ABC, A Viable Algorithm for the Political Districting Problem

Eric-Alfredo Rincón-García, Miguel-Ángel Gutiérrez-Andrade, Sergio-Gerardo de-los-Cobos-Silva, Pedro Lara-Velázquez, Roman-Anselmo Mora-Gutiérrez and Antonin Ponsich

Abstract Since 2004, the Federal districting processes have been carried out using a Simulated Annealing based algorithm. However, in 2014, for the local districting of the state of México, a traditional Simulated Annealing technique and an Artificial Bee Colony based algorithm were proposed. Both algorithms used a weight aggregation function to manage the multi-objective nature of the problem, but the population based technique produced better solutions. In this paper, the same techniques are applied to six Mexican states, in order to compare the performance of both algorithms. Results show that the Artificial Bee Colony based algorithm is a viable option for this kind of problems.

Keywords Districting · Artificial bee colony · Simulated annealing · Heuristic

E.-A. Rincón-García (✉) · R.-A. Mora-Gutiérrez · A. Ponsich
Departamento de Sistemas, Universidad Autónoma Metropolitana - Azcapotzalco,
Av. San Pablo 180. Col, Reynosa Tamaulipas C.P 02200, Mexico D.F.
e-mail: rigaeral@correo.azc.uam.mx

R.-A. Mora-Gutiérrez
e-mail: mgra@correo.azc.uam.mx

A. Ponsich
e-mail: aspo@correo.azc.uam.mx

M.-Á. Gutiérrez-Andrade · S.-G. de-los-Cobos-Silva · P. Lara-Velázquez
Departamento de Ingeniería Eléctrica, Universidad Autónoma Metropolitana - Iztapalapa,
Av. San Rafael Atlixco 186. Col, Vicentina. del Iztapalapa C.P 09340, Mexico D.F.
e-mail: gamma@xanum.uam.mx

S.-G. de-los-Cobos-Silva
e-mail: cobos@xanum.uam.mx

P. Lara-Velázquez
e-mail: plara@xanum.uam.mx

J. Gil-Aluja et al. (eds.), *Scientific Methods for the Treatment of Uncertainty in Social Sciences*, Advances in Intelligent Systems and Computing 377,
DOI 10.1007/978-3-319-19704-3_22

1 Introduction

The zone design problem arises from the need of aggregating small geographical units (GUs) into regions, in such a way that one (or more) objective function(s) is (are) optimized and some constraints are satisfied. The constraints can include, for example, the construction of connected zones, with the same amount of population, clients, mass media, public services, etc. (Bernabé et al. 2012; Rincón-García et al. 2010; Duque et al. 2007; Kalcsics et al. 2005). The zone design is used in diverse problems like school redistricting (Ferland and Guenette 1990; Caro et al. 2001; DesJardins et al. 2006), police district (Dell'Amico et al. 2002), service and maintenance zones (Tavares-Pereira et al. 2007), sales territory (Zoltners and Shina 2005) and land use (Williams 2002).

The design of electoral zones or electoral districting is a well-known case, due to its influence in the results of electoral processes and its computational complexity, which has been shown to be NP-Hard (Altman 1997). In this framework, the GUs are grouped into a predetermined number of zones or districts, and democracy must be guaranteed through the satisfaction of restrictions that are imposed by law (Mehrotra et al. 1998; Ricca et al. 2011). In particular, some generally proposed criteria are population equality, to ensure the "one man one vote" principle; compactness, to avoid any unfair manipulation of the border or shape of electoral zones for political purposes, and contiguity, to prevent from designing fragmented districts.

Nowadays, different meta-heuristics have been reported in specialized literature to produce automated districting plans. The most common techniques include local search algorithms such as Simulated Annealing (SA), Tabu Search and Old Bachelor Acceptance (Macmillan 2001; Bozkaya et al. 2003; Ricca and Simeone 2008, Rincón-García et al. 2013). Although some Genetic (Bacao et al. 2005), Evolutionary (Chung-I 2011) and Swarm Intelligence (Rincón-García et al. 2012) algorithms have been applied.

