An evolutionary approach to the delineation of functional areas based on travel-to-work flows

Francisco Flórez-Revuelta*

Department of Computing Technology, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain

José Manuel Casado-Díaz, Lucas Martínez-Bernabeu

Institute of International Economics, University of Alicante, P.O. Box 99, E-03080, Alicante, Spain

Abstract: This paper presents a new approach to the delineation of local labor markets based on evolutionary computation. The aim of the exercise is the division of a given territory into functional regions based on travel-to-work flows. Such regions are defined so that a high degree of inter-regional separation and of intra-regional integration in both cases in terms of commuting flows is guaranteed. Additional requirements include the absence of overlapping between delineated regions and the exhaustive coverage of the whole territory. The procedure is based on the maximization of a fitness function that measures aggregate intra-region interaction under constraints of inter-region separation and minimum size. In the experimentation stage two variations of the fitness function are used, and the process is also applied as a final stage for the optimization of the results from one of the most successful existing methods, that used by the British authorities for the delineation of Travel-to-Work Areas (TTWAs). The empirical exercise is conducted using real data for a sufficiently large territory which is considered to be representative given the density and variety of travel-to-work patters that it embraces. The paper includes the quantitative comparison with alternative traditional methods, the assessment of the performance of the set of operators which has been specifically designed to handle the regionalization problem, and the evaluation of the convergence process. The robustness of the solutions, something crucial in a research and policy-making context, is also discussed in the paper.

Keywords: Evolutionary Computation, Regional Science, Economics, Optimization methods

1 Delineation of Local Labor Markets

The last years have witnessed a renewed interest in the estimation of economic models at local levels, something which has been fostered by the persistence of substantial local/regional differences in relevant variables such as unemployment, and indicators on cohesion and competitiveness, in a context of general growth of developed countries' wellbeing. When conducting this kind of empirical exercises the election of the spatial framework is a crucial step in the research process. One obvious election is the use of administrative boundaries. This geographical reference is however of little help nowadays since economic activity exceeds traditional administrative limits and spreads throughout wider territories thanks to the improvement in the access to faster means of transport. The patterns of location of places of residence and economic activities are moreover becoming more complex, this resulting in multiple-direction flows that lead to the obsolescence of administrative areas as functional regions. A new concept of 'locality' is therefore needed in which functional coherence dominates, as a crucial first step for meaningful research and policy-making purposes.

Delineating local labor market areas (LLMAs) is an exercise that has become very common in the last decades across developed countries [1]. As stated above, these sets of functional areas are seen as an alternative to the use of local and regional administrative areas as the relevant geography for statistical purposes and for the design, implementation and monitoring of labor market and other public policies in related fields such education and housing markets. It is recognized that administrative areas are defined by boundaries that very frequently were established for historical reasons, and therefore it is not assured that they provide a meaningful insight of the territorial functional reality. The measurement of unemployment, for example, can be seriously biased if conducted on the grounds of administrative areas which on the one hand are frequently composed of very independent regions with different features that are averaged up in the aggregate -something that biases the perception of any external observer- and, on the other hand, are only one part of an upper tier functional reality whose consideration is necessary to fully understand the processes affecting a given territory. One of the scarce pieces of work where these effects have been formally tested is [2]. In this paper the relevance of the choice of a specific regionalization was tested through the use of five alternative geographical frameworks for the estimation of a multiregional model of labor supply. The conclusions of their analysis show that it is not admissible to ignore the potential

effects of the spatial representation choice given the effects that the different geographies have on the estimated parameter values and the model performance.

Some countries have delineated homogeneous LLMAs through the aggregation of basic territorial units such as counties, wards or municipalities that are similar in their characteristics (a review of these procedures can be found in [3]). The most popular strand however consists in the aggregation of building blocks based on the observed interaction among them in terms of travel-to-work.

An intuitive approach to the concept of LLMA leads us to define it as the area where the majority of the interaction between a group of workers seeking jobs and employers recruiting labor occurs (i.e. its boundaries should be defined so that relatively few people travel between home and work across them). This refers to what Goodman [4] called *external perfection* (the boundary of the area is rarely crossed in daily journeys to work) and together with a high degree of intra-market movement (so that the defined market is internally active and so as unified as possible) is the basis of the ideal LLMA. More than a decade ago Eurostat [5] established a code of good practices to guide the selection of a specific procedure through a project led by Prof. Coombes: (a) the ideal map of LLMAs should be based on statistical criteria, thus defined in a consistent way to allow comparison for statistical and policy purposes, (b) the procedure should allow the delineation of boundaries between areas within which most people both live and work, (c) each basic spatial unit should be in one, and only one LLMA, (d) contiguity should be respected, (e) a certain degree of self-containment should be reached, so that most of the LLMA's workers live in that area and most of the area's employed residents should work locally, (f) the map should consist on homogeneous units whose size should overpass a minimum threshold, (g) the areas defined should not be unnecessarily complex from a topographic point of view, (h) the map of LLMAs should respect where possible the standard administrative top tier boundaries, this being considered advantageous from both statistical and policy points of view and finally (i) the procedure should be flexible enough to allow evaluation and adjustment, although the possibility of varying the statistical criteria between regions must be excluded. The preference for detail (delineating as many criteria-meeting LLMAs as possible) is also frequently included as one additional criterion in this kind of exercises.

Despite sharing a common basic view about the ideal features of such an area, current official methods have a very diverse nature and are mostly based in sets of rules whose sophistication substantially varies nationally and, to a certain degree, temporally. Among the diverse possible criteria, and based on previous works by them and other authors, Casado and Coombes [6] classify these official procedures by distinguishing between inductive and deductive methods, on the one hand, and between hierarchical and multi-step, on the other. Deductive methods depart from the identification of the potential centers of LLMAs and proceed by merging residual geographical units with them to form a whole map of LLMAs. On the contrary, inductive methods do not depart from a preconceived idea of which the centers are, and give a central role to commuting flows from the beginning. Hierarchical methods consist on a rule that indicates which basic units should merge first according to the intensity of their relationship in commuting terms and proceed iteratively by applying the same rule until a termination condition is met. Finally, multi-step procedures consist on much more sophisticated sets of rules and frequently have a theoretical model behind that guides the decisions of when and why a certain rule has to be applied. One of the procedures that has a longer history and has been more widely applied is that of Coombes et al. [7] which has been used in the United Kingdom for the delineation of LLMAs (so-called Travel-to-Work Areas, TTWAs) since the decade of 1980. This sophisticated procedure was also used, with minor changes, to define LLMAs in Italy [8-10], Spain [11], New Zealand [12] and Australia [13], among other countries. This is the procedure that inspires the one proposed in the article.

Given the complexity of the problem an exhaustive search of the optimal solution is not possible in many cases. This is the reason why some kind of heuristic is needed. In this paper we opt for an evolutionary approach. Although computational intelligence techniques are becoming quite popular in Economics and Finance [14,15], and there are many techniques for clustering or grouping employing evolutionary computation [16-18], the use of such techniques for the delineation of functional areas has not been attempted.

In our proposal the regionalization problem is presented as the maximization of markets' internal cohesion in terms of travel-to-work subject to a number of restrictions among which stands meeting certain self-containment and minimum size (in terms of occupied population) thresholds, with the aim of identifying as many independent markets as possible, and without making use of geographical distance measures or contiguity constrictions (although the contiguity between the units constituting a functional region is desirable, the use of such conditions during the first steps of the regionalization procedures are likely to bias the results [6]). Unlike other procedures, the method proposed here meets the criteria listed above, and results in a significant improvement in measurable indicators such as the number of identified LLMAs which satisfy the stated criteria. We illustrate our approach using the latest Census data available for Spain [19].

