

Simulated Annealing and Variable Neighborhood Search Hybrid Metaheuristic for the Geographic Clustering

¹María Beatriz Bernábe Loranca, ²David Pinto Avendaño, ³Elias Olivares Benitez, ^{4,6}Javier Ramírez Rodríguez, ⁵José Luis Martínez Flores

^{1,2}Benemérita Universidad Autónoma de Puebla, Puebla, Mexico

^{1,3,5}Universidad Popular Autónoma del Estado de Puebla, Puebla, Mexico

⁴Universidad Autónoma Metropolitana and LIA, Ciudad de Mexico, Mexico

⁶Université d'Avignon et des Pays de Vaucluse, France

beatriz.bernabe@gmail.com

Abstract

In this work we present a new hybrid approach for solving the clustering problem for geographic data, which is known to be NP-hard. Two metaheuristics that have proven efficiency in combinatorial optimization problems have been chosen for the comparison: Simulated Annealing (SA) and Variable Neighborhood Search (VNS). The proposed model is based on the partitioning around the medoids and on P-median. Previous test runs have shown satisfactory results (in terms of quality and time) for instances of 469 geographic objects, but when instances of greater size are used then variability in the results has been detected.

In an effort to achieve better results for the clustering problem, we have incorporated a hybridization of simulated annealing and variable neighborhood search to the geographic clustering problem. We have considered different sizes in the tests runs for distinct groups observing that the solutions obtained with the hybrid approach, named SA-VNS hybrid, overcome SA and VNS when they have been implemented individually.

Finally, with the aim of evaluating the benefits of the meta-heuristic proposed, we have measured the internal connection of the obtained clusters by means of the Dunn Index. The results obtained show that the hybrid SA-VNS performs better than SA and VNS with respect to the compactness feature.

Keywords: Clustering, Dunn, Simulated Annealing, Variable Neighborhood Search, Hybrid Metaheuristic.

1. Introduction

A big amount of combinatorial problems such as the clustering, imply that their solution could be obtained through heuristics techniques. However, once a metaheuristic has been implemented for a real clustering problem, it becomes necessary to compare the results with other methods including the statistical ones. In this process, is possible to distinguish if a determined metaheuristic is better than another one for certain specified factors. On the other hand, when the algorithm is tested using bigger in-

stances, the increment with regard to the response time and even the lower quality solutions is notable. In this point, is reasonable to assume that these instances demand the use of other alternative approximation methods that improve the quality of the solutions incorporating some improvements through different options, one of these is hybridizing 2 metaheuristics as we expose in this paper. The problem we deal with in this work is discrete, combinatorial and consists in grouping geographic data considering partitioning restrictions.

A clustering problem can be seen as an optimization problem that locates the optimal clusters' centroids instead of finding the optimum partition. The clustering algorithm that we employed in this work is partitional: given a database with n objects, a method by partitions builds k data groups, where each partition represents a cluster and $k \leq n$. This is, it classifies the data into k groups where each group must contain at least one element and each element must belong exclusively to one group. Achieving a global optimization of a clustering based on partitions would require an exhaustive enumeration of all the possible partitions. In this point, the importance of incorporating approximation methods for the problem we expose stands out, this problem forms groups of geographic data under partitioning around the medoids properties (Kaufman 1987) and P-Median (Resse 2005). The approximation proposal consists of the combination of the Variable Neighborhood Search (VNS) (Mladenovic 2004) and Simulated Annealing (SA) (Dowland 2003) metaheuristics integrated into an algorithm about partitions and the geographic clustering problem is applied with the objective to improve the results obtained by these two metaheuristics in a separate way. The results obtained with this hybrid proposal have been satisfactory when it has been applied to sets of geographic-spatial data. Finally, the quality of the solutions is tested by using the Dunn index (Dunn 1973, XiuLi 2011). The evidence given by this test revealed that the SA-VNS hybrid that we have implemented obtains better compactness than VNS and SA individually.

The remaining of this paper is organized as follows: Section 2 describes the partitioning as an optimization problem. In section 3 the basic aspects of VNS and SA

are treated. Section 4 presents the hybrid combination of SA-VNS and finally in section 5 we present the application of the Dunn method, which allows evaluating the clustering quality. Finally the conclusions are described in Section 6.