In Mexico, the Federal Electoral Institute[1] (IFE), for the federal districting processes in 2004 and 2013, used a SA based algorithm. However, in 2014, for the local districting problem of the State of Mexico, two automated districting algorithms were proposed: a Simulated Annealing based algorithm, and, for the first time in Mexico, a population based technique was used: an Artificial Bee Colony (ABC) algorithm.

The primary purpose of this paper is to compare the performance of these algorithms in different instances. To address this issue, we provide a description of the problem in Sect. 2 a brief overview of the inner working mode of the SA and ABC algorithms is presented in Sect. 3. Some computational results are detailed in Sect. 4. Finally, conclusions and perspectives for future work are drawn in Sect. 5.

[1]The National Electoral Institute, INE, since April 4, 2014.

2 Problem Description

As mentioned previously, population equality and compactness are important principles that should be promoted in the design of electoral districts. For this reason, the objective functions should guide the search towards regular shaped districts with approximately the same amount of population.

In order to promote the principle "one man one vote", different measures to quantify the population equality have been proposed, for example the difference (or ratio) between the most and the least populated zone, or the sum of the absolute values of the difference between the average and the real number of inhabitants in each zone. In particular, in 2014, the Electoral Institute of the State of Mexico (IEEM) proposed the following measure:

$$C_1(P) = \sum_{s \in S} \left(1 - \frac{P_s}{(P_T/n)}\right)^2 \left(\frac{1}{0.15}\right)^2 \tag{1}$$

where $P = \{Z_1, Z_2, \ldots, Z_n\}$ is a districting plan. Each district Z_s is defined through a set of binary variables x_{is} such that $x_{is} = 1$ if the *i-th* GU belongs to district s and $x_{is} = 0$ otherwise. P_T is the population of the considered state, P_s is the population of district s, 0.15 is the maximum percentage of deviation allowed for the state, $S = \{1, 2, 3, \ldots, n\}$, n is the number of electoral districts that must be generated in the state. Thus, the lower the cost C_1, the better the population equality of a solution. Indeed, the perfect population equality is achieved when all districts have the same number of inhabitants, and in this case the measure assigns a value of zero to C_1.

Regarding the assessment of district compactness, several measures have been proposed, as observed in (Niemi et al. 1990). The IEEM introduced a metric that can be easily computed, in order to improve the runtime performance. This measure of compactness compares the perimeter of each district with that of a square having the same area.

$$C_2(P) = \sum_{s \in S} \left(1 - \frac{PC_s}{\sqrt{AC_s}} - 1\right) \tag{2}$$

where PC_s and AC_s are the perimeter and the area of the considered district s, respectively. Thus, districts with a good compactness will have a compactness value close to 0.

Finally, in order to handle the multi-objective nature of the problem, the IEEM used a weight aggregation function strategy.

$$Minimize\, f(P) = \lambda_1 C_1(P) + \lambda_2 C_2(P) \tag{3}$$

where, λ_1, λ_2, are weighting factors that set the relative importance of equality population and compactness criteria.

This problem formulation therefore seeks for a districting plan that represents the best balance between population equality and compactness, a balance obviously biased by the weighting factors. In 2014 the IEEM proposed $\lambda_1 = 1$, and $\lambda_2 = 0.5$. However, these factors can be modified due to political agreements, thus it is impossible to predict the factors that will be used in future processes. Therefore it has been decided, in the framework of the present work, to produce a set of efficient solutions, using different weighting factors for the computational experiments.

3 Heuristic Algorithms

Since the design of electoral zones is an NP-Hard problem, automated heuristic algorithms are an appropriate strategy to design electoral districting plans. In Mexico, since 2004, a SA strategy has been applied to carry out the federal districting processes. However, new heuristic techniques have proven to get better solutions in these kinds of problems. Whereby the IEEM decided to propose a new strategy, based on Artificial Bee Colony, for districting the state of Mexico. This strategy also includes, nevertheless, a traditional SA algorithm, in order to produce solutions of a quality at least as good as those proposed in last processes.

In this section we provide a brief description of the Simulated Annealing and Artificial Bee Colony strategies used by the IEEM.

