Locating Seismic-Sense Stations Through Genetic Algorithm

Genetic algorithms

Josafath I. Espinosa-Ramos
Intelligent Systems Group
Faculty of Engineering - La Salle University
Benjamín Franklin 47 Col. Condesa CP 06140
México, D.F.
vjier@prodigy.net.mx

Roberto A. Vázquez
Intelligent Systems Group
Faculty of Engineering - La Salle University
Benjamín Franklin 47 Col. Condesa CP 06140
México, D.F.
ravem@lasallistas.org.mx

ABSTRACT

Recent studies warn of a possible major earthquake off the coast of State of Guerrero, Mexico, so that, it turns important to alert the population as long as possible and avoid a great disaster. This requires the construction of a network of seismic sensing stations, located at strategical positions, to detect earthquakes and issue a timely warning. In this research, we investigate how a genetic algorithm can be applied to design this network and determine the optimal location of each seismic sensing station. The number of earthquakes detected by the designed network will be used as a reference point with respect to the currently installed seismic alert system (SAS). This metric will justify the use of the genetic algorithms as a designing tool prior to the construction of the network in different regions of Mexico. The SAS stations and each solution proposed by a genetic algorithm underwent a procedure, in which it is simulated the occurrence of earthquakes obtained from the Mexico's National Seismological Service (SSN) database, to determinate its efficiency in terms of the time to warn Mexico City.

Categories and Subject Descriptors

I.2.1 [**Artificial Intelligence**]: Applications and Experts Systems—*Medicine and science*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search Heuristic methods; J.2. [**Computer application**]: Physical sciences and engineering Earth and atmospheric science

General Terms

Design, Reliability

Keywords

Genetic algorithm, Design of seismic sensor networks, Pattern recognition

GECCO'11, July 12–16, 2011, Dublin, Ireland.

1. INTRODUCTION

Historical information and studies about the seismic activity on the coast of Guerrero State prove that exist a seismic silence between the ports of Zihuatanejo and Acapulco [7]. In this region, an earthquake could occur in similar proportions to that which occurred in 1985 making an extensive damage in Mexico City.

Actually, the seismic alert system (SAS) has 12 seismic stations located along the coast of Guerrero. The primary function of the SAS is to issue a public warning to Mexico City when it detects an earthquake of magnitude greater than 5.0° on the Richter scale. This system is capable of alerting the population up to 60 seconds before the seismic wave reaches Mexico City. It gives us the opportunity to execute procedures and actions that reduce the possibility of having a new earthquake disaster.

Several studies about seismology and geodesy have been published in the last years. Most of them are focused on the analysis of the spread and magnitude of seismic waves [2][5]. Some of these researches are related to the measurement and the instrumentation design [1][3]. Moreover, new techniques based on the statistics and computing intelligence have been focused to the prediction of earthquakes [11][10]. Particularly, the use of neural networks and evolutionary strategies have gain attention in the recent years[12]. Although all these works are of great interest and importance, it is also necessary a designing phase to determine the strategical location and construction of sensor stations aimed to detect earthquakes and prevent the population as fast as possible.

This research focuses on the optimal design for a network of seismic stations finding their best locations through a genetic algorithm that allows us to search the optimal solution to the problem. Subsequently, the solutions provided by the genetic algorithm will be compared against the actual seismic alert system and validated in terms of the warning time.

Sections 2 and 3 will describe some basic concepts needed to solve the problem above. Section 4 describes in details how the solution to the problem is coded into the chromosome (individual) of the genetic algorithm. In the same section, it is also described the proposed objective function which models the solution to our problem. This is followed by the section 5 which presents the experimental results obtained with the proposed method and the simulations to determine its efficiency. Finally, the paper is concluded with some results and future work.

