Aplication of a k-medoids Algorithm to Dengue Mosquito Ovitramps Problem

María Beatríz Bernábe Loranca[1], Marco Antonio Rodríguez Flores[2], Elias Olivares Benitez[3], David Pinto Avendaño[4]and José Luis Martínez Flores[5]

[1,2,4]Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación
beatriz.bernabe@gmail.com

[1,3,5]Universidad Popular Autónoma del Estado de Puebla, Posgrado en Logística y Dirección de la Cadena de Suministro
elias.olivares@upaep.com

Abstract: **The efforts to locate the ovitraps in an homogenous way in a determined community must be redouble still, the reality is that the location of these depends too much of the will of the community and of the use that is given to the piece of land where the ovitrap should be located.**

Given that around the spot of the dengue cases the transmission is more probable (infected mosquitos) and that it's been documented that vertical transmission exists (adult mosquito to eggs), to the viral and entomological vigilance of endemic communities is important to know and to monitor, through the research of the material obtained by the ovitraps, the strains of Dengue virus that circulate between the human population and the mosquito populations.

If the ovitraps were located according to a random representative design and homogenous in the community of study, the vigilance system described above would consist in studying the ovitraps in the range of flight of the mosquito (200 meters approximately). However, in reality this is not the case, therefore a probabilistic approximation is required to establish which ovitraps should be evaluated by the system of sanitary vigilance to have a higher probability of success in the diagnosis, making the process cost-efficient.

In this scenery, in accordance to the mobility of the mosquito, a clustering algorithm has been associated and adapted based in P-means that promises to construct groups where the center of each group is a case and the closest ovitraps are associated to establish a systematic and homogenous configuration of the relationship between a registered case and the ovitraps.

Keywords: **Cases, Clustering, Dengue, Mosquito, Ovitraps**

I. Introduction

The system of entomological vigilance that is carried out by the authorities of national health is based in the called ovitraps; these are the units were the Aedes mosquito females lay eggs after feeding with human blood. In this point, it is assumed that the number of eggs in said devices is proportional to the number of vectors that transmit Dengue virus.

All Vector control programs, which have focused mainly on the patient house and peridomestic areas around dengue cases, have not produced the expected impact on transmission.

To evaluate the assumption that the endemic/epidemic transmission of dengue begins around peridomestic vicinities of the index dengue cases, a prospective cohort study was conducted (in Tepalcingo and Axochiapan, in the state of Morelos, Mexico), using the state surveillance system for the detection of incident cases. Paired blood specimens were collected from both the individuals who live with the incident cases and a sample of subjects residing within a 25-meter radius of such cases (exposed cohort), in order to measure dengue-specific antibodies. Other subjects were selected from areas which have not presented any incident cases within 200 meters, during the two months preceding the sampling (non-exposed cohort) [1].

Symptomatic/asymptomatic incident infection was detected. In the analysis it will be considered as the dependent variable, the exposure to confirmed dengue cases as the main independent variable, and the estimated Aedes. Aegypti abundance and their infection with Dengue Virus (DENV), also socio-demographic and socio-cultural conditions of the subjects will be considered as additional explanatory variables [1].

As it was impossible to monitor and ascertain vector density in the domiciles of the participating subjects prior to and during follow up, we want to estimate this variable based in the data from the ovitrap-based surveillance program. Also, due to the difficult logistics involved in the capture of adult Aedes, the mosquito infection will be analyzed in the individuals that emerge from the eggs collected through the ovitraps closest to the subjects' dwellings, but we don't know how to select these ovitraps because their location into the endemic area is not homogeneous.

To answer to this situation, we propose configuring a structure of the ovitraps under schemes of clustering algorithms analogue to the swarms of mosquitos and in particular to the sample of the behavior of the dengue mosquito.

The study in this work starts with the introduction as section 1. In the section 2 we expose the way of life of the swarms of mosquitos as a bioinspired algorithm. We continue the section 3 with a brief state of the art of the problem of the

Dengue mosquito giving place to describe in section 4, the general aspects of clustering and in section 5 we present a proposal of an algorithmic solution of clustering with analogy to P-medians and the section 6 we expose a new partitioning algorithm based on k-medoids to solve the problem of the cases-ovitraps and finally in section 6 we present a discussion of the results and lastly the conclusions.

II. Mosquito Swarm Algorithm

We can see on the literature a new propose classification of meta-heuristics algorithms not based on swarm intelligence theory but rather on grouping of animals: swarm algorithms, schools algorithms, flocks algorithms and herds algorithms: The swarm algorithms (inspired by the insect swarms and zooplankton swarms): Ant Colony Optimization algorithm – ACO (inspired by the research on the behavior of ant colonies), Firefly Algorithm (based on fireflies), Marriage in Honey Bees Optimization Algorithm - MBO algorithm (inspired by the Honey Bee), Wasp Swarm Algorithm (inspired on the Parasitic wasps), Termite Algorithm (inspired by the termites), Mosquito swarms Algorithm–MSA (inspired by mosquito swarms), zooplankton swarms Algorithm-ZSA (inspired by the Zooplankton) and Bumblebees Swarms Algorithm–BSA (inspired by Bumblebees) [2].

Grouping of animals is a natural phenomenon in which a number of animal individuals are involved in movement as forming a group; there are insect swarms, zooplankton swarms, fish schools, bird flocks, and mammal herds [3]:

1. Swarms. The swarm is a social grouping (of the same species) of insects and marine zooplankton. There are several types of swarms, two of the best known are insects swarms and marine zooplankton swarms. Insects swarms consist of the following insects: honey bees, Africanized Honey Bees, ants, termites, desert locusts, gnats, midges, mosquito, houseflies, African Fly of the Nile river, pine beetles, ladybug, aphids, Monarch butterflies, Bumblebees, Fire Ants, Army Ants, Yellow Jackets, and others). Zooplankton swarms contains animal organism called zooplankters (copepods, mysids, segetids, Scyphomedusae, and others) [2].

2. Swarm Algorithms. The term swarm algorithm refers to any algorithm that models the grouping of insects and zooplankton swarms by social behavior. The swarm algorithms are inspired by the insect swarms and zooplankton swarms. Some of the most popular swarms algorithms are: The Ant Colony Optimization algorithm (ACO) was inspired by the research on the behavior of ant colonies [4]; the Firefly Algorithm is based on insects called fireflies [5]; the Marriage in Honey Bees Optimization Algorithm (MBO algorithm) is inspired by the process of reproduction of Honey Bee [6], the Artificial Bee Colony Algorithm (ABC) is based on the recollection of the Honey Bees [7], the Wasp Swarm Algorithm was inspired on the Parasitic wasps [8], Bee Collecting Pollen Algorithm (BCPA) [9], Termite Algorithm [10], Mosquito swarms Algorithm (MSA) [11], zooplankton swarms Algorithm (ZSA) [12] and Bumblebees Swarms Algorithm (BSA) [13].