The multiple constraints the problem involves (such as the absence of overlapping and the exhaustive coverage of the territory) result in the number of valid solutions being extraordinarily small with regards to the search space. This is the reason why an evolutionary algorithm with standard representation and operators does not lead to a near-optimal solution in a feasible time. A deep knowledge of the problem has allowed the design of an extensive set of crossover and

mutation operators, some of which have similarities with those used in other grouping and clustering problems to which they could be immediately applied, whilst other are much more related to the very specific nature of the problem, i.e. the delineation of LLMAs. This is a strategy that has proved to allow reaching the final solution, and to do so much more rapidly [20].

Next section is devoted to describe the problem formally. Third section is concerned with the development of our evolutionary proposal, including how the individuals are represented, the selection procedure and the explanation of the recombination and mutation operators, most of which have been specifically designed for this problem. In Section IV an application of the method is presented. The Region of Valencia, Spain, is the territory chosen for the experimentation. The election of this case study is justified at the beginning of the section which also includes the assessment of the relative performance of the diverse operators, the analysis of the robustness of the delineation results as well as the evaluation of the convergence process. The concluding section contains some final remarks and sketches a few of the possible extensions of the model in terms both of empirical applications in different research and policy-making processes and regarding methodological future developments.

2 Problem formulation

Let $A = \{A_1, A_2, \ldots, A_n\}$ be a set of areas (the territory to be divided into functional areas). The objective is to obtain the set of regions (LLMAs) $R = \{R_1, R_2, \ldots, R_m\}$ so as $\bigcup_{i=1}^{m} R_i = A$ and $R_i \cap R_j = \varnothing, \forall i, j \in [1, m], i \neq j$, that maximizes a fitness function f based on the interaction index II between an area and the rest of the region it belongs to:

$$II(A_k) = \frac{W_{A_k, R(A_k)-A_k}^2}{W_{A_k, A} \cdot W_{A, R(A_k)-A_k}} + \frac{W_{R(A_k)-A_k, A_k}^2}{W_{R(A_k)-A_k, A} \cdot W_{A, A_k}}, A_k \in A \quad (1)$$

being $R(A_k)$ the region containing area k, and

$$W_{R_s, R_t} = \sum_{\forall A_i \in R_s} \sum_{\forall A_j \in R_t} W_{A_i A_j} \quad (2)$$

where $W_{A_i A_j}$ is the number of commuters from area A_i to area A_j, that is the number of employed residents in area A_i that work in area A_j.

Also, expressed in a more intuitive form as:

$$II(A_k) = PE_{A_k, R(A_k)-A_k} \cdot PJ_{A_k, R(A_k)-A_k} + \\ + PE_{R(A_k)-A_k, A_k} \cdot PJ_{R(k)-A_k, A_k} \quad (3)$$

being

$$PE_{A_k, R(A_k)-A_k} = \frac{W_{A_k, R(A_k)-A_k}}{W_{A_k, A}} \quad (4)$$

$$PJ_{A_k, R(A_k)-A_k} = \frac{W_{A_k, R(A_k)-A_k}}{W_{A, R(A_k)-A_k}} \quad (5)$$

$$PE_{R(A_k)-A_k, A_k} = \frac{W_{R(A_k)-A_k, A_k}}{W_{R(A_k)-A_k, A}} \quad (6)$$

$$PJ_{R(A_k)-A_k, A_k} = \frac{W_{R(A_k)-A_k, A_k}}{W_{A, A_k}} \quad (7)$$

For a better understanding of each factor, in its general form, PE_{R_j, R_i} is the proportion of employees residing in R_i that work in R_j, PJ_{R_j, R_i} is the proportion of jobs in R_j that are held by employees residing in R_i. In our case we calculate the interaction index between an area A_k, that is considered as a single area region, and the region resulting from subtracting A_k to the region $R(A_k)$ it belongs to. This interaction index between an area and its region is a generalization of the interaction index used in [7] and discussed among others in [6]. We have tested our method with two different fitness functions based on (1). First, we intend to maximize the sum of the interaction indexes for all the areas of the territory:

$$f_1 = \sum_{\forall A_i \in A} II(A_i) \quad (7)$$

Second, since one of the criteria stated in Section 1 is *detail*, i.e. reaching the highest possible number of independent LLMAs, we have also considered a variation of that function:

$$f_2 = card(R) \cdot \sum_{\forall A_i \in A} II(A_i) \quad (8)$$

Besides, each region $R_i \in R$ must fulfill two requirements in terms of minimum self-containment (β_1, β_2), $\beta_2 \geq \beta_1$ and minimum size (β_3, β_4), $\beta_3 \geq \beta_4$:

$$\min\left(\frac{W_{R_i, R_i}}{W_{R_i, A}}; \frac{W_{R_i, R_i}}{W_{A, R_i}}\right) \geq \beta_1 \quad (9)$$

$$W_{R_i, A} \geq \beta_4 \quad (10)$$

being $W_{R_s, R_t} = \sum_{\forall i \in R_s} \sum_{\forall j \in R_t} W_{i, j}$ the total amount of employees that reside in R_s and work in R_t.

A trade-off between both constraints has been introduced similarly to [7], but in the formulation proposed by Casado [11]. According to this trade off, the minimum self-containment requirement is relaxed for regions which are sufficiently large following a linear relationship (this allows identifying more separate LLMAs in very urbanized environments, something desirable according to the principles listed in Section 1). This trade-off is considered as follows:

$$\min\left(\frac{W_{R_i, R_i}}{W_{R_i, A}}; \frac{W_{R_i, R_i}}{W_{A, R_i}}\right) \geq a + b \cdot W_{R_i, A} \quad (11)$$

$$a = \beta_2 - b\beta_4 \quad (12)$$

$$b = \frac{\beta_2 - \beta_1}{\beta_4 - \beta_3} \quad (13)$$

We have also included a requisite to guarantee some degree of contiguity by employing only commuting data: an area can only belong to a region if it is reachable from any other area of

that region through the γ highest outgoing/incoming commuting flows (by checking graph connectivity).

3   Evolutionary proposal

The structure of the evolutionary algorithm for the regionalization of the territory follows the next steps:

Step 1. Produce an initial population of a given size. The whole set of areas (the whole territory A) is taken as individual #1 (this is an individual meeting all the requirements). Complete the initial population with n-1 randomly generated individuals (experimentation showed that most of these are invalid solutions).

Step 2. Evaluate fitness of all individuals and sort them accordingly.

Step 3. Generate *nr* new individuals by recombination, selecting the parents from the population by fitness-proportional probability, and selecting the operator to apply according to a pre-established probability.

Step 4. Generate *nm* new individuals by mutation, selecting the original individual from the population by fitness probability, and selecting the operator to apply according to a pre-established probability.

Step 5. Evaluate fitness of all new individuals.

Step 6. Sort the whole population, composed of *n+nr+nm* individuals, by their fitness value.

Step 7. Generate a new population choosing the n best individuals.

Step 8. Stop condition: if the best individual remains without changes for *g* generations, finish. Otherwise, return to step 3.

3.1 Individual Representation

The individuals of the population represent possible solutions, that is, the aggregation of the whole set of areas composing territory A into no over-lapping local labor markets (regions). There are different alternatives for the encoding of a data set grouping [16-18]. In our case, we have chosen a group-number encoding where each individual is represented by a vector of n components, each of which corresponds to an area of A, and takes the value of the identifier of the region the area belongs to (Fig. 1).
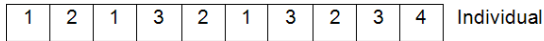
$R_1 = \{A_1, A_3, A_6\}$ $R_2 = \{A_2, A_5, A_8\}$ $R_3 = \{A_4, A_7, A_9\}$ $R_4 = \{A_{10}\}$

Fig. 1. Individual representation

3.2 Selection

The selection of the individuals to be affected by recombination and mutation operations is performed following a ranking method [21], according to which those individuals scoring higher in the fitness function have a larger probability of being selected.