2. GENETIC ALGORITHM

The genetic algorithm leads to the search for an optimal solution to a problem, inspired by inheritance mechanisms observed in nature. This heuristic process keeps the set of solutions (individuals or chromosomes) called population in genetic terms. During each iteration (generation) of the algorithm, the performance (fitness) of all solutions of the population is measured by the objective function f that evaluates a particular problem. Then, some solutions are selected from the population (parents) to create the next generation of solutions. This selection depends on the values of f and can follow several schemes such as the elitist or roulette selection among the most popular [4][8]. The selected solutions undergo a series of combinations, usually consisting of the random exchange of certain parts of the parents. In this process, the useful features of parent solutions are preserved. Thereafter, children were randomly chosen to undergo a mutation. The sequence of evaluation, selection and recombination is repeated until an individual has a satisfactory value for f or until a predefined number of generations is reached. A simple genetic algorithm, algorithm 1, is presented in its algorithmic form.

Algorithm 1 Simple genetic algorithm

1: Choose the initial population of individuals
2: Evaluate the fitness of each individual in that population
3: **repeat**
4: Select the best-fit individuals for reproduction
5: Breed new individuales through crossover and mutation operations to give birth to offspring
6: Evaluate the individual fitness of new individuals
7: Replace least-fit population with new individuals
8: **until** termination (time limit, sufficient fitness achieved, etc.)

3. LINEAR DISCRIMINANT FUNCTIONS

As we will describe in the next sections, it is important to evaluate the location of the stations and validate a feasible region for its construction. This means that we have to use a method that allows us discriminate between feasible and infeasible regions to construct a station. Nowadays, exist several methods that allow discriminates among regions. In our case, we decided to adopt the well-known linear functions widely used in pattern recognition.

The main function of a pattern recognition system is to take decisions on membership of a class of patterns. An easy way to separate a pattern from another is drawing a straight line that divides them and generates a single decision surface.

Following the straight-line equation 1, we could observe that if a pattern is substituted in $d(x)$ and a positive value is obtained, then the pattern belongs to the c_i class; otherwise, a negative value is obtained.

$$d(x) = w_1 x_1 + w_2 x_2 + w_3 = 0 \tag{1}$$

where w_i are the line parameters and x_1, x_2 are the variables of general coordinates.

These principle can be generalized to the n-dimensional case $d(x) = w'x$ where $x = (x_1, x_2, \cdots, x_n, 1)'$ is called an increased pattern vector and $w = (w_1, w_2, \cdots, w_n, 1)'$ is the weight vector.

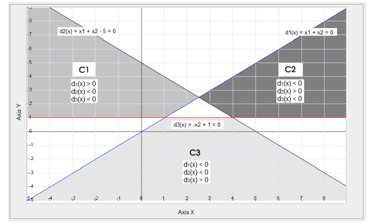

Figure 1: Decision regions that classify 3 classes

Therefore, for the case of identification of two classes, the decision function has the property:

$$d(x) = w'x \begin{cases} > 0 \ if \ x \ \in c_1 \\ < 0 \ if \ x \ \in c_2 \end{cases}$$

Figure 1, shows a three classes identification sample.

4. METHODOLOGY

Since have been explained some general concepts, this section of the article will be devoted to the methodology used to determine the location of a set of seismic stations by means of the genetic algorithm. We will start describing the chromosome representation and after that, we well describe the proposed objective function to solve our problem.

4.1 Chromosome representation

First of all, it must be clear that the solution to our problem is the location of each seismic station. For that reason, the solution (individual) is composed of n pairs of genes where n is the number of seismic-sense stations. Each pair of genes represents the location of a station in terms of its longitude $G_n x$ and latitude $G_n y$. Figure 2 shows the chromosome and gene representation.

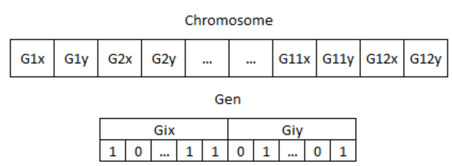

Figure 2: Graphical representation of chromosome

The value (phenotype) of each gene will have real representation in binary encoding, and shall be a vector of bits (alleles). This value is calculated using the equation 2 to encode from binary to real values:

$$v = \frac{(v_{max} - v_{min})}{(2^n - 1)} \cdot v_{bin} + v_{min} \tag{2}$$

where v_{max} and v_{min} are the maximum and minimum gene values, n is the number of bits (alleles) and v_{bin} the real representation in binary encoding (phenotype). The range of values of the genes, correspond to the geographical coordinates of a specific region.