A. Swarm Algorithms

The swarm algorithms were developed by analogy with aspects of the insect swarms and zooplankton swarms. In this section we only show the algorithms: Mosquito swarms Algorithm (MSA), zooplankton swarms Algorithm (ZSA) and Bumblebees Swarms Algorithm (BSA).

Mosquito Swarms Algorithm (MSA)

Ruiz-Vanoye and Díaz-Parra [11] propose the Mosquito Swarm Algorithm (MSA). The MSA is considered as a meta-heuristics algorithm, a bio-inspired, parallel or distributed algorithm based on the research on the social behavior of mosquito swarm [11].

Mosquitoes (gnats) have sensors designed to track their prey: A) Chemical sensors, mosquitoes can sense carbon dioxide and lactic acid up to 36 meters away. Mammals and birds give off these gases as part of their normal breathing. Certain chemicals in sweat also seem to attract mosquitoes. B) Heat sensors, Mosquitoes can detect heat, so they can find warm-blooded mammals and birds very easily once they get close enough. A mosquito swarm exists close to areas with standing water.

Mosquito Swarm Algorithm.

Input: number of mosquitoes (n)

1. Initialize a Mosquito Population with Chemical Sensors (CS) and Heat Sensors (HS).

2. Generating the initial locations (x) of the mosquitoes (n).

3. Initialize the temperature (t) and Maximum Temperature (T).

4. Repeat (total of mosquitoes) //by parallel and/or distributed processing

5. Repeat (maximum temperature)

6. Generate new solutions by adjusting the Heats (HS) and Updating the locations (x).

7. Verify and assign the feasibility of the solution by the Chemical Sensor (CS).

8. Select the best solution (S).

9. While t < T // (Maximum Temperature)

10. While (n total of mosquitoes)

11. Report the best solutions.

Based on the above description of mosquito swarm process, Ruiz-Vanoye and Díaz-Parra propose the Mosquito Swarm Algorithm. The structure of the MSA can see in [11].

III. Dengue

Dengue, the most frequent vector-borne viral disease in the world, constitutes a public health problem in tropical and subtropical countries [1]. In 2010 and 2011 Mexico reported to the World Health Organization 57,971 suspected cases and 22,352 confirmed cases and 67,918 suspected cases and 15,578 confirmed dengue cases, with 20 and 36 deaths, respectively.

The dengue endemic/epidemic cycle is maintained by the vector through mosquito-man-mosquito transmission. Once it has bitten an infected individual, it incubates the virus for a period of 7-14 days, and then starts to infect healthy individuals through its bite. In humans, Dengue virus

(DENV) incubation extends over a period of 3 to 14 (7 on average) days. The actual infectious stage, or time span during which the diseased person spreads the infection to a vector, commences one to two days before the onset of symptoms, at which time high levels of viremia appear and continue to develop throughout a febrile cycle of 2 to 10 days. The DENV infection spectrum covers symptomatic cases (dengue fever, dengue hemorrhagic fever, dengue shock syndrome and death), as well as asymptomatic cases, which, in most studies, represent over half of total infections.

Further, in several countries including Mexico, it has been reported vertical transmission (transovarial transmission) of DENV in A. aegypti and A. Albopictus (infected female mosquito transmits to its brood). It has been evaluated in male mosquitoes and larvae/pupae collected in a field, and in the early stages samples (eggs collected in ovitraps, larvae and pupae collected in breeding sites) then reared up to adult stage. Additionally, it has been found that virus could persist in mosquitoes in successive generations through transovarial passage (even until seventh generation), although the rate is decreased after of second generation [14, 15, 16, 17, 18, 19, 20].

Given the absence of a DENV-specific vaccine or antiviral treatment, health services have focused their prevention efforts on vector control. Government vector control programs offer a number of partially successful examples.

At the government level, vector control measures have concentrated on peridomestic dengue areas. For instance, the Mexican National Health Standard obligates the state health services to execute larval control and dejunking campaigns permanently, as well as nebulization and fumigation procedures during peak transmission periods, targeting the domiciles and peridomestic vicinities of dengue cases detected through epidemiological surveillance. Nevertheless, the results have fallen short of expectations.

On the other hand, there are several methods to monitor the vector population. The ovitrap is a sensitive and economic method to detect the presence of Ae. aegypti and Ae. Albopictus. Also, it has been observed that the ovitraps are more sensitive than the mosquito traps and they are more efficient than larvae traps, even if there are natural breeding sites [21, 22, 23].

Ovitraps have been used in different countries, for example in Argentina, these traps were employed to determinate the date of mosquito population increase and to detect the optimal time to apply vector control measures; and in Brazil evaluated the ovitraps to identify the places where the vector population is consistently concentrated over the time to pinpoint this areas that should be considered as a high priority places to apply vector control activities [24, 25].

However, México is the only country that uses the ovitraps as an entomological surveillance national system in dengue endemics areas. The ovitraps are planted in the first quarter of the year; these are reviewed once a week and are relocated each year.

In each endemic locality must be selected one block per four-six blocks and must be selected six or seven blocks in each cardinal point. After, four ovitraps per each block (one on each side of the block) must be planted; in principle the ovitrap must be located in the center of each block side; however the final location depends of the house owner acceptation and his availability to permit to review the ovitrap each week.

In each house only can be planted one ovitrap outside the home (yard or patio), in a shady spot (protected to sun/rain), at a height lower than 1.20 meters and the place must be out of reach of children and pets. Also, in the locality each ovitrap is situated on a map and recently each ovitraps is georeferenced with a GPS (Operational guidelines for placement of Morelos ovitraps).

Objetive:

The general objective of the present study is to develop a mathematic algorithm to determine which ovitraps could be examined to evaluate the vertical transmission and the abundance of vectors around an index dengue case.

IV. Clustering

The cluster analysis has importance in the feasibility of finding groupings (clusters) directly in the data without using any previous knowledge. The use of grouping (clustering) in diverse areas is beneficial, however, in order to have efficient techniques of cluster analysis it must exist some kind of similarity between data. Several researchers propose its use in spatial data, given the existence of notions of distance and algorithms for partitioning around the medoids, they are adapted with ease to this kind of data. In particular, the model Partitioning Around Medoids (PAM), achieves this purpose determining an object, representative for each cluster to find k clusters (groups) [26]. This representative object, called medoid, is the one located closer to the center within the cluster. Once the medoids have been selected, each unselected object is grouped with the medoid that is more similar to itself. In this point, an evident weakness of PAM is the process of repeated and long local searches in the solution space.

