

A multi-objective proposal for the aggregation of economically inactive population

María Beatríz Bernábe Loranca^{1,2}, David Pinto Avendaño², Jorge A. Ruiz-Vanoye¹, José Espinosa Rosales², Elías Olivares Benitez¹

¹ *Universidad Popular Autónoma del Estado de Puebla UPAEP.*

² *Benemérita Universidad Autónoma de Puebla BUAP.*

beatriz.bernabe@gmail.com

Abstract. As part of the process to support decision making in population problems, with the aim of promoting the selection of projects with funding support for the sector of the economically inactive population, in this paper a custom multi-objective method was applied to produce a set of efficient solutions that correspond to the selection of the best grouping of geographical zones to attend those population sectors without economic activity. Each answer is a partition of geographic zones that meet geometric compactness and homogeneity in terms of the population descriptors of the economically inactive population.

This work presents the basic theoretical aspects of the multi-objective methodology and an outline of a mechanism to analyze their relevance to support the selection of economic projects in the case of the Metropolitan Zone of the Toluca Valley (ZMVT), considering census socioeconomic data for AGEBS (Basic Geostatistical Areas).

Keywords: We would like to encourage you to list your keywords in this section.

1 Introduction

The National Institute of Statistics and Geography, INEGI is a self-governing agency of the Mexican government in charge of performing the population census each ten years. The data from the census is useful in the decision-making process to solve diverse problems of structuring, planning and organizing population. However, those problems are constrained by current policies, available resources and approved social development planning from the government. Ordinarily, the problems involving population data are directly related to territorial-type problem, whose main principle is the analysis of small and strategic groupings according to restrictions that describe the population. Such groupings expedite the job of decision makers by having a number of adequate groups defined. Furthermore, the information contained in the groupings assists the analysis of variability and to relate and locate the data.

The grouping of AGEBS was created applying a Pareto Front [9]. The groups are compact and homogeneous and are well characterized according to census indicators of the economically inactive population. If a project is approved to propose an assistance program for the economically inactive population, the decision maker has diverse options for grouping and partitioning that segment of the population with the purpose of being able to analyze and to propose solution arrangements according to the available resources. The resulting groupings observed in this research have been the product of applying a multi-objective method that was used in previous cases [2].

The problems known as multi-objective are those that deal with the existence of multiple, conflicting criteria. This implies the existence of different solutions for the problem, where a decision must be made based on a series of opposing criteria. The decision making process for solving the problem must create a set of feasible points, i.e., the corresponding set determines the restrictions of the problem by associating a degree of desirability to each alternative, criterion or objective. These possible solutions are those that satisfy the restrictions and preferences, which are executed over the proposed objectives.

The method proposed to obtain the solution set is based on basic principles of order theory, the properties of partial order and non-comparable orders, with the aim of obtaining a set of efficient and optimal solutions that form the Pareto Front. The solutions are interpreted as territorial groups (partitions) characterized as: geometrically compact and homogeneous population-wise. The geographic objects that conform the groups are AGEBS from the INEGI 2000 [8]; population and household census, where the census data for the ZMVT was taken as a source for this paper.

Received Oct 10, 2011 / Accepted Dec 31, 2011

The present work is organized as follows: section 1 contains the introduction to the research, in section 2 the relevant aspects of the multi-objective theory are discussed, section 3 describes the proposed multi-objective methodology, section 4 presents a solution using the proposed multi-objective method to the socio-economic problem of the ZMVT and lastly, section 5 contains the conclusions.

2 Multi-objective programming and Pareto Order

A multi-objective programming problem matches the following formulation

$$(1) \text{OPT} \left(f_1(x), f_2(x), \dots, f_p(x) \right) \text{ subject to } x \in X$$

where $x = (x_1, \dots, x_n)$ are decision variables, X is the set of solutions, f_i are each one of the objectives and $f = (f_1, f_2, \dots, f_p)$ is the vector function with $Y = f(X)$ being the objective space or image space.

To perform comparisons between vectors the following orders were used, where the relation was established noting the value that the objectives take x' at such points.