3.1 Simulated Annealing

Simulated Annealing is a metaheuristic introduced by Kirkpatrick in (Kirkpatrick et al. 1983). The SA algorithm starts with an initial solution P and generates, in each iteration, a random neighbor solution Q. If this neighbor improves the current value of the objective function, it is accepted as the current solution. If the neighbor solution does not improve the objective value, then it is accepted as the current solution according to a probability η based on the Metropolis criterion:

$$\eta = \exp\left(\frac{f(P) - f(Q)}{T}\right) \tag{4}$$

where $f(P)$ and $f(Q)$ represent the objective value of the current and neighbor solutions, respectively. T is a parameter called temperature, which is controlled through a cooling schedule that defines the temperature decrease and the (finite) number of iterations for each temperature value.

3.2 Simulated Annealing Adaptation

A classical implementation of SA was proposed by IEEM, with a geometric decreasing cooling schedule. The initial solution is created using the following strategy. All GUs are labeled as available. The algorithm then selects randomly n GUs, assigns them to different zones and labels them as not available. At this moment, each zone has therefore only one GU. Finally, each zone is iteratively extended by adding an available GU having a frontier with the zone in its current shape. Every time a GU is incorporated to a zone, it is labeled as not available in order to avoid the construction of overlapping zones. The latter step is performed until all the GUs are labeled as not available. This process ensures that the initial solution consists of n connected zones that include all GUs. Note that SA and ABC use the same procedure to create initial solutions.

Regarding now the construction of a neighbor solution, a random zone is chosen and a GU in this zone is moved to a neighbor zone. Therefore, the neighbor solution is identical to the current one, except that one GU is reassigned to an adjacent zone. The new solution is evaluated and accepted or rejected according to the Metropolis criterion. If the neighbor solution is rejected, another GU is randomly selected; this process is repeated until the temperature reaches a predefined lower bound.

3.3 Artificial Bee Colony

Artificial Bee Colony (ABC) is a bio-inspired metaheuristic, originally proposed by Karaboga and based on the natural behaviour of honey bees for finding food resources (Karaboga and Basturk 2007). Artificial bees are classified into three groups of bees: employed, onlookers, and scouts.

In the ABC algorithm, each solution to the problem under consideration is called a food source and represented by a D-dimensional real-valued vector. The fitness of a solution is associated to the amount of nectar in the food source. The algorithm cycle begins with an improvement phase, which is carried out by employed bees. For each solution $Y_i = [y_{i,1}, y_{i,2}, \ldots, y_{i,D}]$ in the population, a solution $Y_k = [y_{k,1}, y_{k,2}, \ldots, y_{k,D}]$ $(i \neq k)$ is randomly chosen to produce a new solution $V_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,D}]$ according to the following equation:

$$V_{i,j} = y_{i,j} + r^*(y_{i,j} - y_{k,j}) \quad (5)$$

where j is an index randomly generated in $\{1, \ldots, D\}$ and r is a uniformly distributed real random number in the range $(-1,1)$. Each produced solution V_i is subsequently evaluated by the employed bees and passes through a greedy selection process: if the new food source has a nectar amount equal to or better than the employed bee's current food source, it replaces by the current solution; otherwise, the old food source is kept.

When the employed searching phase is over, each onlooker bee evaluates the nectar information provided by the employed bees and chooses a food source according to a probability p_i, computed on the basis of its nectar amount (i.e., the corresponding fitness):

$$p_i = \frac{fit_i}{\sum_{j=1}^{M} fit_j} \tag{6}$$

where fit_i is the fitness value of the food source Y_i and M is the number of food sources. Once the food source is chosen, the onlooker tries to improve the corresponding solution generating a new one, in the same way the employed bees do, i.e. through Eq. (5). The new solution is then evaluated by the onlooker bee, its quality is compared with that of the current one and the best solution is selected applying a greedy selection process.

If a solution cannot be further improved through a predetermined number of trials, the solution is abandoned and a scout bee produces a new random solution. The termination criterion is, typically, a fixed number of computed iterations of the previously described cycle.