2. The Partitioning Problem in Combinatory Optimization

The classification techniques are of prominent usefulness in data analysis. Its main objective is to group objects into well differentiated classes or clusters in accordance to the dissimilarities between them and the best internal homogeneity possible, in such a way that the objects that belong to a cluster are closer to each other rather than to the ones in other groups. In this scenario, having a measure D of dissimilarity between objects becomes necessary, which can be expressed as a quality function. The traditional methods that look for a partition of a set of objects in a prefixed number of classes, only obtain partial solutions to the problem, locating barely some local optimums (Piza et al.1999). In the non-hierarchical methods the data set is partitioned into a pre-specified number of clusters k , and then the observations are reallocated to the clusters in an iterative way until a stopping criterion (function to optimize) is met. In this way, the partitioning methods possess the advantage of satisfying an optimality criterion even if it is approximately. On the other hand, the existent clustering methods differ from one another in the way of structuring the clusters. Those that find clusters that correspond to a partition of the set of objects are known as hard-clustering methods or partitional clustering. The method proposed in this work is considered within the hard-clustering or partitional clustering techniques. To that end the partitioning around the medoids methods have been studied as well as P-median and iterative partitioning or optimization methods due to the strong influence in our partitioning algorithm (Carrizosa et al.1998; Resse 2005). The most influential algorithm for the geographic clustering from this work is PAM (Partitioning Around the Medoids) (Kaufman 1987).

2.1. Partitioning

In the classification by partitions we have that being $X = \{x_1, \dots, x_n\}$ the finite set of n objects to classify and $k < n$ the number of classes into which the objects will be classified. A partition $P = (C_1, \dots, C_k)$ of Ω in k classes C_1, \dots, C_k , is characterized by the following 2 conditions for $i \neq j$ the following is met 1) $W = \bigcup_{i=1}^k C_i$ and 2) $C_i \cap C_j = \emptyset$. However, in this case, is eventually possible to allow one of the classes C_i to be empty, such that in reality the partitions $P = (C_1, \dots, C_k)$ being considered are partitions of Ω into k or fewer classes. However

the optimal partitions in accordance to an inertia criterion contain exactly k not empty classes (Piza et al. 1999).

2.2. Description of the Clustering Problem as one of Optimization

In regard to the optimization methods, their main characteristic is that they produce a unique partition of objects into a particular number k , specified beforehand, of non-overlapping clusters, as a result of the minimization or the maximization of some objective function. Regularly, these methods start with an initial partition of the set of objects into k clusters, for each one of them a centroid is defined; every object is located then in the cluster which centroid is the closest one, to calculate later the new centroids and reallocate each object again. So on and so forth until no changes are produced in the clusters. Generally, the objects are represented by D descriptive attributes in form of vectors in the space R^D , and with a similarity comparison measure, such as the distance, and the clusters are created with similar objects. In the groups or clusters formation process, there isn't a previous knowledge about how to form a cluster; for that reason, the clustering process is also known as unsupervised classification. In the clustering, the information of a series of variables for every object is used and in accordance to these variables, the similarity between these objects is measured. Once the similarity has been determined, the objects are grouped into internally homogeneous groups and different to each other. The similarity measures rely on the assumptions and the use that is given to the data; different results of the same, can result from the different similarity measures, where each one can be equally valid for a domain in particular. Given the set of n objects denoted by $X = \{x_1, \dots, x_n\}$ in which $x_i \in R^D$, let K be an unsigned integer known beforehand, the clustering problem consists in finding a partition: $P = \{C_1, \dots, C_k\}$ of X , being C_j a cluster formed by similar objects, satisfying an objective function $f: R^D \rightarrow R$ with the conditions $C_i \cap C_j = \emptyset$ for $i \neq j$ y, $C_j \cup C_i = X$. To measure the similarity between two objects x_a and x_b a distance function denoted by $d(x_a, x_b)$ is used, being the Euclidean distance the most used to measure the similarity. Thus the distance between two different elements $x_i = (x_{i1}, \dots, x_{iD})$ and $x_j = (x_{j1}, \dots, x_{jD})$ is