3.3 Recombination Operators

Due to the large number of constraints that the individuals must fulfill, and very notably to the fact that in a regionalization exercise it is important to guarantee the exhaustive coverage of the territory and the avoidance of overlapping between regions, the usual operator of recombination does not in many cases lead to feasible solutions. This is the reason why we have designed a wide group of specific operators that has proved to allow a more rapid evolution of the population towards acceptable solutions:

1) *Recombination1*: A crossover point is randomly selected. Offspring is generated by taking the initial part of one of the parents and the final part of the other one. This is the standard 1-point crossover operator. However, in this specific case this operator frequently results in invalid offspring due to the lack of a compatible correspondence between the region identifiers of both parents (see Fig. 2 – region 3 in parent #1 and region 4 in parent #2 are identical but they are codified with a different region identifier, resulting in a fragmentation of the region in the offspring).
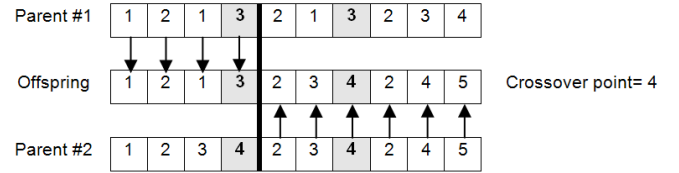

Fig. 2. Operator Recombination1

2) *Recombination2*: A region identifier belonging to parent #1 is randomly chosen. The areas with identifiers lower or equal to the chosen one are inherited by the offspring. Each of the rest of the areas takes the identifier from parent #2, except in those cases when it belongs to a region from which one or more of its constituting areas were already in the offspring. In such cases, the areas take the identifiers from parent #1 (see Fig. 3 – areas in parent #2 belonging to region R_3 must be assigned to R_1 in the offspring).
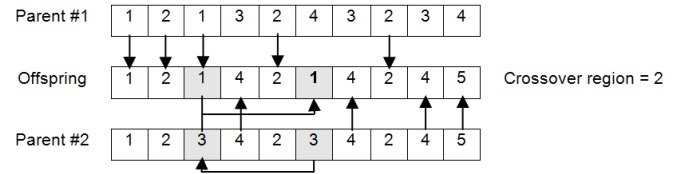

Fig. 3. Operator Recombination2

3) *Recombination3*: a crossover point is randomly selected. For the areas previous to that point, the offspring takes the values of parent #1. From that crossover point, values from parent #2 are inherited, unless this involves a region with an area already set in the offspring; in such cases the identifier of parent #1 is used (see Fig. 4 – areas in parent #2 belonging to R_3 and R_4 must be assigned to R_1 and R_3 respectively.
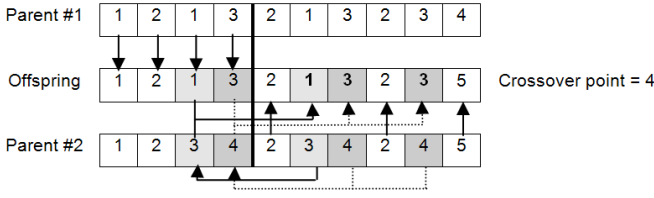
Fig. 4. Operator Recombination3

4) Since the areas characterized by lower identifiers are also assigned to regions with lower identifiers, their probability of being taken from parent #1 is greater than that of areas with high identifiers. To cope with this we have added two recombination operators (*Recombination4* and *Recombination5*), as variations of *Recombination2* and *Recombination3* respectively. In them a random recoding of the regions in the representation of both parents is performed previously to the recombination.

3.4 Mutation Operators

We have designed an extensive set of mutation operators, some of them specifically intended for the delineation of local labour market areas, with the aim to accelerate the obtaining of individuals with adequate fitness. These operators have four main functions: division of regions, fusion of regions, reassignment of single areas, and reassignment of group of areas.

1) *Division1*: This operator divides a region into two. The splitting process is as follows:
   a) A region R_i is randomly selected. It must fulfill two constraints to guarantee that the region is large enough to be divided:

   $$W_{R_i,A} > 2\beta_4 \qquad (14)$$

   $$W_{R_i,A} - W_{focus(R_i),A} > \beta_4 \qquad (15)$$

   where

   $$focus(R_i) = \arg\max_{\forall A_s \in R_i}\left(W_{A_s,A} + W_{A,A_s}\right) \qquad (16)$$

   b) An area belonging to R_i is randomly chosen and assigned to the new region R_i'.
   c) Another area belonging to R_i is randomly chosen. It is then assigned to the new region R_i''.
   d) The rest of the areas belonging to R_i are taken at random, being assigned to the region (R_i' or R_i'') with which each of them has the strongest link according to the interaction index.

2) *Division2*: This operator creates a new region from another one by removing from the latter a number of areas sufficiently large so as to form a valid market:
   a) As in step a) of *Division1*.
   b) An area belonging to R_i is chosen at random, being assigned to the new region R_i'.
   c) If region R_i' does not fulfill the size constraint (10), it

takes the area belonging to R_i with which it has the highest interaction index. This process is repeated until R_i' is large enough.

3) *Division3*: This operator divides one region into two, each with a similar number of areas:
   a) As in steps a) to c) of *Division1*.
   b) R_i' y R_i'', alternately, take the area of R_i with which it has the highest interaction index, until no area remains assigned to region R_i.

4) *Fusion1*: Two randomly selected regions are merged.

5) *Fusion2*: A region is randomly chosen. Each of its constituting areas is then assigned to its optimal region, that is, the region with which it has the highest interaction index:

$$R'(A_i) = \arg\max_{\forall R_j \in R - R(A_i)}\left(\frac{W_{A_i,R_j}^2}{W_{A_i,A} \cdot W_{A,R_j}} + \frac{W_{R_j,A_i}^2}{W_{R_j,A} \cdot W_{A,A_i}}\right) \qquad (17)$$

So, as a result of this operator, the number of regions in the offspring is one less compared to its parent.

6) *Reassignment1*: Similar to the standard mutation operator in EC, randomly reallocating to any region up to 1 per cent of the areas in the territory.

7) *Reassignment2*: This operator is analogous to *Reassignment1*. However, the destination region for each mutated area is its optimal region according to (17) (as in *Fusion2*).

8) *Reassignment3*: An exchange of areas between regions is performed. One area is randomly chosen and it is assigned to its optimal region. One area of that optimal region is then transferred to the region of origin.

9) *GlobalReassignment1*: This operation removes from a region the areas that score lower in the interaction index when measured with regards the rest of the region. Such areas are then assigned to their optimal regions:
   a) As in step a) of *Division1*.
   b) The area to remove is selected as:

   $$A_s = \arg\min_{\forall A_j \in R_i}\left(\frac{W_{A_j,R_i-A_j}^2}{W_{A_j,A} \cdot W_{A,R_i-A_j}} + \frac{W_{R_i-A_j,A_j}^2}{W_{R_i-A_j,A} \cdot W_{A,A_j}}\right) \qquad (18)$$

   c) If $W_{R_i-A_s,A} > \beta_4$ (R_i is large enough), the area A_s is assigned to its optimal region, and step b) is repeated. If that condition is not fulfilled, mutation is finished.

10) *GlobalReassignment2*: This operator is similar to *Reassignment2* since areas are assigned to their optimal regions. In this case, however, instead of a single area a group of them is transferred. Such a group is chosen so that it is composed of highly integrated areas. The process is as follows:
    a) An area i is randomly selected.
    b) The k areas belonging to R_i with which area i has the highest interaction are also selected; k is chosen at random.

c) All the selected areas are assigned to the optimal region for area i.