3.4 Artificial Bee Colony Adaptation

The ABC heuristic was originally designed for continuous optimization problems, and cannot directly be used for discrete cases. Thus, IEEM's algorithm included some modifications to the ABC method based on a recombination strategy.

First, M food sources are generated using the strategy described in Sect. 3.2. The number of onlooker and employed bees is set equal to the number of food sources, and exactly one employed bee is assigned to each food source.

According to Eq. (5), employed and onlooker bees generate new solutions by combining two food sources. However, after some experiments, the performance of the algorithm was improved when employed and onlooker bees used different strategies to explore the solution space. Each employed bee must apply a local search, similar to the strategy used by SA described in Sect. 3.2, while onlooker bees use a recombination technique inspired in Eq. (5).

First, each employed bee, i, modifies its food source, P_i, using the following strategy. A random zone is chosen and a GU in this zone is moved to a neighbour zone in such way that the new solution, V_i. If the new solution V_i has a nectar amount better than or equal to that of P_i, V_i replaces P_i and becomes a new food source exploited by the hive. In other case, V_i is rejected and P_i is preserved.

As soon as the employed bees' process has been completed, each onlooker bee chooses two solutions. The first solution, P_1, is selected depending on the probability given by (6), where the fitness value of the food source is given by (3).

The second solution, P_2, is randomly selected from the food sources exploited by the hive. A new food source, V_1, is produced through a recombination technique described straightforward.

A GU k is randomly selected. Thus, there is a zone Z_i ε P_1 and a zone Z_j ε P_2 such that k ε $Z_i \cap Z_j$. Let us now consider the following sets:

$$H_1 = \{l : x_{li} = 0, x_{lj} = 1\} \quad (7)$$

$$H_2 = \{l : x_{li} = 1, x_{lj} = 0\} \quad (8)$$

Then a GU in H_1 is inserted into Z_i, and a GU in H_2 are extracted from Z_i, and inserted into any randomly chosen zone contiguous to Z_i.

Note that these moves can produce a disconnection in zone Z_i, so that a repair process must be applied. The number of connected components in Z_i is counted after all the moves previously described. If the number of connected components equals 1, then the zone is connected. Otherwise, the algorithm defines the connected component that includes GU k (i.e., the GU used within the above-described recombination strategy) as zone $Z_{i,}$ subsequently, the remaining components are assigned to other adjacent zones.

4 Computational Experiments

The two algorithms described in the previous sections were tested on six Mexican states: Distrito Federal, Guerrero, Jalisco, Mexico, Michoacán and Nuevo León. For each state, the number of inhabitants, sections and zones to be created are presented in Table 1. We must remark that some sections are usually grouped in order to decrease the number of variables, and to reduce the complexity of the districting process. However, we decided that the performance of both algorithms should be tested in difficult problems, thus for these computational experiments the sections were kept separated.

In agreement with the Federal requirements stipulated for México's 2006 and 2013 elections, a maximum percentage of deviation (in terms of population in each district) d = 15 % was considered.

Table 1 Data for the six tackled instances

State	Inhabitants	Geographic units	Number of zones
Distrito Federal	8,605,239	5,535	27
Guerrero	3,079,649	2,784	9
Jalisco	6,322,002	3,326	19
México	13,096,686	5,930	40
Michoacán	3,985,667	2,677	12
Nuevo León	3,834,141	2,135	12

Regarding the settings of parameters shared by the considered techniques, the set $L = \{\lambda_1, \lambda_2\}$ of weighting factors is defined as follows: $\lambda_1 = \{0.99, 0.98, 0.97, \ldots, 0.02, 0.01\}$, while λ_2 is set to $\lambda_2 = 1 - \lambda_1$. Besides, in order to deal with the stochastic effect inherent to heuristic techniques, 990 independent executions were performed for each algorithm on each of the 6 instances (10 runs for each weight couple (λ_1, λ_2)). Each run produces a single solution and the resulting 990 solutions are subsequently filtered through a Pareto sorting procedure, which identifies the final non-dominated solutions.

The approximated Pareto fronts obtained by the algorithms, in Jalisco and Michoacán, are illustrated in Fig. 1. These fronts were obtained combining the solutions produced by the various executions of each algorithm. From the graphical observation of the sets of non-dominated solutions, it seems that ABC provides better solutions, in terms of convergence. However, these observations need to be confirmed by a performance metric.