K-means clustering and are well known techniques for performing non-hierarchical clustering.

Let us describe the clustering problem formally. Assume that S is the given data set $S = \{\vec{x}_1, ..., \vec{x}_n\}$, where $\vec{x}_i \in R^n$. The goal of clustering is to find K clusters $C_1, C_2, ..., C_k$ such that

$$C_i \neq \varnothing \text{ for } i = 1, ..., k \qquad (1)$$

$$C_i \cap C_j = \varnothing \text{ for } i, j = 1, ..., k; i \neq j \qquad (2)$$

$$\bigcup_{i=1}^{k} C_i = S \qquad (3)$$

and the objects belonging into same cluster are similar in the sense of the given metric, while the objects belonging into different cluster are dissimilar in the same sense. In other words, we seek a function $f : S \rightarrow \{1, ..., k\}$ such that for $i = 1, ..., k : C_i = f^{-1}(i)$, where C_i satisfy the above conditions.

The Euclidean metric can be used to measure the cluster quality. The function f is sought such that

$$f = \arg\min_{f} E_{VQ}(\vec{c}_1, ..., \vec{c}_K) \qquad (4)$$

$$f = \arg\min_{f} \sum_{i=1}^{k} \left\| \vec{x}_i - c_{f(x_i)} \right\|^2$$

where $\vec{c}_k = \dfrac{1}{|C_K|} \sum_{x_i \in C_k} \vec{x}_i$, $k = 1,...,K$ (5)

Therefore instead of function f directly, one can search for the centers of the clusters, i.e, vectors $\vec{c}_1,...,\vec{c}_K$, and implement the function f as

$$f(\vec{x}) = \arg\min_{i} \left\| \vec{x} - \vec{c}_i \right\|^2 \quad (6)$$

that is, assign the point to the cluster corresponding to the nearest center.

Considering the properties if clustering described above and PAM, we have adapted to PAM the capability to input a vector of centroids of interest, such that the final solution will be restricted to achieve a cost that sacrifices optimality without losing feasibility but answers to the needs of choosing group representatives in an unusual way in PAM. Two proposals have been implemented: 1) the objects are assigned as usual to the medoids (the closest one), but providing one unique solution, assigning the objects to the medoids input vector and calculating the cost. 2) An improved version of the previous algorithm with the advantage that the medoids input vector works as a not-medoids vector and the usual swap of objects for medoids of PAM is restricted to medoids and the objects in the not-medoids input vector. Therefore this algorithm provides the best solution of a restricted search.

V. Adaptation of a P-means Algorithm and Clustering in the Dengue Problem

Among the location problems that are often used as prototypes to solve other problems, are the set coverage problem, the p-center problem, the quadratic assignment problem and the p-means problem among others, all of them of high complexity [27].

The p-means problem is very important in diverse areas such as the logistics networks design to locate service centers, the decision making in the location of facilities and the assignment of demand spots to these locations; for this reason it is classified as well as a localization-assignment problem, like the problem under study in this work.

The p-means can be interpreted over a related undirected graph where the distance between any pair of vertices can be obtained. It is wished to find a subset of these vertices of cardinality p from where a service can be provided to other vertices that must be the closest ones, this speaking in logistics terms, in such a way that the transport cost is minimal. It can be understood as a discrete optimization problem [28] and can be laid out as a binary integer problem (BIP) in the following way:

Let $x_{ij} = \left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\}$ (x_{ij} is 1 if the vertex j is assigned to vertex i and 0 in any other case).

Let $y_i = \left\{ \begin{smallmatrix} 1 \\ 0 \end{smallmatrix} \right\}$ (y_i is 1 if vertex i is a mean and 0 in any other case).

$$\text{Min } Z = \sum_{i=1}^{k} \sum_{j=1}^{n} C_{ij} x_{ij} \qquad (1)$$

Subject to $\sum_{i=1}^{k} x_{ij} = 1, \quad \forall j = 1,..,n$ (2)

$$\sum_{j=1}^{n} x_{ij} \le n y_i \quad \forall_i = 1,..,k \qquad (3)$$

$$\sum_{i=1}^{k} = p \qquad\qquad (4)$$

Where k is the number of potential vertices where a mean can be located, generally $k=n$. p is the fixed number of required means and the Euclidean distance is C_{ij}

The development of the approach of the p-means problem took place in the 60's; it can be attributed to Hakimi in the direct case and to Weber the continuous case [29]. It is also applied in areas such as the territorial design problem as a clustering analysis tool and data mining.

The problem belongs to the class NP-hard [30] and this means that there isn't an exact algorithm that can solve it in a general way whereby in recent years it has been attempted to be solved by means of several methods of approximation like heuristics, metaheuristics and relaxations. In [30] an efficient genetic algorithm to solve this problem is presented and in [29] a summary is shown about the methods that have been used to solve the p-means problem such as branch and bound, dynamic programming, GRASP, Lagrange relaxation, genetic algorithms and variable neighborhood search.

This algorithm and the clustering by partitions are of iterative partition or optimization nature and produce in the final solution a unique partition of objects into k non-overlapping clusters where the number k is previously specified as a result of the minimization or 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; each object is located then in the cluster that has the closest centroid and later, the new centroids are calculated to reallocate each object again. And so on until there is no change in the clusters or a cost function is reached. The most important restrictions consist in that each group has at least one element and each element belongs to only one group.

However, applying this algorithm to the data of the traps and registered cases of the Dengue mosquito, the resulting groupings are satisfying as an optimal solution but not to the case of study, which requires that each group has a registered case as centroid.

In this point an algorithm has been constructed that allows choosing the centroids as a registered case of dengue.
The algorithm is presented as follows:

A. Clustering with Predetermined Centroids (Algorithm 1)

Input:
k – Number of groups
C – Array of Centroids of size k
D – Dissimilarity matrix of size $n \times n$
A – Array of Geographical Units of size n
S – Array of Set Sizes of size k
1. For i from 1 to n
    1.1. *index* ← closest_centroid(*A(i), C, D*)
    1.2. *G(S(index), index)* ← *A(i)*
    1.3. *S(index)* ← *S(index)* + *1*
    1.4. *Cost* ← *Cost* + *D(C(index), A(i))*
2. End_For

3. Return *Cost*

Description of the algorithm:

Given *k* groups to form, the algorithm requires as input an array *C* that contains *k* centroids given by the user, furthermore it is required the dissimilarity matrix D to determine the distances between objects, the geographical units to group contained in the array *A* and *S* is the array that registers the size of each group, initially 1 is the size of each group due to the fact that the centroids are counted as members.

