): 413–426.
Flischmann, B., and J.N. Paraschis. 1988. Solving a large scale districting problem: A case report. *Computers Operational Research* 15(6): 521–533.
Grafinkel, R.S., and G.L. Nemhauser. 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., B. Jaumard, C. Meyer, B. Simeone, and V. Doring. 2003. Maximum split cluster under connectivity constraints. *Journal of Classification* 20: 143–180.
Hess, S.W., and S.A. Samuels. 1971. Experiences with a sales districting model: criteria and implementation. *Management Science* 18(4): 41–54.
Hess, S.W., J.B. Siegfeldt, J.N. Whelan, and P.A. Zitlau. 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*. Portugal: Instituto Nacional de Estatística.
Lefkovitch, L. 1980. Conditional clustering. *Biometrics* 36(1): 45–58.
Legendre, P. 1987. Constrained clustering. In *Developments in numerical ecology. NATO ASI series*, vol. 14, ed. P. Legendre, and L. Legendre, 289–307. Berlin: Springer.
Maravalle, M., and B. Simeone. 1995. A spanning tree heuristic for regional clustering. *Communications in Statistics-Theory and Methods* 24(3): 625–639.
Mehrotra, A., E.L. Johnson, and G.L. Nemhauser. 1998. An optimization based heuristic for political districting. *Management Science* 44(8): 1100–1114.
Mehrotra, A. 1992. Constrained graph. PhD thesis, Georgia Institute of Technology.
Muyldermans, L., D. Cattrysse, D.V. Oudheusden, and T. Lotan. 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.
Pereira, A. 1995. Regionalism in Portugal. In *The European Union and the regions*, ed. Barry Jones, and Michael Keating, 269–280. Oxford: Claredon Press.
Park, K., K. Lee, S. Park, and H. Lee. 2000. Telecommunications node clustering with the node compatibility and network survivability requirements. *Management Science* 46(3): 363–374.
Portnov, B.A., and M. Schwartz. 2009. Urban clusters as growth foci. *Journal of Regional Science* 49(2): 287–310.
Shanker, R.J., R.E. Turner, and A.A. Zoltners. 1975. Sales territory design: an integrated approach. *Management Science* 22(3): 309–320.
Tavares-Pereira, F., J. Figueira, V. Mousseau, and B. Roy. 2007. The public transportation network pricing system of the Paris region. *Annuals of Operations Research* 154: 69–92.
Vickrey, W. 1961. On the prevention of gerrymandering. *Political Science Quarterly* 76(1): 105–110.
Wise, S.M., R.P. Haining, and J. Ma. 1997. Regionalisation tools for exploratory spatial analysis of health data. In *Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling and Computational Intelligence*, ed. M. Fischer, and A. Getis, 10–83. New York: Springer.
Zoltners, A.A., and P. Sinha. 1983. Sales territory alignment: a review and model. *Management Science* 29(3): 1237–1256.