1*) Pareto Order: Where a point x is preferred to one x' if it is verified that

$$(2) f_j(x) \geq f_j(x') \forall j=1, \dots, p \text{ with at least one } j \text{ such that } f_j(x) > f_j(x')$$

2*) Weak Pareto Order: Where a point x is preferred to one x' if it is verified that:

$$(3) f_j(x) > f_j(x') \forall j=1, \dots, p$$

The Pareto order and the weak Pareto order are partial orders that ascertain that one combination will be preferred to another in the first case whenever it improves all objectives, improving one of them in strict form while in the second case, called weak, whenever all objectives are strictly improved [4].

In general, there will not exist a single combination that satisfies all objectives at the same time. In this way the first concept that is dropped is the optimum as understood in the traditional mono-objective programming, and the solutions sought for the problem will be the so-called *efficient solutions*. The generalization of an optimum gives place to the concept of efficiency, which can be defined for maximization, using the Pareto ordering of the following two forms [4].

Definition 1. A point x^* is efficient if it does not exist x such that x is preferred to x^* . If the Pareto order is used, x^* is efficient if it does not exist x such that

$$(4) f_j(x) \geq f_j(x^*) \forall j=1, \dots, p \text{ with at least one } j \text{ such that } f_j(x) > f_j(x^*)$$

Definition 1.1 A point x^* is not efficient if there exists x such that x is preferred to x^* . That is, x^* is not efficient if there exists x such that $f_j(x) \geq f_j(x^*)$ for some $j=1, \dots, p$

To find non-dominated and non-comparable solutions a variant of the Pareto order has been proposed and is described in section 4.

2.1 Problem description

The goal of this research is to obtain a partition of AGEB spatial data which composition is given by two components: geographical coordinates in the R^2 plane and a vector of descriptive characteristics from the census. The first component provides a distance matrix as input to the process of computing the geometric compactness (one of the objective functions to minimize). The description vector is used to optimize the second objective function, which is the homogeneity of one of the census variables of particular interest. The variable selection will be performed over the data that corresponds to the economically inactive population [2].

According to the INEGI, the economically inactive (or non-active) population (PEI) includes people of 12 years of age or more that in the reference week did not participate in economic activities, nor was part of the unemployed population. PEI is classified as:

- **Available and economically inactive population.** People of 12 years of age or more that did not work nor had a job and were not actively looking for one, due to disenchantment or because they think that no job will be given to them due to their age or lack of studies, among other causes. However, they are willing to accept a job offered to them although they are not actively looking for one.

- **Not available and economically inactive population.** People of 12 years of age or more that did not work nor had a job, are not actively looking for one, and would not be willing to accept one if it was offered to them. That is, people not available to enter the job market because they stay at home or are students, retired, pensioned, physically unable to work, or belong to other inactive groups such as those voluntarily idle or with addiction problems.

Assuming there exist an assistance program for this sector of the population, groupings of zones of economically inactive population are required to analyze and thus propose specific assistance for each grouping. The partition sought is a set of classes where the elements of each class are geographically close and the AGEs (or groups) involved are characterized by variables of non-economic activities and furthermore are balanced by the census variable “Economically inactive population”.

The problem is clearly bi-objective of geographical partitioning. Given the complexity of the problem, it is necessary to use heuristic methods to achieve approximated solutions and the implementation of multi-objective tools to find a set of non-dominated solution pairs. The variable neighborhood search (VNS) was chosen to find efficient solutions to both objective functions [6].

3 Quantization and representation

There exists a physical search space for geographical grouping. The AGEs geographical units are finite, that is, each element is represented by its spatial location and by an array of descriptive variables. The problem is discrete, combinatorial and binary-integer and the groupings are performed under the properties of partitioning.

To achieve compactness, the grouping is solved in such a way that the AGEs composing a group are geographically close, using as the objective function to minimize the sum of distance between AGEs. On the homogeneousness side, the optimization is carried out searching for equilibrium among a census variable of interest. Once the minimizing distance grouping has been done, the homogeneousness of each group is calculated, given that in multi-objective problems the function to optimize has the same dominion for all objectives. In this way the compactness and homogeneousness are optimized over the same partition. Next, the best alternative with respect to the m objectives is chosen. Mathematically, there exists a set X that is a subset of the R^n space such that $f_i: X \rightarrow R_n, i=1, \dots, m$, where m is the number of objectives [12].