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^D (X_{il} - X_{jl})^2}$$

The objects from a cluster are similar when the distances between them is minimal; this allows to formulate the objective function f as

$$\sum_{j=1}^k \sum_{x_i \in C_j} d(x_i, \bar{x}_j)^2$$

; this is, minimizing f is wished; where x_j , known as representative element of

the cluster, is the measure of the elements of the cluster C_j , $x_j = \frac{1}{|C_j|} \sum_{X_i \in C_j} X_i$ and corresponds to the cluster's center. Under these characteristics, the clustering is a combinatory optimization problem and has been proven that it is an NP-hard one (Vicente et al. 2005).

2.2.1 Combinatory Nature of the Partitioning Algorithms

Given that the clustering can be approached as an NP-hard combinatory optimization problem, is similar to say that the non-hierarchical grouping is a combinatory problem. The computational complexity of this problem is of the NP-hard kind, since is in fact a generalization of the known NP problem, which originated the terminology for this complexity rank for analogue problems (Laarhoven, 1988). Is necessary to note that when is wished to obtain a partition into K classes of a set with n individuals, it doesn't make sense to examine all the possible partitions of the set of individuals into K classes. In effect, we are before a very complex combinatory problem. Just for illustrative purposes, we mention that the number of partitions of a set with 60 elements into 2 classes is about 10^{18} , and for 100 elements into 5 classes is close around 10^{68} .

In fact, it can be proven that the number $S(n, K)$ of different partitions of a set of n individuals into K classes, fulfills the recurrence equation

$(n, K) = S(n-1, K-1) + nS(n-1, K)$. This leads to

$$S(n, k) = \frac{1}{K!} \sum_{i=1}^K (-1)^{K-i} \binom{K}{i} i^n$$

(Piza et al. 1999). From the above, the need for methods and algorithms that provide a satisfactory solution to this problem is deduced. The metaheuristics incorporated to the partitioning algorithm for geographical data are presented in the following section.

3. Simulated Annealing and Variable Neighborhood Search Metaheuristics in Geographic Clustering

One simple and logical technique that can be used to solve an optimization problem consists in examining all the feasible solutions of the problem, evaluating them with the objective function and then choosing the best one. This technique called enumeration despite it is applied in different computing problems, can be impractical even for instances of regular size due to the big amount of possible solutions (Cristian 2011). However, given the underlying difficulty in the NP-hard optimization problems, as the partitioning is, the inclusion of heuristic methods is necessary even though they can't ensure achieving optimal solutions but they can be reached in a reasonable time. In this work the metaheuristics VNS and SA are combined along a K-medoids algorithm into one single algorithm which is applied to the geographic clus-

tering problem with the goal to improve the results achieved by both metaheuristics in an independent way. The results obtained with this hybrid proposal have been satisfactory when applied to sets of geographic data. In the following lines, SA and VNS are described briefly.

3.1. Simulated Annealing

The SA algorithm is a neighborhood-based search method characterized by an acceptance criterion for neighboring solutions that self-adapts at run time. It uses a parameter called temperature (T), that according to its value determines the degree in which worse neighboring solutions can be accepted. The temperature variable is initialized with a high value, called initial temperature T_0 and is reduced with each iteration through a cooling temperature mechanism (cooling factor α) until a final temperature (T_f) is reached. During each iteration a specific number of neighbors $L(T)$ is generated, which may be fixed for all the execution time or may change for each iteration. Each time a neighbor is generated, an acceptance criterion is applied to determine if it will substitute the current solution (Kirkpatrick et al. 1983). If the neighbor solution is better than the current one, it is automatically accepted, as a classic local search would do (LS). On the contrary, if it is worse, there still exists the possibility for the neighbor to substitute the current solution. This allows the algorithm to escape from local optima where LS would get trapped. The higher the temperature, the more likely it is to accept worse solutions. In this way, the algorithm accepts solutions much worse than the current one at the beginning of the execution but not at the end. By incorporating this procedure to the geographical clustering model, an appropriate algorithm for geographical clustering with simulated annealing is obtained (Bernábe et al. 2009). The general procedure for simulated annealing is presented as follows (Dowsland 2003):

```

SIMULATED ANNEALING
INPUT (To, α, L(t), Tf)
T ← To /* Initial value for the control parameter */
Scur ← Generate initial solution
WHILE T ≥ Tf DO /* Stopping condition */
  BEGIN
    FOR cont ← 1 To L(T) DO /* Cooling speed(T) */
      BEGIN
        Scand ← Select solution N(Scur) /* Creation of a new solution */
        δ ← cost(Scand) - cost(Scur) /* Computation of cost difference */
        IF (U(0,1) < e(-δ/T) OR δ < 0)
          THEN Scur ← Scand /* Acceptance criterion */
        END
        T ← α * T /* Cooling mechanism */
      END
    END
  /* WRITE as solution the best of the visited Scur */