In step 1, the array of geographical units is scanned, afterwards in 1.1 the index of the closest centroid to the geographical unit *A(i)* is obtained, this index is the same in every structure related to the clusters, that is, *C, S and G,* this last structure is a matrix of size $n \times k$ in which the geographical units are stored under the column of its corresponding centroid, this action is realized in step 1.2, on the other hand, the index corresponding to the row, that will store the geographical unit, of matrix G is determined by size of the current group (centroid), that is, *S(index)*. After that the size of the group is increased and the distance between the centroid *C(index)* and the geographical unit *A(i)* is added to the cost of the solution. Finally the algorithm returns the cost and the clustering is stored in matrix G.

B. Results

The geographical data under study is given in latitude and longitude; corresponding to 99 ovitraps plus 29 registered cases, then 29 groups have been created to form groups where each center of the groups is a case and the closest ovitraps are assigned to each center with the end of achieving an objective structure and reliable to place ovitraps strategically, where the cases have been registered.

Figure 1 shows the case of study in Tepalcingo Morelos, the zone where the ovitraps have been previously placed.

The picture seen in figure 1 has been extracted with the help of a Geographical Information System GIS. The coordinates of the cases and ovitraps have made easier different tasks needed for the development of this work: a) the map of Tepalcingo with the marks of the cases and ovitraps, b) the creation of the distances matrix (needed in the input of the algorithms) and the construction of the graphs.

Figure 1. Case of stuy in Tepalcingo Morelos

We have done 29 groupings, from 1 element to 29. Some groups only contain the centroid, that is, one registered case. These groups must be analyzed in accordance to the geographical zone to know the situation of the Dengue mosquito appearance. For illustrative effects, we show the results for 12 groups. Due to the fact that the algorithm has been built as an algorithm that optimizes the objective function of distances minimization between ovitraps and registered cases (centroid), we have calculated the cost function, the computing time, the centroids and the objects that belong to the group. When the size of the group is 1, the centroid (case) is the only element:

Ovitraps grouping for 12 groups

Number of Groups: 12, Cost: 201.5455, Time: 573.
Cluster 1: Size 1, Centroid 101
Cluster 2: Size 1, Centroid 102
Cluster 3: Size 1, Centroid 103
Cluster 4: Size 1, Centroid 104
Cluster 5: Size 9, Centroid 105, Elements 85, 86, 87, 88, 89, 90, 91, 92
Cluster 6: Size 1, Centroid 106
Cluster 7: Size 7, Centroid 107, Elements 73, 74, 93, 94, 95, 96
Cluster 8: Size 7, Centroid 108, Elements 72, 75, 76, 97, 99, 100
Cluster 9: Size 14, Centroid 109, Elements 53, 61, 62, 63, 64, 67, 68, 77, 81, 82, 83, 84, 129
Cluster 10: Size 12, Centroid 110, Elements 1, 2, 3, 45, 46, 69, 70, 71, 78, 79, 80
Cluster 11: Size 34, Centroid 111, Elements 4, 5, 6, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 47, 48, 49, 50, 51, 52, 54, 55, 56, 57, 58, 59, 60, 65, 66, 98, 124, 123, 114, 115
Cluster 12: Size 41, Centroid 112, Elements 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 128, 127, 126, 125, 122, 121, 120, 119, 118, 113, 116, 117.

In the remaining groupings, the groups with centroids 101, 102, 103, 104 y 106, only contain as element the centroid itself.

From algorithm 1, we obtain the associated graph to each solution. The following figure 2 shows the result for the test run of 12 groups with algorithm 1 where it is observed the sought solution: the centroids are the cases just as it is described in the membership of the clusters. The structure of the graph obeys as well the coordinates (longitude and latitude) of the ovitraps and the cases. However, this result for 12 groups signifies a distribution that proposes that the ovitraps must be placed in association with the case (centroid) to which it corresponds as it is shown in the graph, in this way is possible to have a configuration that indicates the form in which the reported cases get surrounded by the most adequate ovitraps that favor the study of the Dengue exposed in this paper.
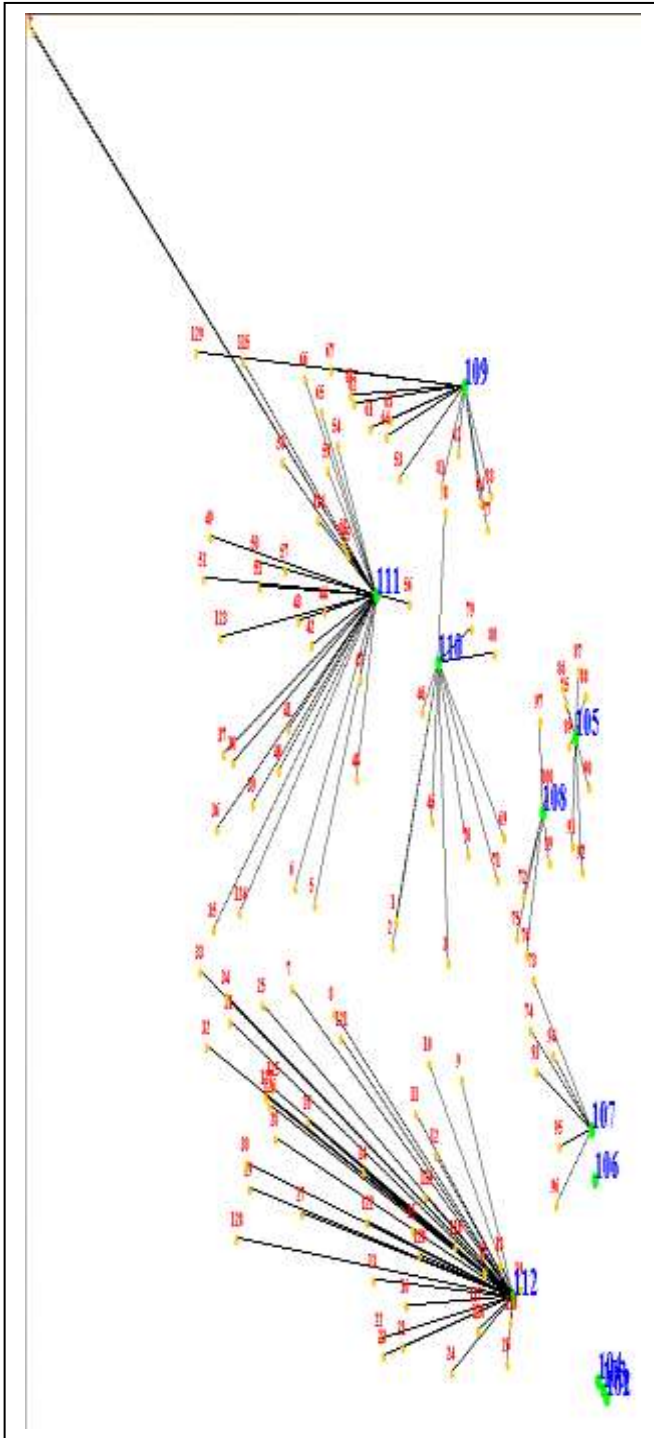
Figure 2. Ovitraps-cases Graph for Algorithm 1

Now the challenge consists in placing the ovitraps in accordance to a chosen grouping and observing the results.