3.1 Model

Let AGE be a spatial datum defined by its components in space and a description of census variables given by a vector. In the following definitions: M = territory map, T = territory, DE = census data, MS = dissimilitude matrix, UGB = basic geographic unit (AGEB), GT = territory group, m = GT number, n = UGB number with $m < n$, I = territory group index, j = basic geographic unit index, C_j = centroid of the i th territory group.

The model in question is mixed integer and uses the binary variables for models of this kind. The proposed formulation for the selection of the census variables and its boundaries is as follows:

VA_{kj} is the value of the k -th attribute contained in the j th UGB $_j$

α_k, β_k tolerance parameters for VA_k in any UGB

$(VA_{ki}) = \sum_{j=1}^n VA_{kj} X_{ij}$ is the value for the k th GT

$\frac{X_{ij}}{VA_k} = i / m \sum_{j=1}^n VA_k$ if the goal value for the k th attribute in any UGB

$d_{ij} = d(C_j, UGB) X_{ij}$ is the distance from the j th UGB in the i -th GT to its centroid

Thus the computation of the distance between AGEs can be expressed as:

$$a) D_{ij} = \sum_{j=1}^n d(C_j, UGB) X_{ij}$$

And the value of homogeneousness for variables is expressed as

$$b) H_{AGEB-var\ iable} = \sum_{j=1}^m (\overline{VA}_k - VA_{kj})$$

Considering the preceding, a proposal for the model is:

$$c) \text{Minimize } y = f(x) = (A, Z)$$

where

A : is the cost of minimizing the distance between AGEBs according to equation a) which must be formulated as a function, and
 Z : is the cost of minimizing the homogeneousness of an AGEB census variable. This function must be expressed from equation b.

The formulation of functions a) and b) will be presented later.

The functions f1 and f2 presented in c) are subject to:

$$GT_i \neq \emptyset \text{ for } i=1, \dots, K \text{ (the groups are non empty)}$$

$$GT_i \cap GT_j = \emptyset \text{ for } i \neq j \text{ (there are not repeated AGEBs in different groups)}$$

$$\bigcup_{i=1}^n GT_i = UGB \text{ (the union of all groups is contains all AGEBs)}$$

$$\alpha_k \leq VA \leq \beta_k \text{ (bounds for the variables)}$$

$$\sum_{j=1}^m X_{ij} = 1 \text{ is the AGEBs assignment}$$

$X_{ij} = 1$ if $UGB \in GT_i$ or $X_{ij} = 0$ if $UGB \notin GT_i$ are the decision variables

$y = (y_1, y_2) \in Y \subset R^2$ is the objective vector.

Given that a) and b) are the implicit equations en the objective functions, the goal now is to present both as functions to be optimized during the partitioning process. The optimization heuristic being variable neighborhood search (VNS), it is convenient to briefly describe its role in the method being developed. The VNS heuristic finds a random initial solution (current solution).

This solution is a partition where compactness and homogeneousness have been minimized producing a pair (c_i, h_i) . The next solution is generated (c_{i+1}, h_{i+1}) and compared with the previous one according to the Pareto non-comparable order. If this solution (c_{i+1}, h_{i+1}) is non-comparable with respect to (c_i, h_i) it is labeled as a "suboptimal" solution and takes the place of the current solution to be compared with the next. This process is repeated until the parameters of VNS allow it, thus obtaining the set of non-dominated solutions that form the Pareto Front [2].

The non-comparable Pareto order draws from the Pareto order to generate non-comparable and non-dominated solutions [2, 9].

That is, a Pareto order implies:

Given a solution (a, b) the next solution (a', b') is accepted if and only if

$$(1) (a' > a \wedge b' = b) \vee (b' > b \wedge a' = a) \vee (a' > a \wedge b' > b) \vee (a' = a \wedge b' = b)$$

When the comparison of the pair of solutions is performed through an expression, only one point from the Pareto frontier is reached.