11) *GlobalReassignment3*: As in some cases there is a high level of interaction between regions, this operator tries to redistribute areas in such regions. The procedure is:

a) A number $k \geq 2$ of regions to be mutated is randomly chosen.

b) A region R_i is selected at random.

c) The k-1 regions that have a higher degree of interaction with R_i are selected.

d) These regions are then disintegrated into their constituting areas.

e) k areas from this new group are selected at random. These areas act as seeds for the new regions.

f) The rest of unassigned areas are individually taken at random and merged with their optimal region among those k new regions.

12) *GlobalReassignment4*: This operator is very similar to the previous one. Only the way in which new regions are generated has been modified:

a) As in steps a) to d) of *GlobalReassignment3*.

b) For each one of the k new regions:

   i. A random area is selected as its seed.

   ii. This new region takes the areas with which it has the highest interaction, until the size of the region is larger than β_4 –fulfilling (10).

c) Unassigned areas are merged with their optimal region among those k new regions.

4 Experimentation

Our proposal has been implemented for the delineation of local labor markets in the Region of Valencia, Spain, using travel-to-work data derived from the Spanish Census of Population [19]. This data allowed us to build a 541x541 matrix (541 is the number of municipalities that integrate the Region), where each cell captures W_{ij}. The Region of Valencia is an appropriate case study since it can be considered a geographical laboratory where many of the typical commuting 'types' in any territory can be found. It will allow moreover the analysis of longitudinal changes in the areas identified thanks to the availability of relevant data which do not exist for other territories.   To set the case study context, the Region of Valencia includes metropolitan territories such as those around the capital city of Valencia, and the surroundings of the capital of the Northern Province of Castellón, where an industrial cluster specialized in the tile industry is located. These areas are characterized by a dense and complex multi-directional pattern of commuting flows which make it very difficult to identify separate functional areas, something that is however desirable from the governance and statistical point of view. The region also embraces rural areas inland where agriculture is still overrepresented and where a dual pattern of commuting is observed which distinguishes between those employed in the primary sector and those commuting longer distances to the service centers closer to the coast. The Southern province of Alicante is constituted, except for the coastal line, to a large extent by medium-size cities that are specialized in diverse manufacturing activities of the Marshallian industrial-district type, and attract workers from surrounding municipalities for which they act as employment centers. The coast is finally very service sector-focused, with a high specialization in tourism and second-residence building activities.

Parameters employed in the following examples are: size population = 100, offspring size = 170 (*nr*=50 and *nm*=120) where each recombination or mutation operator generates 10 new individuals, generations *g* without changes in the best individual to stop the process = 250. Parameter γ of 'functional neighborhood' is 5.

In the experimentation conducted for this paper we decided to test two different sets of the parameters related to self-containment and minimum size conditions, (9) to (11). In the first one, only the minimum self-containment requisite (9) is considered and we decided to fix the parameters values at the level most widely used in the relevant literature (see [6]): $\beta_1 = \beta_2 = 0.75$. The trade-off (11) to (13) does not apply in this case. The second set of values includes self-containment as well as minimum size constraints, (10) and (11); the values were in this case taken from the official method applied in the UK for the delineation of the official sets of Travel-to-Work Areas (TTWAs), i.e. $\beta_1 = 0.7$, $\beta_2 = 0.75$, $\beta_3 = 20000$ and $\beta_4 = 3500$.

Three different exercises have been conducted using both sets of parameters. The first two follow the steps described in Section 1 using fitness function f_1 (7) and f_2 (8). The third exercise is the optimization of delineations obtained through the application of the TTWAs method [7]. This procedure originally included as a final stage the possibility of using a heuristic to reassign some areas and therefore improve the results, although it was not finally applied for the official delineation of British TTWAs. In this third exercise, one of the individuals in the first generation is the result of applying Coombes' method. In this particular application we have added an extra constraint to ensure that no new regions (LLMAs) are created by our algorithm. To do so we do not allow the centers of each area to be altered. As there is not need of creating/removing regions, operators of division and fusion are not employed, so offspring size = 120, and fitness function is f_1.

We have conducted 50 independent runs of each test in order to validate our approach.

In the remaining of this section we start by discussing the results and comparing them in terms of the score reached in the fitness function and the number of LLMAs resulting from each test. We then discuss the convergence process and we finally conclude the section with the assessment of the relative relevance of each of the operators developed in getting the best individual.

Table 1   Results with minimum size constraints

	TTWAs method	Optimization of TTWAs method (f₁ is applied)			f₁			f₂		
		Best individual	Mean	Standard deviation	Best individual	Mean	Standard deviation	Best individual	Mean	Standard deviation
Number of labor markets	**44**	**44**	44	0	**54**	52.78	0.67	**61**	58.76	1
Fitness function	**f_1=2.73, f_2=120.23**	**2.86**	2.84	0.0086	**3.14**	3.11	0.014	**185.42**	179.41	3.56

Table 2   Results without minimum size constraints

	TTWAs method	Optimization of TTWAs method (f₁ is applied)			f₁			f₂		
		Best individual	Mean	Standard deviation	Best individual	Mean	Standard deviation	Best individual	Mean	Standard deviation
Number of labour markets	**95**	**95**	95	0	**114**	109.4	2.3	**124**	120.35	1.58
Fitness function	**f_1=3.11, f_2=295.633**	**3.34**	3.33	0.004	**3.67**	3.62	0.024	**447.04**	425.76	8.36

4.1 Results

The main goal of our research is to develop a method to obtain a near-optimal division of a given territory into functional areas according to the criteria expressed in Section I. So, besides performing a qualitative study some statistics have been generated to facilitate the comparison of the diverse alternatives.

Fig. 8 to 11 show that all the solutions are roughly similar in territorial terms. However, when results are analyzed in detail, we can observe the contribution of our evolutionary approach to the improvement (increase) in the number of markets and in the value of the fitness function. The main difference in the results obtained in the three exercises described above in graphical terms is of course the degree of division of the territory. Our proposal allows reaching finer-grained outcomes (for comparison purposes none of the regionalization exercises depicted there have been modified to achieve contiguity between the municipalities constituting each market; this can be easily done through a final optimization step that assigns non-contiguous municipalities to the contiguous regions with which they have the highest interaction index, provided that the rest of statistical constraints is satisfied).

Tables 1 and 2 show the results of the division of the Region of Valencia into functional areas according to each of the three exercises, and with two variants depending on whether the minimum size constraint is included or not. For comparison purposes also the results of applying the TTWAs' method [7] are included in the first column in both tables (this method produces 44 LLMAs when minimum size constraints are considered – Table 1- and 95 otherwise – Table 2). The values of f_1 and f_2 for both sets of markets are also included in that column. In the rest of columns in bold, and under the heading 'Best individual' the results of the best individual of the 50 independent runs that were conducted for each combination *exercise/ set of parameters* are displayed. As shown by Table 1, the use of our proposal as a final optimization step in the TTWAs' method increases f_1 by 4.8% (Fig.10). When our alternatives are applied using f_1 and f_2 the number of identified

LLMAs reaches 54 and 61 respectively, and the fitness improvement is apparent (from 2.73 to 3.14 –f_1- and from 120.23 up to 185.42 –f_2, i.e. 15% and 54.2% higher respectively; Fig. 8). As expected the use of fitness function f_2 results in a considerable increase in the number of markets (i.e. an enhancement in one of the criteria stated in Section 1, Detail). Besides, averages and standard deviations show that
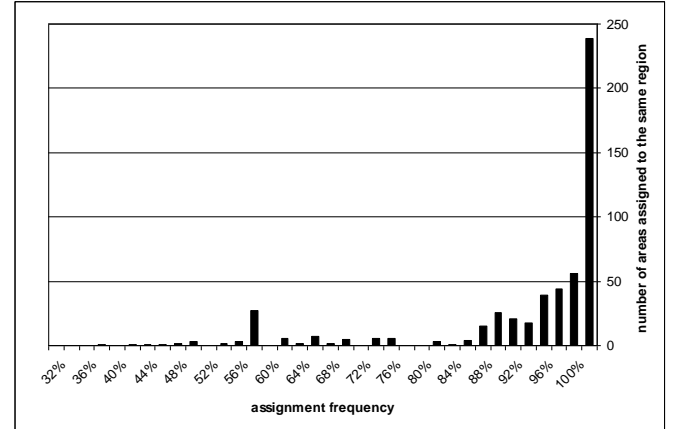