4.2 Objective function

The main idea is to design a configuration where each station covers the major number of epicenters. Therefore, the objective function has to maximize the fitness of each individual assessed by the sum of the number of detected epicenters.

To compute the fitness of the individual, we first determine whether the epicenters are within the coverage of the station's area using the geographical distance.

To calculate the geographical distance between two points on the earth's surface, we apply the cosine theorem of spherical geometry [6] and the proportion of 360° to 40000 kms. (Ecuador diameter). This gives the approximated equation:

$$d_g(p,p') = \frac{40000}{360}\cdot \cdot \arccos\left(\cos\theta\cdot\cos\theta' + \sin\theta\cdot\sin\theta'\cdot\cos\left(\lambda-\lambda'\right)\right) \quad (3)$$

where p is the first point, θ the latitude of p, λ the longitude of p, p' the second point, θ' the latitude of p' and λ' the longitude of p'.

Let $g_i \in \mathbb{R}^2$ be the location of a station i, and $\alpha_j \in \mathbb{R}^2$ be the location of an epicenter j. To determinate if an epicenter is into the area covered by a station, we have to calculate the geographical distance between the epicenter and the desired station. If the distance is less than the radius of the covered area, then the epicenter is detected by the station. This relationship will be represented by the equation 4

$$h_{ij} = \begin{cases} 1 \text{ if } d_g(g_i,\alpha_j) < \rho \\ 0 \text{ otherwise} \end{cases} \quad (4)$$

where $i,j \in \mathbb{N}$, $i \neq j$, $i = 1,...,n$, $j = 1,...,r$, n is the number of stations, r is the total number of epicenters and ρ is the radius of the area covered by the station.

The number of epicenters detected $s_i \in \mathbb{N}$ by each station i is computed using the equation 5

$$s_i = \sum_{j=1}^{r} h_{ij} \quad (5)$$

where r is the total number of epicenters.

Finally, the individual's fitness $f(x)$ is the sum of epicenters detected by whole stations, which is computed by means of the equation 6

$$f(x) = \sum_{i=1}^{n} s_i \quad (6)$$

where n is the number of pairs stations.

In most optimization problems, the objective function is subject to constraints. For this research, we include two constrains: the first one is related to the overlap between the areas covered by the stations, and the second one for the event that any station is located in a not desired area as at the Pacific Ocean.

These constraints directly affect the fitness of individuals. The idea is to extend the domain of the objective function, which will be affected according to the equation 7:

$$f(x) = f(x)\cdot\prod_{i=1}^{n} c_i(x) \quad (7)$$

for $i = 1,2,3,...,n-1,n$ where n is the total number of constraints and $c_i(x)$ is a constraint. In this particular case, we have two constraints, then the next equation 8 will be applied:

$$f(x) = f(x)\cdot\delta(x)\cdot\gamma(x) \quad (8)$$

where $\delta(x)$ and $\gamma(x)$ are penalty functions corresponding to the first and second constraints respectively.

As it was described above, the objective function maximizes the number of epicenters detected by the individual. This would mean that geographical distance among stations must be greater to cover a larger area and thus provoke that more epicenters will be detected. On the contrary, if the geographical distance among stations is less than a specific radius, then two or more stations could detect the same epicenters, which does not help to maximize the individual's fitness.

Therefore, the first constraint is the condition in which the geographical distance between stations is greater than the double coverage of a station's area $d_g(p_i,p_j) > 2\rho$ for $i,j = 1,2,3,...,n-1,n$ where n is the stations' number, $i \neq j$ and ρ is the radius of the area covered by a station.

Considering that the geographical distance between some SAS stations is little less than 20 km, at least a radius of 10 km between stations was chosen. Hence, if the distance between some of the stations does not meet that condition, the death penalty is applied to the individual, assigning a value of zero to its fitness and declaring it not feasible for the solution. Otherwise, the fitness will be one.

$$\delta(x) = \begin{cases} 0 \text{ if } d_g(p_i,p_j) < 2\rho \\ 1 \text{ otherwise} \end{cases} \quad (9)$$

The second constraint will be applied to prevent that any of the stations are located in a reject region where it is not feasible for its construction; for example, when a station is located at the ocean.