```

In this work from 2009, a statistical experiment was developed to achieve a calibration of the best parameters for SA where 469 geographic data were used as input, which have been processed as well with the hybrid of the present paper. The main test in 2009 focused in forming 24 groups and we pick again this instance for the experimental phase of this article. This has been an important test because it's been an inflection point for groups between 10 and 200. The cost of the solutions obtained in Bernábe (2009) was satisfactory; however, measuring the solutions was necessary, in such a way that the results were compared with Variable Neighborhood Search (Bernábe et al. 2011).

The general model for geographical partitioning that has derived the SA and VNS implementations, is presented next.

3.2. Partitioning Model for Geographic Data

1) Data: \mathbf{U}_g = total of objects. Let the initial set of n geographical units be $\mathbf{U}_g = \{x_1, x_2, \dots, x_n\}$ where x_i is the geographical unit ($i = \mathbf{U}_g$ index); k is the zone number (group). The following variables are defined to refer to the different groups: Z_i is the set of geographical units that belong to the i^{th} zone (group), n is the number of geographical units, C_i is the centroid and $d(i, j)$ is the Euclidean distance from node i to node j (from one object to another);

2) Constraints: $Z_i \neq \emptyset$: for $i = 1, \dots, k$ (non empty groups), $Z_i \cap Z_j = \emptyset$ for $i \neq j$, (the same objects cannot

be in different groups), $\bigcup_{i=1}^k Z_i = \mathbf{U}_g$ (the union of all the groups are all the objects) and

3) Objective Function: Once the number of centroids (k) is decided (C , with $i = 1, \dots, k$), the centroids will be randomly selected and the objects will be assigned to the nearest centroids. Then, for each object, the objective function is defined as the minimum of the sum of the distances between the centroids (for each k), and the objects assigned to them (each object is assigned to the closest centroid). For every k (where $k = 1, \dots, n$) the sum of the distances from every object assigned to each centroid is calculated, and the minimum is selected. Therefore the objective function f can be written as:

$$\text{Min} \left\{ \sum_{t=1}^k \sum_{i \in C_t} d(i, c_t) \right\} \text{ where } n_{it} \text{ is the number of iterations}$$

3.3 Variable Neighborhood Search VNS

The VNS metaheuristic, proposed by Hansen and Mladenovic (1997) is based on the observation that local minima tend to cluster in one or more areas of the search

space. Therefore when a local optimum is found, one can get advantage of the information it contains. For example, the value of several variables may be equal or close to their values at the global optimum. Looking for better solutions, VNS starts exploring, first the nearby neighborhoods of its current solution, and gradually the more distant ones. There is a current solution S_a and a neighborhood of order k associated to each iteration of VNS. Two steps are executed in every iteration: first, the generation of a neighbor solution of S_a , named $S_p \in N_k(S_a)$, and second, the application of a local search procedure on S_p , that leads to a new solution Sol. If Sol improves the current solution S_a , then the searching procedure will restart now from Sol using $k = 1$. Otherwise, k is incremented and the procedure is repeated from S_a . The algorithm stops after a certain number of times that the complete exploration sequence $N_1; N_2; \dots; N_{kmax}$ is performed. The following algorithm shows how the solutions are obtained.