Fig. 5. Assignment frequency of each area (municipality) to the same region (LLMA) (50 runs; f_2 as fitness function; minimum size constraints included).
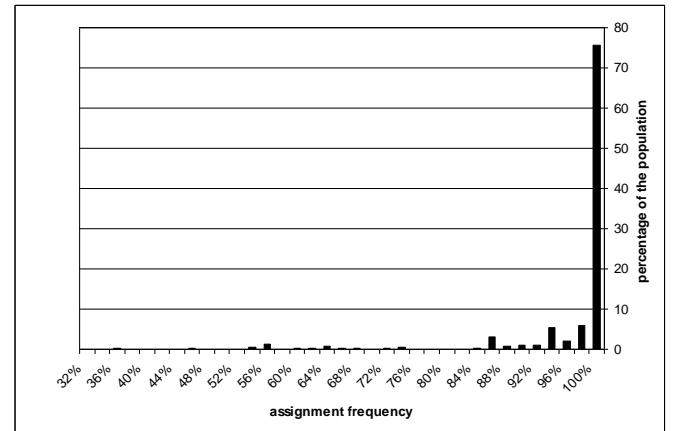

Fig. 6. Assignment frequency of each area (municipality) to the same region (LLMA) in terms of population (50 runs; f_2 as fitness function; minimum size constraints included).

the different runs of the algorithm reach comparable solutions in terms of quality, what gives robustness to our procedure. The outcomes of the exercises conducted using the alternative set of parameters (i.e. excluding minimum size constraints – Table 2, Fig. 9 and 11) are very similar.

Fig. 5 and 6 illustrate that conclusion. After running the procedure 50 times using f_2 and the set of parameters that include minimum size constraints (last group of columns in Table 1), the majority of areas (and a large proportion of total population) is systematically assigned to the same region (LLMA). According to Fig. 5 most areas (417 out of 541, 77.1 per cent) are assigned in more than 90 per cent of cases to the same market in all the independent runs. Among them, 239 are always assigned to the same region. These results give support to the capacity of the method to reach a near-optimal solution. Moreover, if we move the focus from the municipalities to the share of the population being assigned to the same market (Fig. 6), this conclusion becomes even more apparent. Eighty per cent of the population has been systematically (100 per cent of cases) assigned to the same market, and 90.73% is assigned to
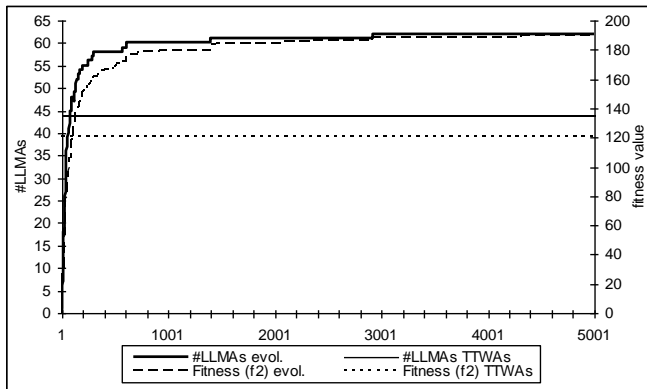

Fig. 7. Change in the number of LLMAs and fitness throughout the evolutionary process (including minimum size constraints; fitness function is f2) compared with the values from the TTWAs method.

the same market in 90 per cent of the runs. Those cases where the percentages of assignment are lower typically correspond to small areas whose commuting flows are similarly distributed between several regions, and where dominance cannot be clearly established.

4.2 Convergence Process

Convergence of the evolutionary process is rapidly achieved; and our approach improves the results obtained with the TTWAs method in less than 100 generations. Only a few more generations are needed so that the best individual almost reaches its final value. In the last 50 per cent of generations the solution is not improved in more than 1 per cent. Therefore the proposed procedure rapidly gives place to a good coarse division of the territory into functional areas, which is later refined by the application of -mainly- the recombination operators. These observations are illustrated here in Fig. 7, where our evolutionary proposal (as an example we have only included the case where f_2 is the fitness function and minimum

Table 3
Frequency of appearance of the operators in the best individual

	Operator	With size constraint		Without size constraint	
f_1	Recombination1	9.60%		9.30%	
	Recombination2	9.10%		11.50%	
	Recombination3	3.80%	48.80%	2.50%	45.50%
	Recombination4	12.60%		14.80%	
	Recombination5	13.70%		7.40%	
	Division1	12.60%		15.70%	
	Division2	4.60%	21.00%	6.40%	26.20%
	Division3	3.80%		4.10%	
	Fusion1	1.30%	3.50%	1.80%	5.10%
	Fusion2	2.20%		3.30%	
	Reassignment1	0.20%		0.10%	
	Reassignment2	7.90%	9.20%	5.00%	8.40%
	Reassignment3	1.10%		3.30%	
	GlobalReassignment1	9.60%		9.20%	
	GlobalReassignment2	4.40%	17.60%	2.80%	14.80%
	GlobalReassignment3	1.30%		1.30%	
	GlobalReassignment4	2.30%		1.50%	
f_2	Recombination1	10.20%		9.50%	
	Recombination2	8.70%		10.10%	
	Recombination3	4.00%	49.30%	2.30%	42.90%
	Recombination4	12.60%		13.20%	
	Recombination5	13.80%		7.80%	
	Division1	10.70%		14.90%	
	Division2	5.50%	19.90%	6.20%	26.40%
	Division3	3.70%		5.30%	
	Fusion1	0.70%	1.60%	1.00%	2.90%
	Fusion2	0.90%		1.90%	
	Reassignment1	0.20%		0.20%	
	Reassignment2	8.60%	11.30%	5.90%	10.20%
	Reassignment3	2.50%		4.10%	
	GlobalReassignment1	10.40%		11.80%	
	GlobalReassignment2	5.00%	17.90%	3.00%	17.60%
	GlobalReassignment3	0.90%		1.50%	
	GlobalReassignment4	1.60%		1.30%	
*	Recombination1	17.20%		17.10%	
	Recombination2	10.50%		15.30%	
	Recombination3	7.80%	80.50%	5.20%	77.90%
	Recombination4	19.70%		20.90%	
	Recombination5	25.30%		19.40%	
	Reassignment1	0.60%		0.20%	
	Reassignment2	7.50%	8.80%	6.60%	9.50%
	Reassignment3	0.70%		2.70%	
	GlobalReassignment1	6.00%		9.50%	
	GlobalReassignment2	3.40%	10.60%	2.10%	12.50%
	GlobalReassignment3	0.40%		0.50%	
	GlobalReassignment4	0.80%		0.40%	

* Optimization of TTWAs method through our evolutionary approach

size constraints are considered) is compared with the results from the official TTWAs method.

4.3 Behavior of the Operators

As stated above, a crucial part of our proposal includes the designing of specific operators that improve the search for a good solution, in terms of fitness function, but also in consumed time. Table 3 depicts the results of analyzing their trajectory within the final solution (best individual of the last generation), i.e. the sequence of operators that led from an individual of the first generation to that solution in the 50 independent runs for each combination *exercise/set of*

parameters.