To determine the rejection region, 60 geographical points along the coast of Guerrero and a point called the origin located in the ocean, were taken. The location of this origin point is -103.0° longitude and 16.0° latitude, which corresponds to lower limits of the State of Guerrero.

As it was previously shown, a decision region with linear functions is defined by a set of straight lines R. Knowing that the straight-line equation through two points $P_0(X_1,Y_1)$ and $P_i(X_i,Y_i)$ where P_0 is the origin and P_i any point limiting the coast of Guerrero, the set of lines that defines the reject region is $R = \{r_1(P_0,P_i), r_2(P_0,P_{i+1}), r_3(P_i,P_{i+1})\}$ for $i = 1,2,3,...,n-2,n-1$ where n is the number of points that limits the coast.

Therefore, a point $p(x_1,x_2)$ will be located in the reject region when it is evaluated in each $r_i \in R$, in its general form $r = Ax_1 + Bx_2 + c = 0$, and satisfies the condition $r_1(p) > 0 \wedge r_2(p) < 0 \wedge r_3(p) > 0$.

Finally, the individual will be penalized according to

$$\gamma(x) = \frac{m}{n} \quad (10)$$

where m is the number of stations that are in the reject region and n the total number of stations.

5. EXPERIMENTAL RESULTS

We performed several experiments to test the effectiveness of the network provided by the genetic algorithm. We vary

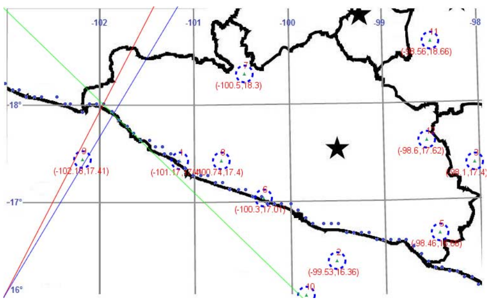

Figure 3: Detection of a station belonging to a reject region

the amount of earthquakes, the crossover and the mutation rates during the designing phase.

These experimental results are divided into three parts. The first one will focus on the settings and conditions of the genetic algorithm. The second one will describe the conditions to simulate all the earthquakes. And the last one shows the results and some observations about the behavior of the genetic algorithm. Particularly, we perform all experiments using the earthquakes registered from 1998 to 2006 in the State of Guerrero, Mexico.

5.1 Genetic algorithm settings

We pretend to use the genetic algorithm to cover as much territory as possible, considering the greater amount of earthquakes around the covered area of the proposed stations.

The range values used for genes are from $-102.11\,^\circ$ to $-98.00\,^\circ$ longitude and from $16.19\,^\circ$ to $18.53\,^\circ$ latitude. These values correspond to the geographical coordinates of the State of Guerrero and thus limits the stations to be located in this range of coordinates.

To ensure that the fittest individuals of the population survive and continue evolving to the next generation, we chose the elitist selection. In this case 20% of the individuals of the population with the highest scores will continue in the process of evolution considering the penalties on the constraints.

In previous section was indicated that the alleles have a binary encoding, for that reason n crossing points crossover method was used and for simplicity, one point crosses was selected. The crossover rates used were 90%, 70% and 50% for exploring the search space. With each crossover rate, we used the 1%, 3%, 5% and 10% of mutation rate to generate solutions that the cross cannot produce and thus achieve the optimum value. The number of generations done by the genetic algorithm was set to 200.

The epicenters used to evaluate individuals in the objective function, correspond to all the earthquakes recorded in México from 1998 to 2006, for a total of 5974. We selected the 2539 records belonged to the territory of interest located between $-103.0\,^\circ$ and $98.0\,^\circ$ longitude and $16.0\,^\circ$ to $19.0\,^\circ$ latitude. This information was obtained from the database of Mexico's National Seismological Service (SSN) [9].

In order to have almost the same conditions in which SAS was designed, we choose the 100%, 50% and 10% from the 2539 earthquakes (the entire sample space), for the training phase of the genetic algorithm. It is worth mentioning that this research does not pretend to discredit the labor of building SAS stations, but to have a real parameter to compare and demonstrate that this work could be feasible.

Based on the parameters already described, 36 configurations were created to run the genetic algorithm. In order to observe the behavior of the genetic algorithm and select the best optimal configuration, we tested over 20 times each configuration.