Procedure Variable Neighborhood Search

```
BEGIN
/*Nk: k=1,...,kmax, neighborhood structures */
/* Sa: current solution */
/* Sp: neighbor solution of Sa */
/* Sol: local optima solution */

REPEAT UNTIL (END) DO
k ← 1
  REPEAT UNTIL (k ← kmax) DO
/* Generate neighbor Sp of the kth neighborhood of Sa (Sp ∈ Nk(Sa)) */
Sp ← GetNeighbor (Sa, Nk);
Sol ← LocalSearch (Sp);
  IF (Sol is better than Sa) THEN Sa ← Sol;
  ELSE k ← k + 1
  END DO
END DO
END
```

Comparing the quantitative and qualitative behavior of the geographical partitioning between VNS and SA was a needed task, therefore, in a separate way, each metaheuristic has been incorporated to the exposed model where a response surfaces analysis in a factorial experiment was validated and it was reported that VNS throws better results than SA for the 24 groups test with 469 geographic objects (Bernábe et al. 2011). From this result, we assumed that better solutions were possible if we hybridized VNS and SA. In this point, an additional effort has been combining VNS and SA into one hybrid heuristic with the goal to achieve approximations that surpass VNS and SA when independently implemented to solve this geographic clustering problem. The computing experience is shown below.

4. SA-VNS Hybrid for Geographic Clustering

Although trajectory methods only deal with a single solution, they have shown their potential in exploiting the promising regions in the search space with high quality solutions such as VNS. VNS is a simple and effective trajectory metaheuristic. The basis of VNS involves using more than one neighborhood structure and systematically changing the neighborhood within a local search process. Unlike many other metaheuristics, the basic scheme of VNS and its extensions require few and sometimes no parameters. However, it is still prone to premature convergence traps due to the limited exploration ability. Hence, it is a likely choice to consider the hybridization of metaheuristics. It has been decided to start the clustering process with SA given its high capability to escape local optimums, then when it has “convergence” in its last iteration, this final solution is given as an argument to VNS, where VNS only searches for even better solutions. A summarized version of the algorithm is presented as follows:

Algorithm SA-VNS Hybrid

Input k number of groups

Input for SA T_0 initial temperature, T_f final temperature, α cooling factor and $L(t)$.

Input for VNS: ni number of iterations for neighborhood structures (NS in VNS), nbl number of iterations for Local Search (LS in VNS).

Initialize $SA(T_0, T_f, \alpha, L(t))$ and $VNS(ni, nbl)$

$initSol \leftarrow$ Generate a Random Initial Solution(k)

$saSol \leftarrow$ Execute SA ($initSol$)

$Sol^* \leftarrow$ Execute VNS ($saSol$)

Generate Clusters (Sol^*)

4.1. Computing Experience

To measure the hybrid’s behavior against other competitive implementations, we have done different test runs gathering them into tables 1 and 2, which show the different number of groups formed for 2 geographic zones: Toluca Valley Metropolitan Zone (TVMV) with 469 geographic units U_g (objects) and Mexico Valley Metropolitan Zone (MVMV) with 5023 geographic units.

One of the general characteristics metaheuristics is the randomness of the initial solution. In our hybrid algorithm, we have kept such property and just as it is stochastic, is recommended to carry out at least 20 runs for each test and to calculate the average of each one of them. We register these results in tables 1 and 2.

On the other hand, we have statistically calibrated the combined parameters for the hybrid algorithm, then, we randomly select the following parameters for the tests: for SA the parameters for all the test runs have been $T_0=5000$, $T_f=.01$, $\alpha=.95$, $L(t)=5$ and for VNS $LS=20$, $NS=2$, $T=time(sec.)$ and H is Hybrid. In the tables, 1), 2), 3), 4)

and 5) represent Hybrid, SA, VNS, P-median with GAMS (General Algebraic Modeling System) and PAM respectively.