The negation of expression (1) enables the production of approximations to the Pareto frontier by several chains. However, it is also necessary to iteratively examine whether the non-comparable solutions satisfy the Pareto dominancy. Lastly, the solution set thus obtained is a set of non-dominated and non-comparable solutions (Pareto frontier). A pair of solutions is not comparable in a determined partial order if it does not fulfill the trichotomy property, which in this case means that tow pairs (a, b) and (a', b') are Pareto non-comparable if

$$\neg((a, b) < (a', b')) \wedge \neg((a', b') < (a, b)), \text{ this is}$$

$$(2) (a > a' \vee b > b') \wedge (a' > a \vee b' > b)$$

Under this strict partial order called Pareto non-comparable, conveniently combined with the Pareto order, a set of non-dominated solutions is obtained.

To ensure that the solutions thus obtained are form a Pareto Front, the application Nodom has been used, which finds minimal solutions from a given set of solution pairs (Nodom 2007) [11]. This way the generated solutions generated by the method

presented are compared against those found by Nodom. Note that the minimals are a set of non-comparable and non-dominated points in a Hasse diagram.

These could be presented in tabular or graph form, with appropriate statistical evaluation, discussion of results, statement of conclusions drawn from the work.

4 Application of a multi-objective method to a socio-economical problem

There are several aspects to be considered when a project proposal is made regarding economical and social aspects of the population, some of them are: organizational structure, project plan, budget evaluation, among others. This type of problems requires groupings of zones that meet the problem restrictions. The chief point to focus on is to offer a set of groups of economically inactive population that reflects different scenarios. These groups are composed of spatial objects known as AGEBS. A vector of variables that originate from census variables constitutes each AGEBS.

An important decision to make is whether the assignment of resources to a sector of the vulnerable population is viable. The economically inactive population has been selected as a case study. Assuming that there exist a government program to sponsor assistance programs for this sector of the population, a set of intelligent groupings is required to expose the distribution of the population in terms of the following census variables, which naturally are strongly correlated [2].

- Population 12 years or older economically inactive whom are students
- Unemployed population
- Population 12 years or older economically inactive and dedicated to housework
- Economically inactive population

4.1 Application description

The model exposed so far has been applied to a socio-economical problem, where it was assumed that it requires 8 compact groups where AGEBS are located very near each other in order to expedite travel. The groups must also be integrated by the four economically inactive population variables. Homogeneous is maintained in this regard.

The Pareto Fronts for six tests of eight groups each is shown next. The values for VNS are 15 for local search and 2 for neighborhood structure. The value of homogeneousness was maintained stable for the six tests and the response time oscillated between 311 and 315 seconds. For all tests the solutions generated by the proposed method are shown in blue while the ones generated by *Nodom* are identified in red. In each figure it can be clearly seen an intersection between both methods, which ensures that the method proposed in this paper generates all the non-dominated solutions and several additional ones, which in turn implies a revision to the method exposed here in order to detect the reason for its presence in the Pareto Front. However, the additional solutions are very near to the Pareto Front, which may favor the selection of alternatives when taking multi-criteria decisions.

In each graph the \mathcal{X} -axis represents the compactness against the homogeneousness on the \mathcal{Y} -axis. If the decision makers are interested in sacrificing the compactness or homogeneousness they should choose the solution among the points shown. The values in each table associated with the graph are the costs of the minimum value for both objectives. The decision makers may not be interested in these values since they only exhibit the cost corresponding to the mathematical computation of the objective function. However, if it desired to have an inflection point for compactness or homogeneousness then for interpretation purposes it is pertinent to have a map that describes the solution (this map can be seen together with the graph or the Pareto Front).

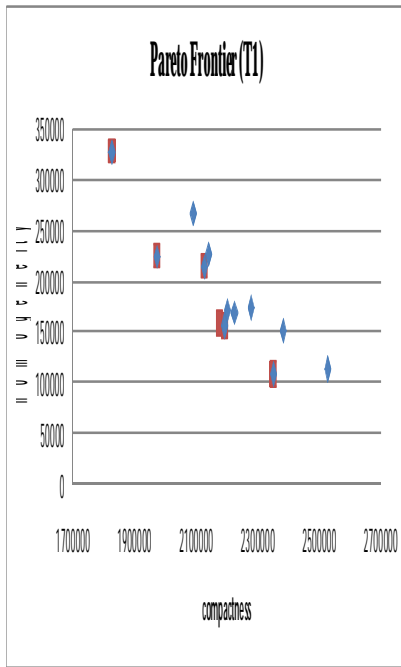
The map is a diagram obtained by a MapX interface (MapX 1982 [11]) and [14]. On the other hand, if a more detailed work is required, it is necessary to “move” the diagram to a Geographic Information System (GIS) to work over a layer of census population and know in detail the distribution of information, which is a topic for future research.