In addition, we also compare the genetic algorithm against a simple greedy search. For that purpose, one more group of 20 tests was done to observe if the greedy search produced a better configuration than either the SAS or the genetic algorithm.

5.2 Settings for simulation

In this section, we compare the results of 20 solutions obtained using the proposed genetic algorithm against the current SAS configuration.

To determine the efficiency of the algorithm in terms of the time to alert Mexico City, the SAS stations and each solution proposed by the genetic algorithm underwent into a procedure in which the occurrence of the 2539 earthquakes was simulated. We obtain the warning time computing the difference between the time that the seismic stations detect an earthquake and the time that the seismic wave reaches Mexico City. This approach was subject to the next considerations during the simulation process:

1. The same 2539 earthquakes registered in the database from 1998 to 2006 were used.

2. The speed of propagation of seismic waves varies between 4 km/s and 8 km/s. Fortunately, most energetic waves are transmitted to the lower speed, so we propose the following proportions.

   (a) If an earthquake has a magnitude greater than $5.0\,^\circ$ Richter, is propagated at a speed of 4 km/s.

   (b) For an earthquake of magnitude between $4.0\,^\circ$ and $5.0\,^\circ$ Richter will propagate at a speed of 6 km/s.

   (c) And for an earthquake of magnitude less than $4.0\,^\circ$ Richter, propagation velocity will be 8 km/s.

3. To determine the number of stations that detects an earthquake and issue a warning, the following rules were proposed.

   (a) For an earthquake of magnitude lower than $4.0\,^\circ$ Richter was determined that at least three stations have to detect the earthquake in order to reduce the likelihood that the sensor is stimulated at a station for reasons unrelated to seismic activity, possibly caused by heavy goods vehicles that cause a movement of the earth. This means that when the seismic wave reaches the third nearest station the epicenter, is the time to issue a warning.

   (b) In the event that the magnitude of the earthquake is between $4.0\,^\circ$ and $5.0\,^\circ$ Richter and continuing with the above criteria, at least two stations must detect the occurrence of the earthquake. In this case when the seismic wave reaches the second nearest station, a warning is issued.

   (c) Finally, if the earthquake's magnitude is greater than $5.0\,^\circ$ Richter (highly risky), it is enough that

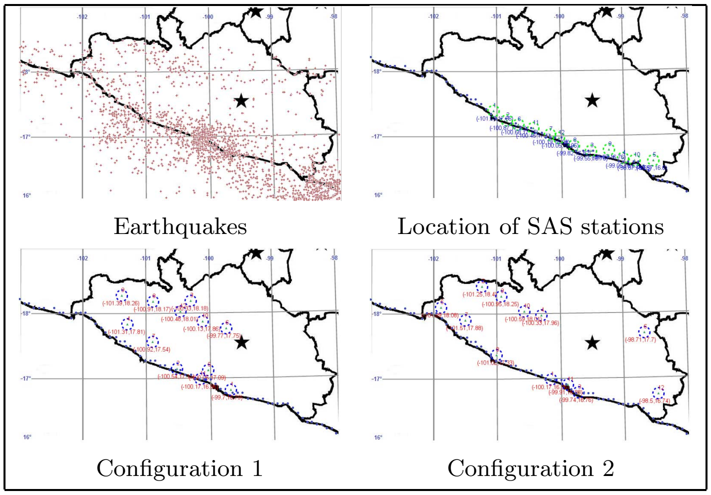

Figure 4: Configurations proposed by genetic algorithm

the first nearest station detects it and issues a warning. In this case, there is a low probability that the sensor is stimulated by activities other than seismic.

The speed formula $v = \frac{d_g(pp')}{t}$ was used to calculate the time that a station issues a warning signal and the time for seismic waves spreads and reaches Mexico City.

5.3 Results

The solutions obtained with the genetic algorithm show that the stations are divided into a northern region and southern region along the coast, locating them where most of the earthquakes are concentrated, similar to earthquakes recorded. Figure 4 shows the current distribution of SAS and some of the solutions proposed by the genetic algorithm.