Table 1. Test runs gathering for Geographic Clustering (ZMVT)

	Instance					
	TZMV					
	24		47		94	
	Cost	T	Cost	T	Cost	T
1)	11.59	3	8.11	2	4.97	4
2)	11.89	9	8.36	15	5.17	21
3)	11.85	1	8.19	1	5.02	2
4)	9.20	936	5.73	1123	3.29	93
5)	9.19	79	5.73	431	3.20	71

Table 2. Test runs gathering for Geographic Clustering (ZMVM)

	Instance					
	MVMV					
	255		508		1018	
	Cost	T	Cost	T	Cost	T
1)	73.01	628	50.42	1387	31.01	2434
2)	72.49	191	51.94	554	30.82	843
3)	70.93	555	50.74	1240	29.24	1337
4)	--	--	--	--	--	--
5)	--	--	--	--	--	--

As we can observe for small instances of 469 objects, the hybrid algorithm responds with better solution quality in a very reasonable amount of time, but still far from an “optimum” like the one provided by PAM and GAMS, of course these 2 with a high computing cost. However for big instances of 5023 objects, only for 508 groups, the hybrid algorithm is the best one. For GAMS and PAM it wasn’t possible to obtain a solution in a running time of 44 hours.

The following figure shows the behavior of the best solution obtained for 24 groups, with an objective function cost of 11.5958.

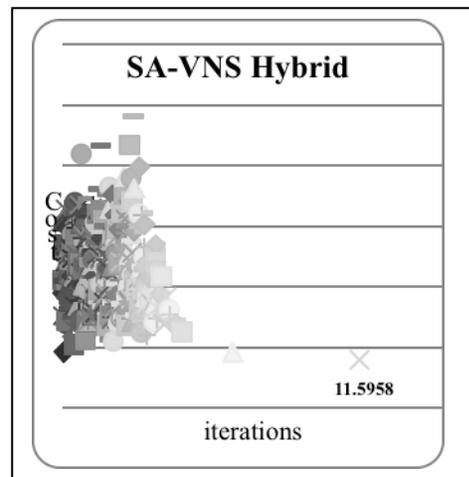


Figure 1. Plotting of the SA clustering: cost of the solution vs. number of iterations

An implementation with a Geographic Information System GIS eases the numerical interpretation of the previous figure, which makes possible to see the result in the map (the data under study are geographic units that belong to a territorial zone), (Bernábe, 2013).

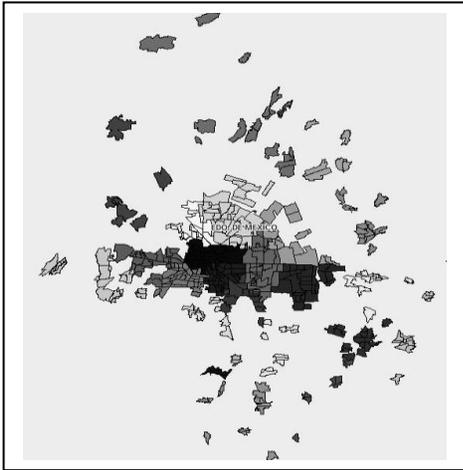


Figure 2. Graphic associated to the result of the clustering for 24 groups.

Once that it has been proven that hybrid SA-VNS partitioning achieves better results than the implementations already reported (Bernábe et al. 2011), the next step consist in developing another test: knowing the connection of the algorithm; this is, it's important to prove that the algorithm improves the geometric compactness, a property always sought in territorial design and partitioning problems over territorial or geographic units (Bernábe et al. 2011). The next step resides in having another algorithm that measures such connection. In the literature we can find an algorithm that has been proven to measure the connectedness among the group members but is possible to adapt it to geographical data (XiuLi et al. 2011). In the following section, the application of the Dunn index to the groups obtained from the test runs is presented.

5. Evaluating compactness with the Dunn Index validity measure

In this section we perform an assessment of the presented approaches: SA, VNS and SA-VNS using the Dunn Index metric. The aim is to evaluate the quality of each clustering method, and use this metric for determining the approach that performed better, from the point of view of compactness. The Dunn index is a widely used metric for evaluating clustering algorithms (Dunn, 1974). It is basically, an internal validity measure which aims to identify sets of clusters that are compact, with a small variance between members of the cluster, and well separated, where the means of different clusters are sufficiently far apart, as compared to the cluster variance within. The higher the Dunn index, the better the clustering quality. Let C_i be a cluster of m objects, let x and y be any two objects assigned to the same cluster C_i . The Dunn Index (DI) is defined as follows:

$$DI = \min_{1 \leq i \leq m} \left\{ \min_{1 \leq j \leq m, j \neq i} \left\{ \frac{\partial(C_i, C_j)}{\max_{1 \leq k \leq m} \Delta_k} \right\} \right\}$$

where $\partial(C_i, C_j)$ is the intercluster distance metric between clusters C_i and C_j ,

whereas

$$\Delta = \frac{1}{|C_i||C_{i-1}|} \sum_{x, y \in C_i, x \neq y} (d(x, y))$$

calculates the average distance between all pairs.