Each Pareto Front designates a set of non-dominated, non-comparable solutions; for each one the initial solution (IS) was random.

For test 1, the IS was 2554725 in compactness and 290759 in homogeneousness. In tests 2 to 6 the values for the IS in compactness and homogeneousness were respectively: 2:(2308354, 225704); 3:(2557228, 170824); 4:(2770870, 10412); 5:(2777916, 286505), and 6:(2738487, 150301).

4.2 Pareto Fronts (PF) for six tests over four variables about economic inactivity and geographic compactness for AGEBs

Test 1 (T1)



Graph 1. PF for T1

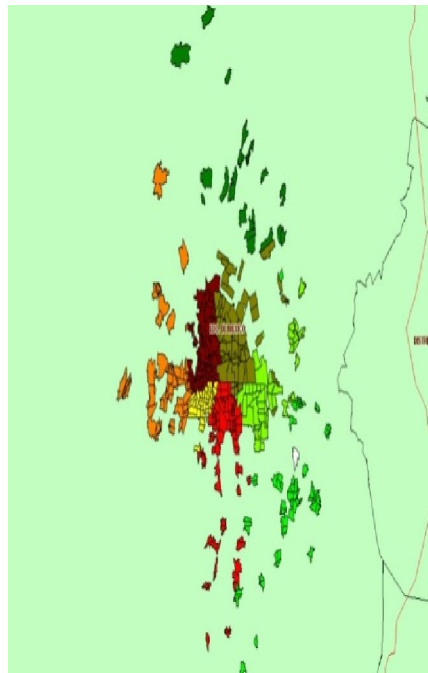
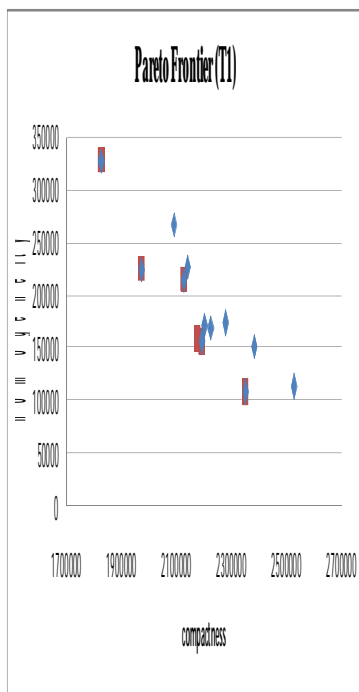


Fig. 1. Map associated to the PF for T1

compactness	homogeneity
2201586	171330
2383132	150992
2090895	266770
2278380	173204.5
2126642	213928.5
2528561	112620
1826806	327043
2225974	169870.5
2193505	155127
2141082	227650.5
1972880	224382.5

Table 1. 11 Solutions

Test 2 (T2)