It was observed that, when the algorithm starts, the number of earthquakes detected and the individual's fitness was different. This occurred because some stations were located at the Pacific Ocean and the fitness was affected by a penalty caused by the second constraint. In average, until the 31th generation, all the stations are located within the feasible region. After this phase, the genetic algorithm continues with the optimization of the individual's fitness. In average, the optimum value was reached after the 97th generation. Figure 5 shows the behavior of the genetic algorithm.

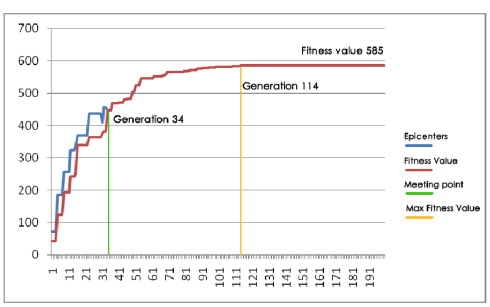

Figure 5: Genetic algorithm's behavior

On the other hand, the greedy algorithm has a not homogenous behavior and locates the stations along the territory of the State of Guerrero without a pattern or a specific order. Figure 6 shows the behavior of greedy algorithm.

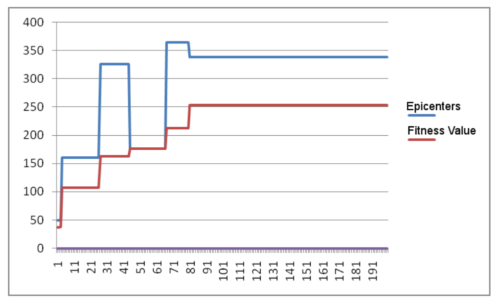

Figure 6: Greedy algorithm's behavior

The following tables show the results obtained by simulating the earthquakes in the current stations of the seismic alert system (SAS), in the stations proposed by the genetic algorithm and finally in the stations proposed by the greedy algorithm.

In the simulation performed with the entire sample space, the current SAS configuration gives the following results shown in table 1.

Table 1: Evaluation for SAS configuration

All earthquakes		Earthquakes>5.0 °		Earthquake
Maximum	Average	Maximum	Average	1985
80	34	80	65.23	57

These results will be taken as a starting point to compare the results provided by the genetic and greedy search algorithms.

After the simulation for all earthquakes recorded of any magnitude, the maximum time value achieved by the best configuration was 99 seconds and an average time of 31.32 seconds. These results are shown in tables 2, 3 and 4.

In a comparison against the results obtained by SAS, we can observe that the maximum time is 19 seconds over the maximum time registered by SAS, but in the average time,

Table 2: Evaluation of the training phase using the 100% of the total sample space (2539 records) for all earthquakes

Crossover rate	Mutation rate	Maximum warning time	Average warning time
90%	1%	98	31,51
	3%	88	31,7
	5%	88	31,78
	10%	98	32,18
70%	1%	97	31,87
	3%	98	31,53
	5%	98	31,66
	10%	99	31,47
50%	1%	96	30,92
	3%	98	31,04
	5%	98	31,09
	10%	96	31,38

Table 3: Evaluation of the training phase using the 50% of the total sample space (1269 records) for all earthquakes

Crossover rate	Mutation rate	Maximum warning time	Average warning time
90%	1%	98	31,72
	3%	96	31,96
	5%	98	31,34
	10%	97	31,4
70%	1%	88	31,53
	3%	95	31,73
	5%	99	31,43
	10%	97	31,55
50%	1%	99	31,08
	3%	86	30,66
	5%	95	31,39
	10%	96	30,28

Table 4: Evaluation of the training phase using the 10% of the total sample space (253 records) for all earthquakes

Crossover rate	Mutation rate	Maximum warning time	Average warning time
90%	1%	94	31,08
	3%	87	31,34
	5%	98	31,46
	10%	99	31,15
70%	1%	95	30,68
	3%	91	31,34
	5%	96	30,87
	10%	99	31,48
50%	1%	93	30,57
	3%	88	31,24
	5%	94	30,62
	10%	97	31,36

Table 5: Evaluation of the training phase using the 100% of the total sample space (2539 records) for earthquakes greater than 5 richter

Crossover rate	Mutation rate	Maximum warning time	Average warning time
90%	1%	98	67,83
	3%	88	68,76
	5%	88	67,19
	10%	98	68,4
70%	1%	97	68,88
	3%	98	68,88
	5%	98	69,24
	10%	99	68,15
50%	1%	96	70,2
	3%	98	67,6
	5%	98	68,75
	10%	96	67,94

SAS configuration is better than the stations proposed. Although these results could be significant, they consider the earthquakes with a magnitude less than 5.0 ° Richter, which are not of interest on this research, since SAS does not issue a warning when earthquakes of that magnitude occur.