VI. Adaptation of a K-Medoids Algorithmand clustering in the Dengue Problem

After different tests it was assumed that 12 groups could be a good configuration for the problem we've dealt with, however, the algorithm 1 being deterministic with the given medoids input vector, it only builds a partition based on this input vector being its weakness that the partition formed is only a local optimum, because the initial choice of medoids forbids the algorithm to look for better alternatives (lower cost solutions) for the given amount of groups.

At this point, knowing the best grouping for the chosen test run is a need, therefore we have supported on the aspects of k-medoids and PAM to modify the algorithm 1 to obtain the best centroids (medoids) among the 29 cases possible. The input of the algorithm is the dissimilarity matrix of cases and ovitraps and an input vector formed by the codes of the cases (restricted set of medoids).

On the other hand the number of groups can only be less or equal than 29. The algorithm restricts the search or swaps to the objects in the array of valid centroids received as input and thus returns as result the best solution with respect to distance minimization between cases and ovitraps, in contrast with P-median and PAM that provide an optimal solution from among all of the objects, this is, that even the ovitraps can be centroids giving a very good and optimal solution but impractical for this problem.

Based on PAM, the improvement of the algorithm 1 consists of a small alteration to achieve groupings with a maximum of 29 (cases) groups where the centroid is a case among over a hundred objects:

A. Clustering with Predetermined Centroids with K-medoids (Algorithm 2)

Input: C – Array of valid centroids
Input: k – Number of groups
Input: D – Dissimilarity matrix of size nxn

    1: Initialize: select k objects from C
    2: Associate each object and not centroid to the closest centroid
    3: for each centroid c do
    4:   for each element o of C do
    5:   exchange c with o and compute the total cost of the configuration
    6:   end for
    7: end for
    8: Select the lowest cost configuration
    9: Repeat steps 2 to 8 until there's no change in the centroids

Result for 12 groups whit algorithm 2
Number of groups: 12 Cost: 201.43918
Time: 777
Cluster 1: Size of Group 14, Centroid 105, Elements 85, 86, 87, 88, 89, 90, 91, 92, 97, 99, 100, 107, 108
Cluster 2: Size of Group 11, Centroid 112, Elements 17, 18, 19, 20, 24, 95, 96, 113, 118, 117
Cluster 3: Size of Group 2, Centroid 104, Elements 106
Cluster 4: Size of Group 11, Centroid 109, Elements 53, 63, 64, 67, 77, 81, 82, 83, 84, 129
Cluster 5: Size of Group 6, Centroid: 116 Elements 9, 12, 74, 93, 94,
Cluster 6: Size of Group 3, Centroid 119, Elements 10, 11,
Cluster 7: Size of Group 24, Centroid 121, Elements 1, 2, 5, 6, 7, 8, 14, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 127, 126, 124, 128, 125,
Cluster 8: Size of Group 14, Centroid 123, Elements 4, 36, 37, 38, 39, 40, 41, 49, 50, 51, 52, 98, 123,
Cluster 9: Size of Group 3, Centroid 101, Elements: 102, 103,
Cluster 10: Size of Group: 7, Centroid: 120 Elements 13, 15, 16, 21, 22, 23,

Cluster 11: Size of Group: 17, Centroid: 110 Elements 3, 45, 46, 48, 56, 69, 70, 71, 72, 73, 75, 76, 78, 79, 80, 111,
Cluster 12: Size of Group 17, Centroid 114, Elements 42, 43, 44, 47, 54, 55, 57, 58, 59, 60, 61, 62, 65, 66, 68, 115
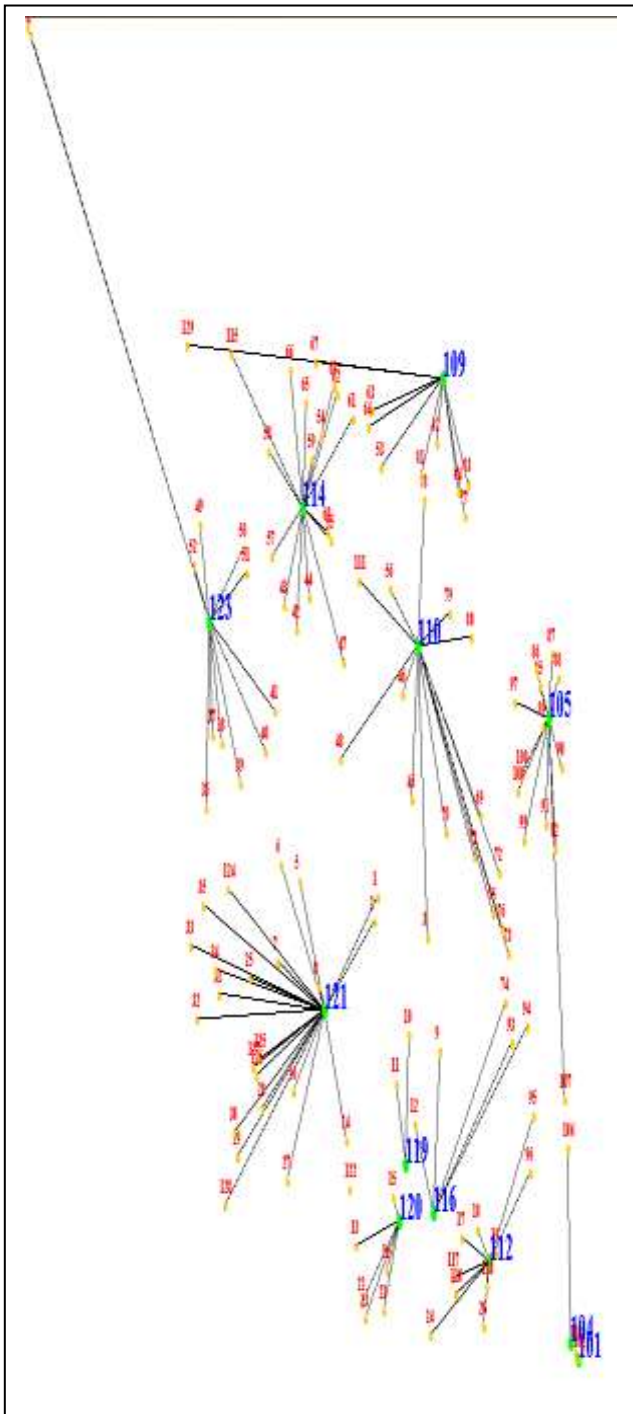