Graph 2. PF for T2

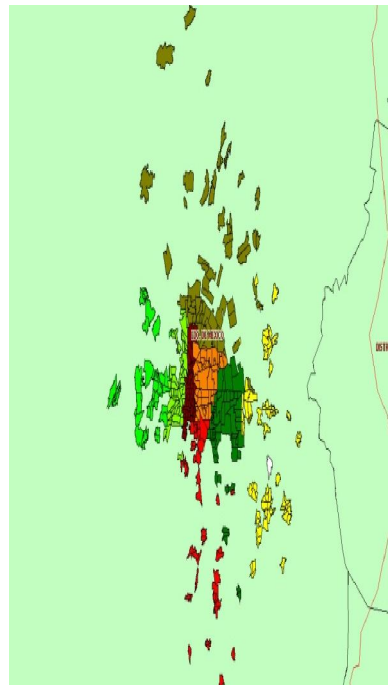
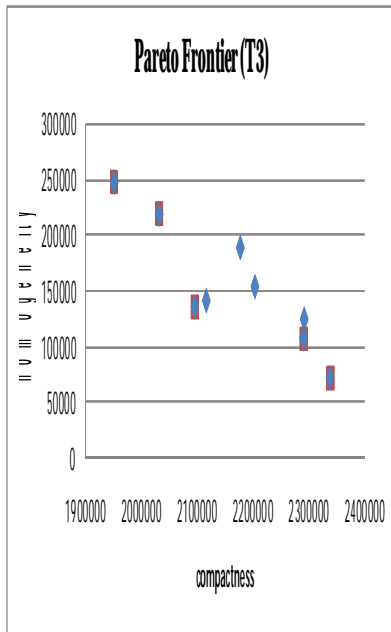


Fig. 2. Map associated to the FP for T2

compactness	homogeneity
2478504	95432
2175403	146120
2016134	156139
2002053	250353
2169142	155797
2629265	80696
2144708	167558
2183735	131310
2337024	70078
2115412	182776
2311345	120129

Table 2. 11 Solutions

Test 3 (T3)



Graph 3. PF for T3

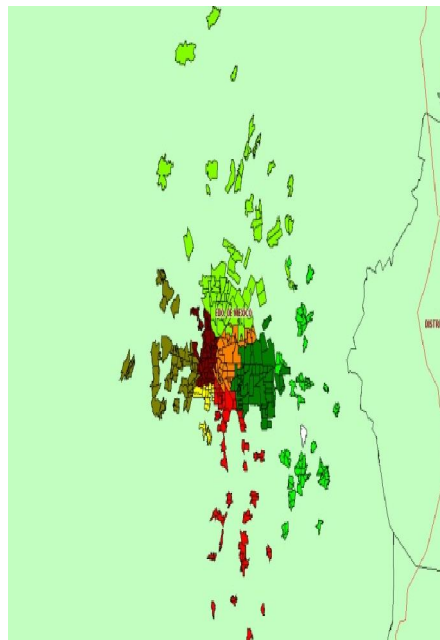
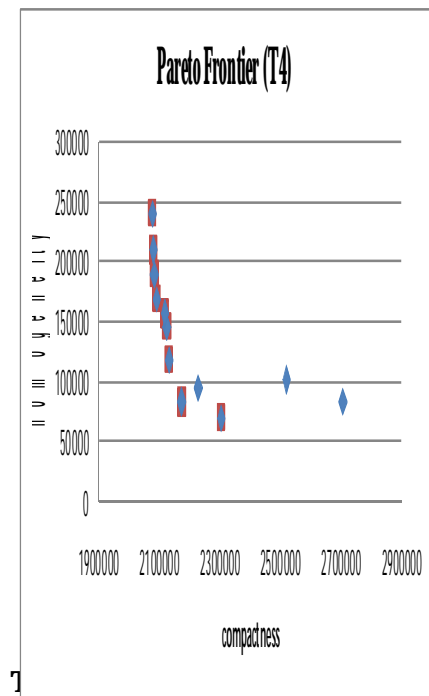


Fig. 3. Map associated to the PF for T3 **Table 3.** 8 Solutions

compactness	homogeneity
2290551	107211
2337873	72143
2203225	153745
2291454	124653
2177142	188700
2095605	135723
1951097	247358
2031878	218777

Test 4 (T4)