In Table 3 we can see the obtained Dunn index values for each approach. The values obtained by de internal validity measure show that the SA-VNS approach performed better than the other two approaches.

Table 3. Evaluating compactness by means of the Dunn Index metric

Approach	Dunn Index	Ranking
Hybrid	0.923	1st
SA	0.584	2nd
VNS	0.465	3rd

The results suggested by the Dunn Index metric can be visually corroborated when the 24 clusters obtained by each approach are plotted. Figure 3, 4 and 5, show the distribution of similarities among objects for the VNS, SA and Hybrid approach, respectively.

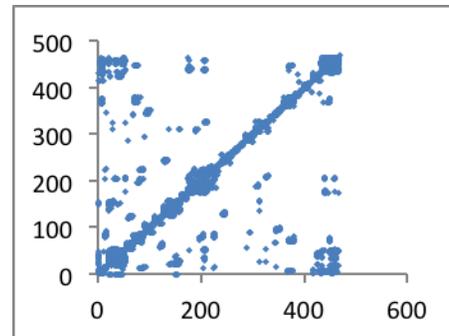


Figure 3. Plotting of the SA clustering approach.

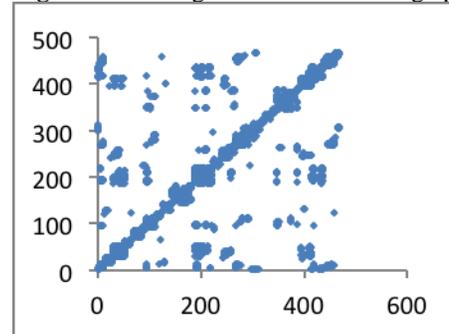


Figure 4. Plotting of the VNS clustering approach.

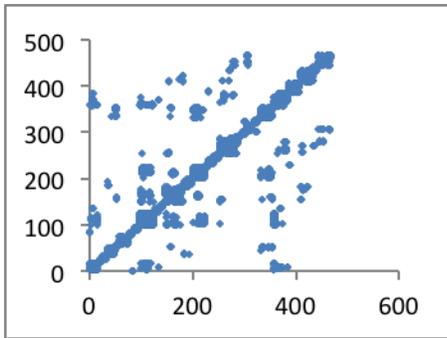


Figure 5. Plotting of the Hybrid clustering approach.

We have re-ordered the real objects, so that consecutive objects now belong to the same cluster. In theory, the ideal clustering algorithm should obtain a perfect object grouping which will show a graphic with squares of points in the diagonal of the plotting. In other words, in a perfect clustering there will be not points outside the diagonal.

The maximum Euclidean distance obtained between two objects was 0.577332, whereas the minimum distance was 0.000447214. Let us consider that the lower the Euclidean distance, the more similar the pair of objects. Being similar is a subjective issue, however, we may propose a threshold for determining a positive degree of similarity, which in this case was empirically established as 0.1730. Future work will analyze the behaviour of each approach with different thresholds of similarity.

In this research work, all those object pairs that share a distance lower than the empirical threshold are plotted in the above mentioned figures. By observing the number of points plotted outside the diagonal, we can easily conclude that the hybrid approach performed better than the other two approaches, a fact that confirms the results obtained with the Dunn Index.

6. Conclusions

In this work we have reported an experience with SA and VNS combined in an effort to achieve better solutions than the ones already reported. For small instances, is possible to rely on our hybrid proposal and as can be seen in table 2, for a bigger sized problem the results are not very encouraging, at least for the number of groups we have chosen.