Due to the multi-objective nature of the problem, we decided to compare the algorithms using the front coverage measure.

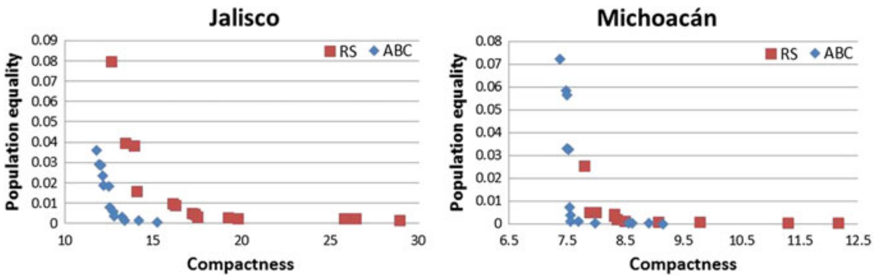

Fig. 1 Non-dominated solutions for the tested algorithms

The front coverage $C(A_1; A_2)$ is a binary metric that computes the ratio of efficient solutions produced by one algorithm A_2 dominated by or equal to at least one efficient solution produced by another competing algorithm A_1. Note that, commonly, $C(A_1; A_2) \neq C(A_2; A_1)$, so that both values must be computed. The expression of $C(A_1; A_2)$ is provided straightforward:

$$C(A_1; A_2) = \frac{|s_2 \in PF_2; \exists s_1 \in PF_1 : s_1 \succ s_2|}{|PF_2|} \tag{20}$$

where PF_1 and PF_2 are the approximated Pareto sets obtained by algorithms A_1 and A_2 respectively. If $C(A_1; A_2)$ is equal to 1, then all the efficient solutions produced by A_2 are dominated by efficient solutions produced by A_1.

The front coverage metric results are presented in Table 2 for all instances. The ABC algorithm seems to be the best option since its solutions dominate in a great extent those produced by SA; simultaneously, a low percentage of its solutions are dominated. For example, all SA solutions are dominated in Jalisco, México, Michoacán and Nuevo León, while only the 38.89 % of ABC solutions are dominated in Distrito Federal.

Table 2 Front coverage metric

Distrito Federal		SA	ABC	México		SA	ABC
A1	SA	–	0.3889	A1	SA	–	0.0
	ABC	0.6364	–		ABC	1.0	–
Guerrero		SA	ABC	Michoacán		SA	ABC
A1	SA	–	0.0	A1	SA	–	0.0
	ABC	0.7273	–		ABC	1.0	–
Jalisco		SA	ABC	Nuevo León		SA	ABC
A1	SA	–	0.0	A1	SA	–	0.0
	ABC	1.0	–		ABC	1.0	–

5 Conclusions

In this paper, the performance level of an ABC based algorithm, previously developed for the districting process of the state of México in 2014, is studied. In addition, a SA based technique, traditionally used by IFE, is employed as a reference to evaluate the population-based algorithm. The respective performances of both algorithms were evaluated in terms of the quality of the approximated Pareto front and of efficiency. The computational experiments proved that the ABC algorithm produces better quality efficient solutions than its counterpart.

Despite the computational experiments were carried out over six instances, it is likely that the overall performance level of the ABC algorithm can be meaningfully generalized to this formulation of the districting problem.

Finally, it is clear that this work can be enhanced focusing on some global guidelines such as implementing multiobjective versions of these algorithms to exploit the characteristics of this bi-dimensional problem. These guidelines provide perspectives for future work.