Graph 4. PF for T4

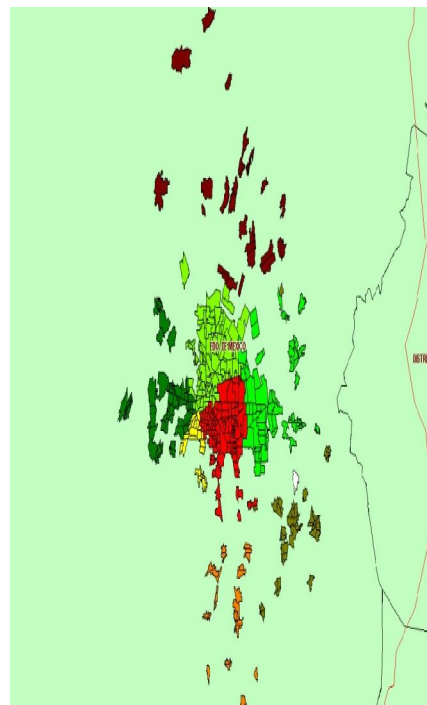
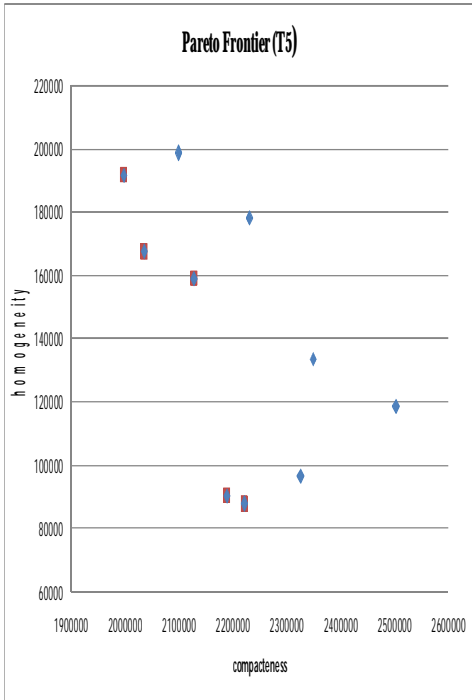


Fig. 4. Map associated to the PF for T4 **Table 4.** 11 Solutions

compactness	homogeneity
2130871	117075
2078086	209036
2076279	240148
2518534	100088
2089252	168654
2228036	95155
2706083	83106
2123688	145806
2082765	189411
2173516	81880
2116561	155758



Graph 5. PF for T5

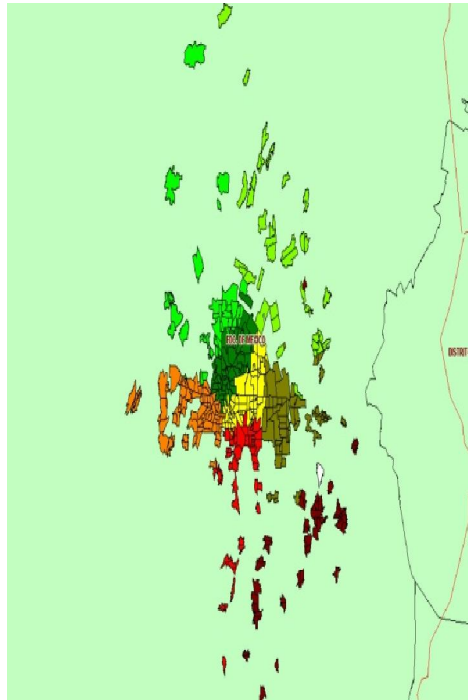
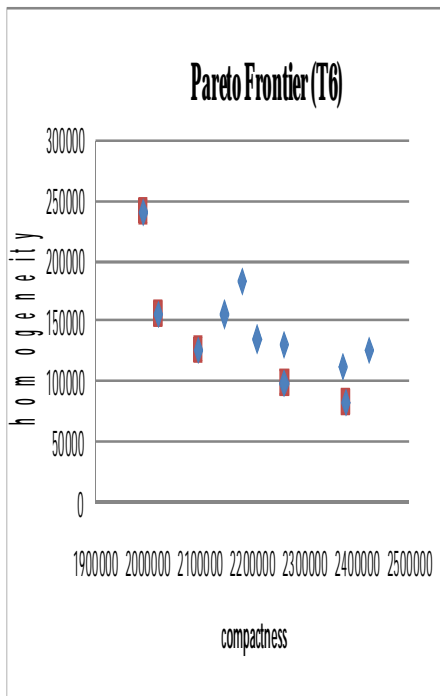


Fig. 5. Map associated to the PF for T5

compactness	homogeneity
2348532	133452
2098661	198669
2325588	96706
2126979	159006
2034534	167489
1996625	191661
2229619	178216
2220882	87682
2502750	118695