The internal validity measure has shown that the hybrid approach has performed better than the other two approaches. This fact was also visually confirmed by plotting the groups obtained by each approach when a threshold of similarity is established.

As future work, we propose a factorial experiment design for VNS, SA and the hybrid algorithm proposed, and then the number of groups and the parameters of the metaheuristics would be variable and would indicate the behavior of the metaheuristics with more reliability to determine their performance. On the other hand, PAM and

GAMS result untreatable for big instances. This issue implies proposing a Lagrange Relaxation scheme to compare our results.

In previous works, we have applied VNS partitioning to the location of ovitraps for a Dengue mosquito problem and with the results obtained in this article we ensure that is possible to apply SA-VNS to the Dengue problem with bigger size instances obtaining satisfactory solutions in regard to time and quality, furthermore SA-VNS can be applied to similar problems of bioinspired nature (Bernábe 2012).

Currently we are working with the geographic partitioning with homogeneity restrictions in a multiobjective context.

7. References

- [1] B. Bernábe, J. Espinosa, J. Ramírez, M. Osorio, "A Statistical comparative analysis of Simulated Annealing and Variable Neighborhood Search for the Geographical Clustering Problem", *Computación y Sistemas*, Vol. 14, pp. 295-308, 2011.
- [2] Bernábe M. B., Espinosa J. E. & Ramírez J. (2009), "Evaluación de un Algoritmo de Recocido Simulado con Superficies de Respuestas", *Revista de Matemáticas Teoría y Aplicaciones*, 16(1), pp.159-177, 2009.
- [3] B. Bernábe, M. A. Rodríguez, R. A. Martínez, J. Ramos, E. Olivares, "Adaptation of a Clustering Algorithm and Mosquito Swarm to a problem of ovitraps for the Dengue Mosquito Vector", *Online Proceedings on Trends in Innovative Computing*, pp.108-113, 2012.
- [4] B. Bernábe, "Integración de un Sistema de Información Geográfica para Algoritmos de Particionamiento", unpublished 2013.
- [5] E. Carrizosa, E. Muñoz, & J. Puerto, "The Weber Problem with Regional Demand", *European Journal of Operational Research*, Vol. 104, pp.358-365, 1998.
- [6] A. Cristian Martínez, "Metaheurísticas híbridas aplicadas al Problema de Ruteo de Arcos Capacitados", Tesis doctoral, *Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Computación*, pp. 12-20, 2011.
- [7] K. A. Dowsland, & Adenso Díaz, "Diseño de Heurística y Fundamentos del Recocido Simulado", *Revista Iberoamericana de Inteligencia Artificial*, Vol. 7 (19), pp. 93-102, 2003.
- [8] GAMS, Optimization Software <http://www.gams.com/>.
- [9] L. Kaufman, P. J. Rousseeuw, "Clustering by means of medoids", *Statistical Data Analysis based on the L1 Norm, North-Holland Amsterdam*, pp. 405-416, 1987.
- [10] Laarhoven, E. Aarts, J. Korst, "Simulated Annealing: Theory and Applications", *Kluwer Academic Publishers Dordrecht*, 1988.

- [11] N. Mladenovic & P. Hansen, "Variable Neighborhood Search", *Computers & Operations Research*, Vol. 24(11), pp. 1097-1100, 1997.
- [12] E. Pizza, A. Murillo & J. Trejos, "Nuevas técnicas de particionamiento en clasificación automática", *Revista de Matemáticas Teoría y Aplicaciones*, Vol. 16(1), pp. 51-66, 1999.
- [13] J. Resse, "Methods for Solving the p-Median Problem: An Annotated Bibliography", pp. 1-4, 2005.
- [14] E. Vicente, L. Rivera, D. Mauricio, "Grasp en la resolución del problema de cluster", Vol. 2(2), pp. 16-25, 2005
- [15] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics*, Vol. 3 (3), pp. 32-57, 1973
- [16] Z. XiuLi & X. WeiXiang, "A New Measurement Method to Calculate Similarity of Moving Object Spatio-Temporal Trajectories by Compact Representation", *International Journal of Computational Intelligence Systems*, Vol. 4(6), pp. 1140-1147, 2011.