According to Table 3 the operators of fusion and Reassignment1 almost never contributed to the final best individual. The other operators are more or less used depending on the specific form of the fitness function used and on the algorithm parameters. We have also tested that their use is also affected by the territory over which the procedure is applied. Division operators allow reaching the final number of markets from the original individual where all the areas were merged in a single region. The experimentation stage shows that the operators of recombination perform a local search in the space of solutions, contributing with small improvements in the value of the fitness function. On the other hand, mutations for the reassignment of individual areas -or groups of them- result in local searches as well as in great jumps which allow to departure from local minima. These are the reasons why we consider all of them necessary so as to reach a good solution, despite some of them being successful only rarely.

5   Conclusion

The degree of success in the delineation, implementation and monitoring of public policies in different contexts (statistics, labor markets, housing markets, transportation, urban planning…) heavily depends on the adequateness of the geographical reference. The delineation of functional regions consists on the aggregation of the basic spatial units constituting a territory into regions according to certain criteria. Very frequently this aggregation is based on information about the spatial interaction between such units. Examples of such datasets are commuting or migration origin-destination matrices between pairs of municipalities. One way of dealing with this problem is to maximize intra-region interaction under constraints of inter-region separation and (eventually) minimum size. The use of functional areas for policy making and statistical purposes makes it necessary to consider at least two additional requirements: absence of overlapping between regions and exhaustive coverage of the territory (every spatial basic unit must be allocated to one region).

Official methods for the delineation of functional areas have until now rely on procedures that very frequently were designed some decades ago and that can now be improved through the use of new procedures as evolutionary computation. The use of these techniques has allowed us to model the regionalization problem as one of optimization which is then solved through a genetic algorithm that builds upon one of the most widely and successfully applied procedures, the official method used for the delineation of so-called Travel-to-Work Areas (TTWAs) in Britain. Given the complexity of the specified requirements (notably the absence of overlapping between the groups and the need to assign every basic unit to one group) conventional operators hardly ever generate valid solutions. Therefore our genetic algorithm approach includes developing and testing a new set of operators which has been specifically designed to cope with the complex requirements typical of any functional regionalization, and that could be easily generalized for their use in other grouping problems based on interaction data.

The experimental results show that the proposed method out-performs current official procedures, since it manages to (a) identify a larger number of Local Labour Market Areas (LLMAs) -the preference for detail (delineating as many criteria-meeting LLMAs as possible) is one of the principles that guide these exercises- and (b) improve the fitness (i.e. the internal integration between the areas that constitute these markets) without failing to fulfill the rest of statistical requirements as listed in codes of good practices like that of the Statistical Office of the European Union [5] (in particular all LLMAs meet the minimum self-containment condition).

The major concern in this policy making context is undoubtedly the fact that the use of our evolutionary approach does not guarantee that the results of the regionalization exercise would remain unaltered in different sets of runs. Despite giving place to worse results in the referred terms, traditional methods are consistent through different applications. In this paper we provide evidence that indicates that this 'uncertainty' problem is very limited. The discrepancies observed in different runs of the whole procedure are rare, have a low relevance in terms of the population affected, and are in any case constrained to a specific set of municipalities typically situated on nebulous boundaries between functional areas which are equally 'attractive' for them (and whose assignment by traditional procedures is to a certain extent arbitrary).

Although the particular application in this paper has been the analysis of commuting flows, other applications which constitute typical Regional and Urban Economics topics are also possible, including grouping based on shopping trips, communication flows, and trips derived from the use of public services, among others. More generally this procedure is potentially useful for the delineation of clusters in which the guiding principle is not the *similarity* between the units to be grouped but the *interaction* between them.

Our work currently focuses in several extensions. First, the changing success of the operators during the process of generation of good individuals is leading us to consider the use of a self-adaptive variation of the method, i.e. the proportion of application of each operator will depend on previous results. Therefore each operator would be applied depending on several criteria, such as success, temporal cost or improvement in the best individual, among others. Second, we are designing a parallel implementation of the algorithm to improve convergence in large problems. Third, other representations that have been used in grouping or clustering problems such as the Grouping Genetic Algorithms [16] are currently being explored in order to measure their convergence time. Finally, as the problem has different components that could eventually be evaluated, a re-formulation of the problem based on multi-objective optimization is also being considered.

6 References

[1] OECD, *Redefining Territories. The Functional Regions*. Paris: OECD Publications, 2002

[2] J. H. Baumann, M. M. Fischer, and U. Schubert, "A multiregional labour supply model for Austria: The effects of different regionalisations in multiregional labour market modelling," *Papers of the Regional Science Association* 52, pp. 53-83, 1983.

[3] J. C. Duque, "Design of homogeneous territorial units: A methodological proposal and applications," Unpublished PhD dissertation, Universidad de Barcelona, Spain, 2004.

[4] J. F. B. Goodman, "The Definition and Analysis of Local Labour Markets: Some Empirical Problems," *British Journal of Industrial Relations*, 8(2), pp. 179-196, 1970.

[5] EUROSTAT, and M. G. Coombes: "Étude sur les zones d'emploi." Document E/LOC/20. Office for Official Publications of the European Communities, Luxembourg, 1992.

[6] J. M. Casado-Díaz, and M. G. Coombes, "The Delineation of 21st Century Local Labour Market Areas (LLMAs)," presented at the 8th Nectar Conference, Las Palmas de Gran Canaria, Spain, 2005.

[7] M. G. Coombes, A. E. Green, and S. Openshaw, S., "An efficient algorithm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain," *Journal of the Operational Research Society*, 37, pp. 943-953, 1986.

[8] ISTAT-IRPET, *I mercati locali del lavoro in Italia.* Milan: Angeli, 1989.

[9] ISTAT-IRPET, *I sistemi locali del lavoro in Italia.* Rome: ISTAT, 1997.

[10] ISTAT (2005): *I sistemi locali del lavoro 2001*. Available: http://www.istat.it

[11] J. M. Casado-Díaz, "Local Labour Market Areas in Spain: A Case Study," *Regional Studies*, 34(9), pp. 843-856, 2000.

[12] K. Papps, and J. O. Newell, "Identifying Functional Labour Market Areas in New Zealand: A Reconnaissance Study Using Travel-to-Work Data," Institute for the Study of Labor (IZA), Bonn, Discussion Paper nº. 443, 2002

[13] M. Watts, "Local Labour Markets in New South Wales: Fact or Fiction?," Centre of Full Employment and Equity, The University of Newcastle, Australia, WP 04-12, 2004.

[14] S.-H. Chen (ed.), "Evolutionary Computation in Economics and Finance," in *Studies in Fuzziness and Soft Computing*, n. 100. Physica-Verlag, 2002.

[15] S.-H. Chen, and P. P. Wang (eds.), *Computational Intelligence in Economics and Finance*. Springer-Verlag, 2004.

[16] E. Falkenauer, *Genetic Algorithms and Grouping Problems*. John Wiley & Sons, 1998

[17] A. Tucker, J. Crampton, and S. Swift, "RGFGA: An Efficient Representation and Crossover for Grouping Genetic Algorithms," *Evolutionary Computation*, 13(4), pp. 477-499, 2005.

[18] R. M. C., "Clustering with Genetic Algorithms," Department of Computer Science, University of Western Australia, M.Sc. Thesis, 1998.

[19] Instituto Nacional de Estadística (2001). Censo de población. Available: http://www.ine.es

[20] Y. Jin (ed.), *Knowledge Incorporation in Evolutionary Computation*, Studies in Fuzziness and Soft Computing, 167, Springer-Verlag Berlin Heidelberg, 2005.

[21] K. de Jong, *Evolutionary Computation: A Unified Approach*. MIT Press, 2006.

Francisco Flórez-Revuelta received the M.Sc. and Ph.D. (*summa cum laude*) degrees in Computer Science from the University of Alicante, Spain, in 1994 and 2001, respectively.