Table 5. 11 Solutions

Test 6 (T6):



Graph 6. PF for T6

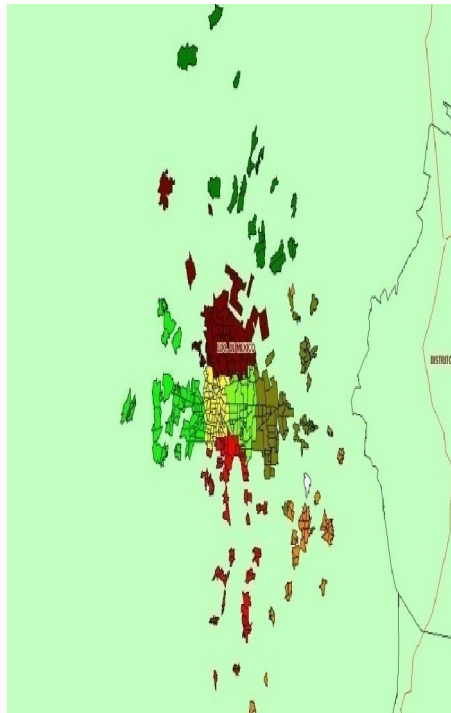


Fig. 6. Map associated to the PF for T6

compactness	homogeneity
2018144	156066
2422432	125965
2095144	126709
2259491	131332
2179832	183214
2371782	111986
2208783	134218
2260711	99255
2377838	82245
1989526	241205

Table 6. 11 Solutions

References

1. Altman, M.: The Computational Complexity of Automated Redistricting: Is Automation the Answer?. Rutgers Computer and Technology Law Journal. 23, (1) 81-142 (1997)
2. Bernábe, B.: Diseño y desarrollo de un modelo para la zonificación óptima. Tesis Doctoral, Universidad Nacional Autónoma de México, Posgrado de Ingeniería (2000).
3. Bernábe B., López, R.: Proceso de Análisis de Correlaciones para Identificar Patrones de Asociación Sobre Datos Censales. Research on Computing Science, CIC – IPN (2005).
4. Caballero, R., Gomez, T., Gonzalez, M., Muñoz, M., Rey, M., Ruiz, F.: Programación Matemática para Economistas. Universidad de Málaga, ISBN 84-7496-651-5, Málaga (1997).
5. Kung, H., Luccio, F., Preparata, F.: On Finding the Maxima of a Set of Vectors. Journal of the ACM JACM, 22, (4) 469-476 (1975).
6. Hansen P., Mladenovic, N.: Variable neighbourhood search. In Fred Glover and Gary A. Kochenberger editors, Handbook of Metaheuristics, Kluwer (2003).
7. Ingaramo, D., Pinto, D., Rosso, P., Errecalde, M.: Evaluation of Internal Validity Measures in Short-Text Corpora. Proceedings de la conferenciaCICLing (2008).
8. INEGI.: Sistema para la consulta de información censal 2000, (SCINCE), XII Censo General de Población y Vivienda 2000. Retrieved from http://www.inegi.org.mx/prod_serv/contenidos/espanol/catalogo/Default.asp?accion=2&upc=702825496371.
9. Lara, L.: Un estudio de las Estrategias Evolutivas para problemas Multiobjetivo, Tesis de Maestría en Ciencias en la especialidad de Ingeniería Eléctrica Opción Computación, Cinvestav IPN (2003).
10. MapX, <http://geospatialmethods.org/mapx/>
11. Nodom, <http://www.cs.cinvestav.mx/~emoobook/nodom/nonodom.html>
12. Ríos, D.: Sobre soluciones óptimas en problemas de optimización multiobjetivo. Trabajos de Investigación Operativa. Editor Springer Berlin/ Heidelberg, ISSN 0213-8204, 2 (1) 49-67 (2008).
13. Zamora, E.: Implementación de un algoritmo compacto y homogéneo para la clasificación de zonas geográficas AGEBS bajo una interfaz gráfica. Tesis de Ingeniería en Ciencias de la Computación, Facultad de Ciencias de la Computación, Benemérita Universidad Autónoma de Puebla (2008).