Figure 3.Graph for 12 groups in accordance with algorithm 2

PAM (Partitioning Around Medoids) is the partitioning algorithm that we have considered to incorporate the modifications of the algorithms 1 and 2 previously described. PAM being an algorithm that works over Medoids, has advantages with respect to other methods, for example, it is different to K-means because PAM builds an initial solution of k medoides or centroids, from which the medoides/centroids are swapped for the other elements, and the solutions that lower the total cost are kept whereas k-means works based on an initial solution randomly generated. This stage when the initial solution is generated is known as "BUILD-step" and the stage of exchange between selected objects (medoids) for unselected objects is known as "SWAP-step". This last stage is successively executed until the solution can't be improved anymore.

Algorithm PAM:
BuildStep();
cost ←objectiveFunction();
SwapStep();

PROCEDURE SwapStep()
REPEAT
FOR each medoid m
        FOR each non-medoid data point o   swap(m, o);
   newcost←objectiveFunction();
     IF newcost< cost
     save(medoids);
     END IF
   END LOOP
END LOOP
UNTIL there's no change in the medoid.
END PROCEDURE

The advantage that PAM offers with respect to the algorithms 1 and 2 resides in the optimality of the solution found:

Number of Groups: 12 Cost: 0.16410002
Time: 774
Size of Group: 10, Cluster 1: Centroid: 38 Elements: 5, 6, 36, 37, 39, 40, 49, 52, 123
Size of Group: 8, Cluster 2: Centroid: 95 Elements: 74, 93, 94, 96, 104, 106, 107
Size of Group: 16, Cluster 3: Centroid: 44 Elements: 41, 42, 43, 46, 47, 48, 50, 51, 55, 56, 57, 58, 60, 111, 114
Size of Group: 12, Cluster 4: Centroid: 71 Elements: 3, 45, 69, 70, 72, 73, 75, 76, 79, 80, 110
Size of Group: 19, Cluster 5: Centroid: 64 Elements: 53, 54, 59, 61, 62, 63, 65, 66, 67, 68, 77, 78, 81, 82, 83, 84, 109, 115
Size of Group: 17, Cluster 6: Centroid: 127 Elements: 7, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 127, 125, 126, 128
Size of Group: 13, Cluster 7: Centroid: 89 Elements: 85, 86, 87, 88, 90, 91, 92, 97, 99, 100, 105, 108
Size of Group: 13, Cluster 8: Centroid: 11 Elements: 1, 2, 121, 8, 9, 10, 122, 12, 14, 15, 120, 116
Size of Group: 14, Cluster 9: Centroid: 118 Elements: 13, 16, 17, 18, 19, 20, 21, 22, 23, 24, 112, 113, 117
Size of Group: 3, Cluster 10: Centroid: 102 Elements: 101, 103
Size of Group: 2, Cluster 11: Centroid: 129 Elements: 129
Size of Group: 2, Cluster 12: Centroid: 4 Elements: 98

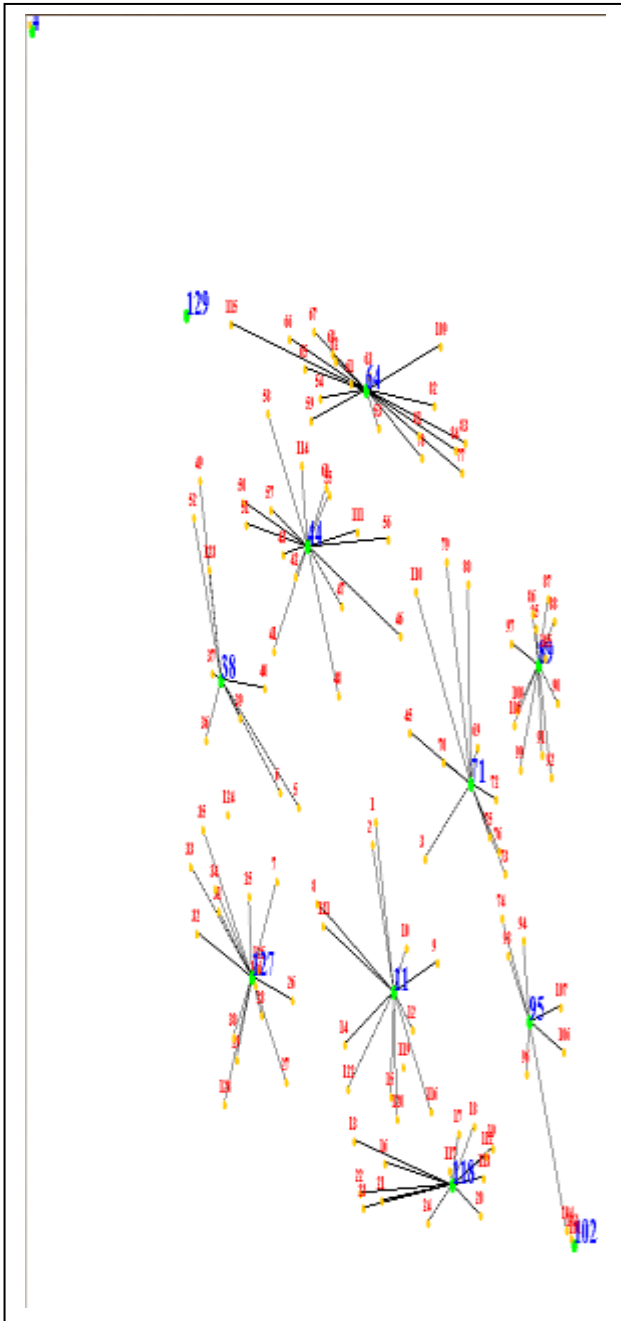The following figure 3 shows the generated graph.

Figure 3. Graph for 12 groups in accordance with PAM

In table 1 a test run for 12 groups is shown. The nomenclature is the following: CF (Function Cost), CT (centroids) y MD (Medoids).