References

Altman, M.: Is automation the answer: the computational complexity of automated redistricting. Rutgers Comput. Law Technol. J. **23**, 81–141 (1997)

Bacao, F., Lobo, V., Painho, M.: Applying genetic algorithms to zone design. Soft. Comput. **9**, 341–348 (2005)

Bernabé, B., Coello-Coello, C.A., Osorio-Lama, M.: A multiobjective approach for the heuristic optimization of compactness and homogeneity in the optimal zoning. J. Appl.Res. Technol. **10** (3), 447–457 (2012)

Bozkaya, B., Erkut, E., Laporte, G.: A tabu search heuristic and adaptive memory procedure for political districting. Eur. J. Oper. Res. **144**, 12–26 (2003)

Caro, F., Shirabe, T., Guignard, M., Weintraub, A.: School redistricting: embedding GIS tools with integer programming. J. Oper. Res. Soc. **55**, 836–849 (2001)

Chung-I, C.: A Knowledge-based Evolution Algorithm approach to political districting problem. Comput. Phys. Commun. **182**(1), 209–212 (2011)

Dell'Amico, S., Wang, S., Batta, R., Rump, C.: A simulated annealing approach to police district design. Comput. Oper. Res. **29**, 667–684 (2002)

DesJardins, M., Bulka, B., Carr, R., Jordan, E., Rheingans, P.: Heuristic search and information visualization methods for school redistricting. AI Mag. **28**(3), 59–72 (2006)

Duque, J.C., Ramos, R., Suriñach, J.: Supervised regionalization methods: a survey. Int. Reg. Sci. Rev. **30**(3), 195–220 (2007)

Ferland, F., Guenette, G.: Decision support system for the school districting problem. Oper. Res. **38**, 15–21 (1990)

Kalcsics, J., Nickel, S., Schröder, M.: Towards a unified territorial design approach: applications, algorithms and GIS integration. TOP **13**(1), 1–74 (2005)

Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J. Glob. Optim. **39**, 459–471 (2007)

Kirkpatrick, S., Gellat, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**, 671–680 (1983)

Macmillan, W.: Redistricting in a GIS environment: an optimization algorithm using switching-points. J. Geogr. Syst. **3**, 167–180 (2001)

Mehrotra, A., Johnson, E.L., Nemhauser, G.L.: An optimization based heuristic for political districting. Manage. Sci. **44**(8), 1100–1114 (1998)

Niemi, R.G., Grofman, B., Carlucci, C., Hofeller, T.: Measuring compactness and the role of a compactness standard in a test for partisan and racial gerrymandering. J. Polit. **52**, 1155–1181 (1990)

Ricca, F., Simeone, B.: Local search algorithms for political districting. Eur. J. Oper. Res. **189**(3), 1409–1426 (2008)

Ricca, F., Scozzari, A., Simeone, B.: Political districting: From classical models to recent approaches. J. Oper. Res. **204**(1), 271–299 (2011)

Rincón-García, E.A., Gutiérrez-Andrade, M.A., Ramírez-Rodrígez J., Lara-Velázquez, P., de-los-Cobos-Silva, S.G.: Applying simulated annealing to design compact zones. Fuzzy Econ. Rev. **15**(2), 11–24 (2010)

Rincón-García, E.A., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., Lara-Velázquez, P., Mora-Gutiérrez, R.A., Ponsich, A.: A Discrete Particle Swarm Optimization Algorithm for Designing Electoral Zones. Methods for Decision Making in an Uncertain Environment. World Scientific Proceedings Series on Computer Engineering and Information Science, Reus, Spain, vol. 6, pp. 174–197 (2012)

Rincón-García, E.A., Gutiérrez-Andrade, M.A., de-los-Cobos-Silva, S.G., Lara-Velázquez, P., Mora-Gutiérrez, R.A., Ponsich, A.: A multiobjective algorithm for redistricting. J. Appl. Res. Technol. **11**(3), 324–330 (2013)

Shirabe, T.: A Model of contiguity for spatial unit allocation. Geogr. Anal. **37**, 2–16 (2005)

Tavares-Pereira, F., Rui, J., Mousseau, V., Roy, B.: Multiple criteria districting problems: The Public Transportation Network Pricing System of the Paris Region. Ann. Oper. Res. **154**, 69–92 (2007)

Williams, J.C.: A Zero-one programming model for contiguous land acquisition. Geogr. Anal. **34** (4), 330–349 (2002)

Zoltners, A.A., Sinha, P.: Sales territory design: thirty years of modeling and implementation. Market. Sci. **24**(3), 313–331 (2005)