He is currently an Associate Professor in the Polytechnic University College, University of Alicante, where he belongs to the Department of Computing Technology and the Institute for Computing Research. His research is devoted to the application of computational intelligence techniques, such as evolutionary computation and neural networks, to problems of different disciplines: computer vision, smart homes, robotics, economics, earth sciences…

Dr. Flórez-Revuelta is a member of the IEEE Computer and the IEEE Computational Intelligence societies. He is also member of ACM, belonging to the Special Interest Groups on Artificial Intelligence, Computer-Human Interaction, and Genetic and Evolutionary Computation.

José M. Casado-Díaz received the M.Sc. (First-Class Honours with Distinction) and Ph.D. (*summa cum laude*) degrees in Economics and Business from the University of Alicante, Spain, in 1993 and 1998, respectively.

He is currently an Associate Professor in the Faculty of Economics and Business, University of Alicante, where he belongs to the Department of Applied Economic Analysis and to the Institute for International Economics. His research interests include the delineation and analysis of local labour market areas, the geographical mobility of workers, and the application of evolutionary computation techniques to the identification of functional areas. Among his publications, J.M. Casado-Díaz, "Spatial Labour Markets in Spain: A Case Study", *Regional Studies*, vol. 34, No. 9, pp. 843-856, 2000; J.M. Casado-Díaz, "Who are the commuters? A microdata analysis of travel-to-work" in F. Columbus (Ed.) *European Economic and Political Issues*, New York, NY: Nova Science Publishers, pp. 25-50; J.M. Casado-Díaz and H. Simón, "Industry Wage Premia and Collective Bargaining Revisited: Evidence from Spain", *Applied Economic Letters*, forthcoming.

Lucas Martínez-Bernabeu received the M.Sc. degree in computer sciences from the University of Alicante, Spain, in 2003.

He is currently a research fellow in the Institute of International Economics of the University of Alicante. He is also a Ph.D. student in computer sciences in the Department of Computing Technology of the University of Alicante. His research interests are metaheuristics, evolutionary algorithms, artificial neural networks and computer architecture.
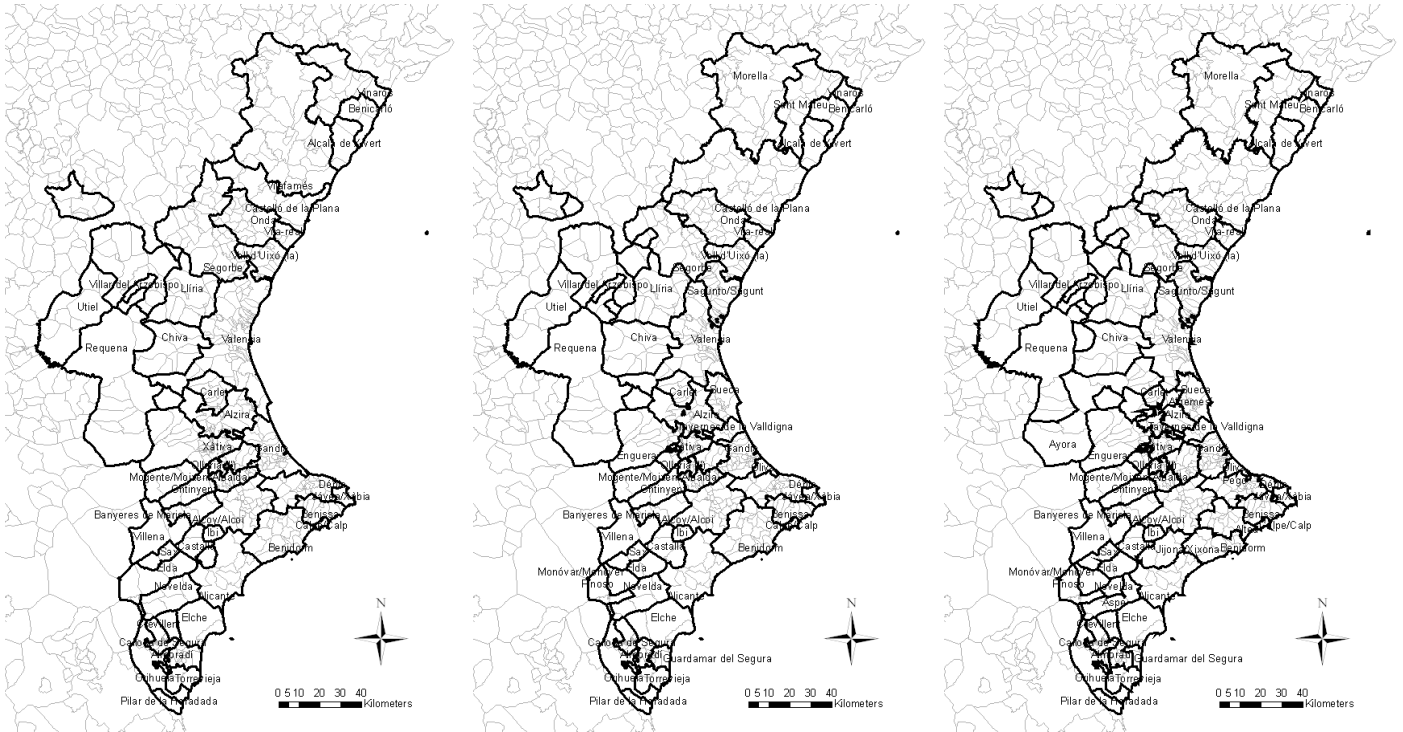
Fig. 8. Regionalization of the Region of Valencia with minimum size constraints applying TTWAs method (left); and our proposal: f_1 (center) and f_2 (right).
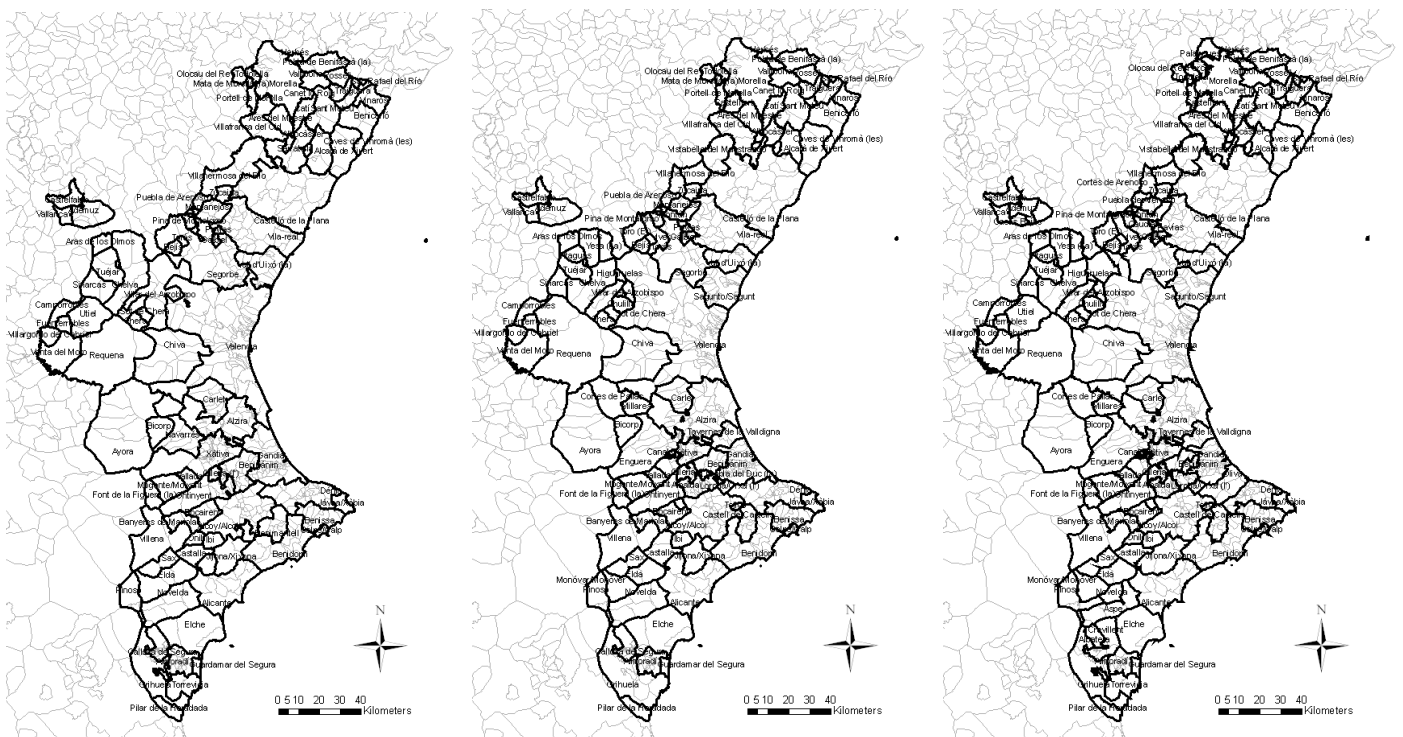