TABLE I.   RESULTS FOR 12 GROUPS

Test runs for 12 groups with 3 algorithms					
PAM		ADAPTATION OF A P-MEANS		ADAPTATION OF A K-MEDOIDS	
CF	CT	CF	CT	CF	MD
0.1641	38 95 44 71 64 127 89 11 118 102 129 4	201.545	101 – 112	201.4391	105 112 104 109 116 119 121 123 101 120 110 114

TABLE II.   RESULTS FOR 24 GROUPS

Test runs for 24 with 3 algorithms					
PAM		ADAPTATION OF A P-MEANS		ADAPTATION OF A K-MEDOIDS	
CF	CT	CF	CT	CF	MD
0.08859	21 71 1 110 38 104 62 28 34 8 55 119 52 93 108 105 101 43 84 113 129 4 106 102	201.3711	101 – 124	201.267	122 113 109 102 101 112 117 118 110 114 103 104 120 126 129 127 115 123 119 108 111 121 116 106

As can be seen in table 1, the cost function of PAM is better due to the fact that the algorithm does an exhaustive search of Medoids from all of the objects, including cases and ovitraps, this is a disadvantage for this problem because several of the selected centroids are ovitraps.

The adaptation of p-median gives as result one unique solution where the centroids are only cases from 101 to 112 with the drawback that said solution is a local optimum with a cost function really far from the optimum and finally, the adaptation of k-medoids improves the cost of p-median.

The table 2 shows results for 24 groups, that were not shown explicitly in this work, however, just like the results for 12 groups, we can observe that the best solution is for PAM but it's not useful for the researchers of the Dengue, due to the fact that some of the selected centroids are ovitraps.

VII. Conclusions

In this work we have achieved adapting a well-known partitioning algorithm to the behavior of the Dengue mosquito with the end of reassigning ovitraps in correct places where a case of dengue has been registered. The grouping algorithm it's been supported by the works where bioinspired algorithms of mosquitos swarms have been proposed. The results of our algorithm produce adequate configurations to place the ovitraps in correct coordinates and thus continue with the study of the Dengue mosquito.

The algorithmic proposals are easy to implement, however, it has been necessary to know the aspects of partitioning over Medoids, P-median, and clustering in general to propose schemes that solve the problem of the Dengue that we have discussed [31]. The researchers interested in the association of the ovitraps with the reported cases have now 3 configurations available to study the best structure according to the mobility of the Dengue mosquito, where at least 3 aspects must be analyzed:

1. The best graphic homogeneous distribution of cases and ovitraps where the distance between them is minimum. In this case regular PAM offers the best solution but the researchers will have to decide if the centroids selected by PAM are useful due to the fact that several of them are ovitraps.

2. A simple structure of cases in the centroids where the solution is local but feasible, with the disadvantage that it is

not the best solution with respect to the value of the objective function.

3. In the same way than the previous one, but iterating and improving the algorithm 2, we have achieved a better cost function.

The results we have exposed in a graph, can be visible in a Geographic Information System GIS and from there is possible to count with geographic information that helps the researchers to make a decision in function of the results we have proposed.

On the other hand, due to the small data in this work, there haven't been problems in the computational complexity, however, a broader study, where the coordinates are greater in different states of the Mexican Republic implies the incorporation of heuristic methods in its solving.

Finally the importance of adapting the behavior of the mosquito to a real problem, has given place to propose a bioinspired clustering algorithm to solve the problem of practical configurations for the Dengue mosquito problem that has been described in this work.

Acknowledgment

References

[1] R.A Martínez-Vega, R Danis-Lozano, J. Velasco-Hernández, F.A Díaz-Quijano, M. González-Fernández, R.Santos, "A prospective cohort study to evaluate peridomestic infection as a determinant of dengue transmission: Protocol", *BMC Public Health*, pp. 12-262, 2012.

[2] J.A. Ruiz-Vanoye, O. Díaz-Parra, F. Cocón, A. Soto, M. A. Buenabad-Arias, G. Verduzco-Reyes, R. Alberto-Lira, "Meta-Heuristics Algorithms based on the Grouping of Animals by Social Behavior for the Traveling Salesman Problem", *International Journal of Combinatorial Optimization Problems and Informatics,* Vol. 3, No. 3, ISSN: 2007-1558, pp. 104-123, 2011.

[3] A. Okubo, "Dynamical aspects of animal grouping: Swarms, schools, flocks, and herds", *Advances in Biophysics*, Vol. 22, pp. 1-94, 1986.

[4] M. Dorigo, "Optimization, Learning and Natural Algorithms", Ph.D. Thesis, *Politecnico di Milano*, Italy, 1992.

[5] X.S Yang, "Firefly Algorithm", chapter 8, Nature-inspired Metaheuristic Algorithms, *Luniver Press*, British Library Cataloguing-in-Publication Data, British Library, 2008.

[6] H.A Abbass, "MBO: Marriage in Honey Bees Optimization-a Haplometrosispolygynous Swarming Approach", Proceedings of the 2001 Congress on Evolutionary Computation, vol.1, 2001, pp. 207-214.

[7] D. Karaboga, "An idea based on Honey Bee Swarm for Numerical Optimization", *Technical Report-TR06*, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.

[8] P.C. Pinto, T. A Runkler, T.A and J. M Sousa, "Wasp Swarm Algorithm for Dynamic MAX-SAT Problems". *In Proceedings of the 8th international conference on Adaptive and Natural Computing Algorithms*, Part I, 2007.

[9] L. Xueyan and Z. Yongquan,: "A Novel Global Convergence Algorithm: Bee Collecting Pollen Algorithm", Advanced Intelligent Computing Theories and Applications, With Aspects of Artificial Intelligence, Lecture Notes in Computer Science , Vol. 5227, pp. 518-525, 2008.

[10] M. Roth, "Termite: A Swarm Intelligent Routing Algorithm For Mobile Wireless Ad-Hoc Networks", *Theses and Dissertations, Cornell University Graduate School*, 2005.

[11] J.A. Ruiz-Vanoye O. and Díaz-Parra, "A Mosquito Swarms Algorithm for the Traveling Salesman Problem", *unpublished manuscript*.

[12] J.A. Ruiz-Vanoye and O. Díaz-Parra, "A zooplankton Swarms Algorithm for the Traveling Salesman Problem", *unpublishedmanuscript*.

[13] J.A. Ruiz-Vanoye and O. Díaz-Parra, "A Bumblebees Swarms Algorithm for the Traveling Salesman Problem", *unpublished manuscript*.

[14] D.T. Mourya, Gokhale, A. Basu, P. V. Barde, G. N. Sapkal, V. S. Padbidri, M. M. Gore, "Horizontal and vertical transmission of dengue virus type 2 in highly and lowly susceptible strains of Aedesaegypti mosquitoes", *ActaVirol*, Vol. 45, pp. 67-71, 2001.

[15] V. Joshi, D.T. Mourya, R.C. Sharma, "Persistence of dengue-3 virus through transovarial transmission passage in successive generations of Aedesaegypti mosquitoes". *Am J. Trop Med Hyg*, Vol. 67, pp. 158-61, 2002.

[16] N. Arunachalam, S. C. Tewari, V. Thenmozhi, R. Rajendran, R. Paramasivan, R. Manavalan, "Natural vertical transmission of dengue viruses by Aedesaegypti in Chennai", *Tamil Nadu, India, Indian J. Med Res*, Vol. 127, pp. 395-397, 2008

[17] G. Le Goff, J. Revollo, M. Guerra Cruz, Z. Barja Simon, Y. Roca, "Natural vertical transmission of dengue viruses by Aedesaegypti in Bolivia", Vol. 18, pp. 277-280, 2011.

[18] J. Günther, J.P. Martínez-Muñoz, D.G. Pérez-Ishiwara, Salas-Benito, "Evidence of vertical transmission of dengue virus in two endemic localities in the state of Oaxaca", *Mexico Intervirology*, Vol. 50, pp. 347-52, 2007.

[19] A.B. Cecílio, E.S. Campanelli, K.P. Souza, L.B. Figueiredo, M.C. Resende, "Natural vertical transmission by Stegomyiaalbopicta as dengue vector in Brazil", *Braz J. Biol*, Vol. 69, pp.123-127, 2009.

[20] A.P. Vilela, L.B. Figueiredo, J.R. Dos Santos, A..Eiras, C.A. Bonjardim, P.C. Ferreira, E.G. Kroon, "Dengue virus 3 genotype I in Aedesaegypti mosquitoes and eggs",*Brazil, 2005-2006. Emerg Infect Dis*., Vol. 16, pp. 989-992, 2010.

[21] WHO, TDR, "Dengue guidelines for diagnosis, treatment, prevention and control", 2009.

[22] N.A. Honório, C.T. Codeço, F.C. Alves, M.A. Magalhães, R. Lourenço-De-Oliveira, "Temporal

distribution of Aedesaegypti in different districts of Rio de Janeiro, Brazil, measured by two types of traps", *J. Med Entomol*, Vol. 46, pp. 1001-114, 2009.

[23] C.C. Marques, G.R. Marques, M. de Brito, L.G. Santos Neto, V. de C. Ishibashi, F. de A. Gomes, "Comparative study of larval and ovitrap efficacy for surveillance of dengue and yellow fever vectors", *Rev SaudePublica*, Vol. 27, pp. 237-241, 1993.

[24] E.L. Estallo, F. F. Ludueña-Almeida, A. M. Visintin, C.M. Scavuzzo, M.V. Introini, M. Zaidenberg, W. R. Almirón, "Prevention of dengue outbreaks through Aedesaegyptioviposition activity forecasting method", *Vector Borne Zoonotic Dis*, Vol. 11, pp. 543-549, 2011.

[25] L. Regis, A.M. Monteiro, M.A. Melo-Santos, J.C. Silveira, A.F. Furtado, R.V. Acioli, "Developing new approaches for detecting and preventing Aedesaegypti population outbreaks: basis for surveillance, alert and control system", *MemInstOswaldo Cruz,* Vol. 103, pp. 50-59, 2008.

[26] Kaufman L., Rousseeuw P. J.,"Clustering by means of medoids", Statistical Data Analysis based on the L1 Norm, North-Holland, Amsterdam , pp. 405-416.,1987

[27] M.S. Dasking, "Network Discrete Location, Models, Algorithms, and Applications", *John Wiley & Sons*, pp. 198-208, 1985.

[28] R. L. Church, "COBRA: A New Formulation of the Classic p-Median Location Problem", Annals of Operations Research 122, 2003, pp. 103-120.

[29] J. Reese, "Solution methods for the p-median problem: An annotated bibliography", Networks, october 2006, vol. 48, no. 3, pp. 125-142.

[30] O. Alp and E. Erkut, "An Efficient Genetic Algorithm for the p-Median Problem", Annals of Operations Research, 2003, 122, pp. 21-42.

[31] Z. Detorakis, G.Tambouratzis. **"**Clustering Techniques for Establishing Inflectionally Similar Groups of Stems", *International Journal of Computer Information Systems and Industrial Management Applications*, ISSN 2150-7988, Vol. 4, pp. 219-227, 2012.

Author Biographies

MaríaBeatríz Bernábe Loranca was born in the city of Puebla, Mexico. He received the B.S. degree in Computer Science from Benemérita Universidad Autónoma de Puebla (BUAP), Mexico. In 2010, she received the Doctorate degree in Operations Research from the Universidad NacionalAutónoma de México (UNAM). Since 1995, she has been a professor at the School of Computer Science of BUAP, where she works in databases and statistics. She belongs to the National System of Researchers with Level Candidate (SNI). Her research interests are: combinatorial optimization, territorial design and multiobjective techniques

Marco Antonio Rodríquez Floreswas born in the city of Acapulco, Mexico. He received the B.S. degree in Computer Science from Benemérita Universidad Autónoma de Puebla (BUAP), Mexico. Her research interests are: combinatorial optimization, clustering, Data Base and Metaheuristics.

Elias Olivares Benítez Professor in the areas of Supply Chain Management, Industrial Engineering, and Materials Engineering. His research interests are in modeling and optimization of logistics, manufacturing, supply chain, and service systems, with multiple objectives. He has used Scatter Search, Tabu Search, GRASP, CPLEX, and hybridizations. He is member of the Institute for Operations Research and the Management Sciences, the Mexican Society for Operations Research, and the Mexican Association of Logistics and Supply Chain.

David Eduardo Pinto Avendaño obtained his PhD in computer science in the area of artificial intelligence and pattern recognition at the Polytechnic University of Valencia, Spain in 2008. At present he is a full time professor at the Faculty of Computer Science of the Benemérita Universidad Autónoma de Puebla (BUAP). His areas of interest include clustering, information retrieval, crosslingual NLP tasks and computational linguistics in general.

José Luis Martínez Flores is professor and researcher in logistic network design and belongs to SNI, level 1 (National Sistem of Researchers). He works as academic coordinator in the Logística y Dirección de la Cadena de Suminstro grade, at Postgraduate Interdisciplinary Center, Research and Consultancy of the Universidad Popular Autónoma del Estado de Puebla. As a guest has issuedconferences in differentinstitutions and fórums such as Universidad de Bogotá Jorge Tadeo Lozano, Universidad Autónoma de Nuevo León, Universidad Del Caribe, Universidad Autónoma de Aguascalientes and Corporación Mexicana de Investigación en Materiales (COMIMSA).