Fig. 9. Regionalization of the Region of Valencia without minimum size constraints applying TTWAs method (left); and our proposal: f_1 (center) and f_2 (right).
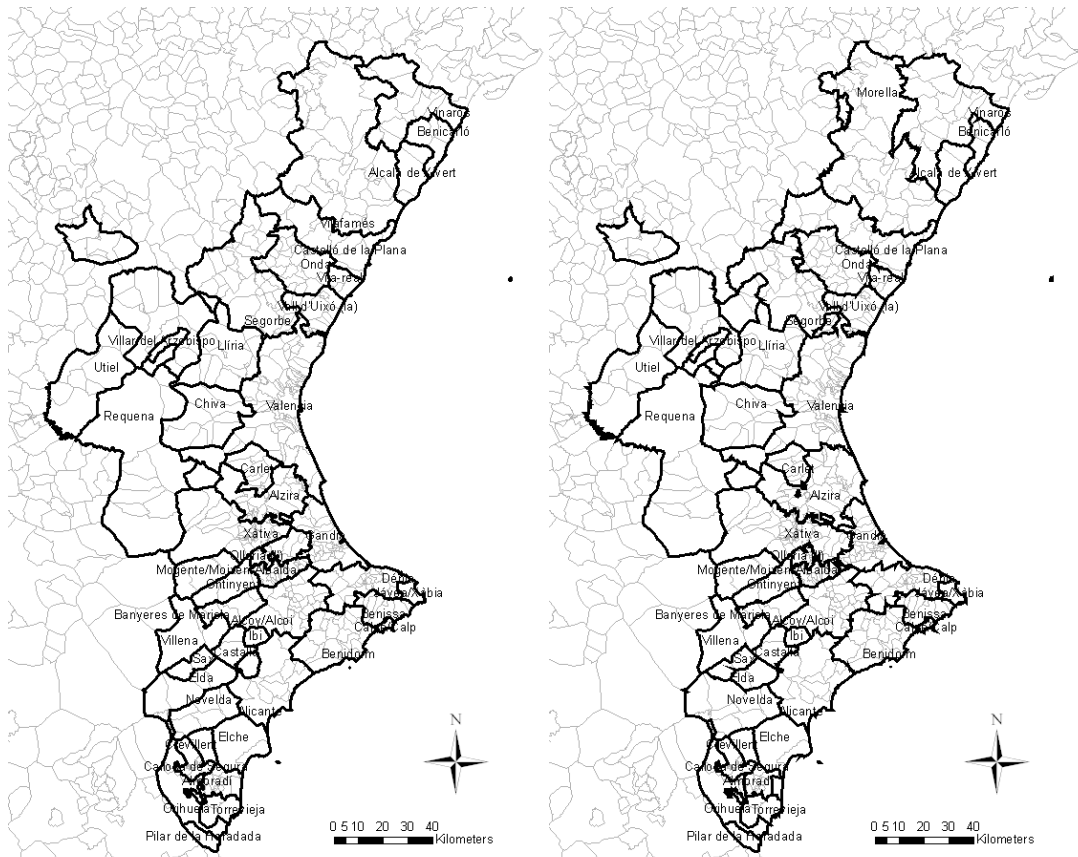
Fig. 10. Optimization with minimum size constraints (right) of the results from the TTWAs method (left) with our evolutionary approach.
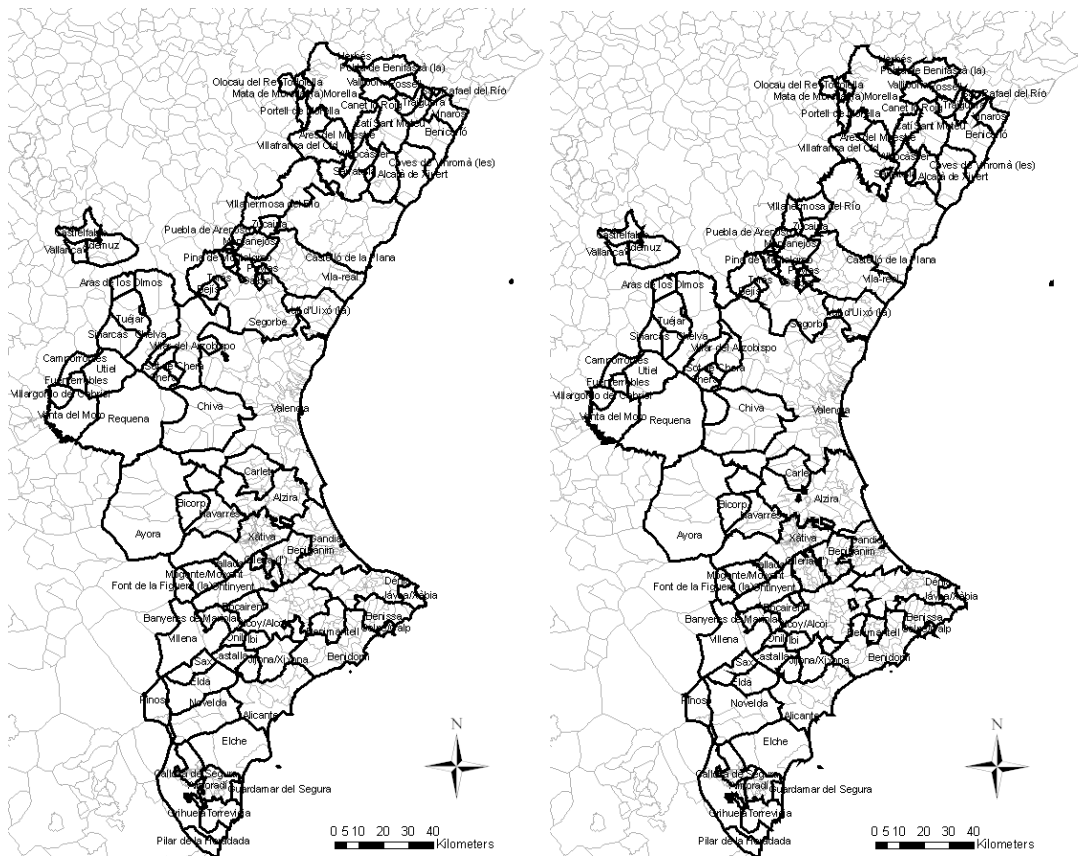

Fig. 11. Optimization without minimum size constraints (right) of the results from the TTWAs method (left) with our evolutionary approach.