In tables 5, 6 and 7 we show the results for a simulation with the earthquakes greater than 5.0 ° Richter.

From these tables we can observe that the maximum and the average times for all experiments in each configuration are similar to each other. Furthermore, it is observed that with a rate of 70% for crossover and a rate of 5% or 10% for mutation, it was achieved a maximum value of 99 seconds to alert the population. This behavior remains constant even using a small number of samples during training phase. Considering all of this, the best results are achieved using the 50% of the samples, a crossover rate of 70% and a mutation rate of 5%. These results are interesting for this research, because the seismic alert system must issue a warning alarm when an earthquake over 5.0 ° Richter has occurred. The warning time is 19 seconds over the current SAS configuration, in other words, 23% more effective.

The last set of experiments performed to evaluate the accuracy of the genetic algorithm includes the simulation of the earthquake occurred in Mexico City, which caused a major disaster in 1985. These results are shown in tables 8, 9 and 10.

Once again, we observed that the genetic algorithm provides a better time. Here the maximum time value achieved was 90 seconds. These are 33 seconds and 57% more effective than current SAS. In addition, we can see that as in all cases before, the settings for the best results of the genetic algorithm are a crossover rate of 70%, and in general any mutation rate is good.

Finally, we perform a comparison against the greedy algorithm. Basically, this algorithm produces individuals with random values, but some of the stations could be located in a not feasible region to build them, as in the ocean. To get a more equitable way for comparison, we considered only those

Table 6: Evaluation of the training phase using the 50% of the total sample space (1269 records) for earthquakes greater than 5 richter

Crossover rate	Mutation rate	Maximum warning time	Average warning time
90%	1%	98	67,97
	3%	96	67,79
	5%	98	68,31
	10%	97	67,17
70%	1%	88	67,67
	3%	95	68,49
	5%	99	68,94
	10%	97	67,77
50%	1%	99	67,93
	3%	86	67,71
	5%	95	66,8
	10%	96	68,82

Table 7: Evaluation of the training phase using the 10% of the total sample space (253 records) for earthquakes greater than 5 richter

Crossover rate	Mutation rate	Maximum warning time	Average warning time
90%	1%	94	66,53
	3%	87	66,82
	5%	98	67,54
	10%	99	67,64
70%	1%	95	66,94
	3%	91	67,86
	5%	96	67,1
	10%	99	67,11
50%	1%	93	67,38
	3%	88	65,98
	5%	94	67,27
	10%	97	67,78

stations that were within the State of Guerrero to evaluate the effectiveness of this algorithm. Moreover, we test with the entire sample space of earthquakes, 100 individuals and 200 iterations. These are the same parameters used in the genetic algorithm.

The table 11 show the results obtained with the greedy algorithm.

Certainly, the results seem to be good, since they are better than the current SAS. However, it is important to mention that in all tests, any configuration produced all stations in a feasible region.

As we mention in 5.2, the behavior of the greedy algorithm is no homogenous, this means that most of the earthquakes detected were concentrated in only one station while other stations detect few or no earthquakes.

A final result summary is shown in table 12.

This results indicates that in an earthquake greater than 5.0 ° Richter or in an earthquake like that which occurred in 1985 (7.0 ° Richter), a configuration proposed by the genetic

Table 8: Evaluation of the training phase using the 100% of the total sample space (2539 records) for the earthquake of 1985

Crossover rate	Mutation rate	Average warning time
90%	1%	88
	3%	83
	5%	84
	10%	90
70%	1%	90
	3%	89
	5%	90
	10%	87
50%	1%	84
	3%	84
	5%	89
	10%	90

Table 9: Evaluation of the training phase using the 50% of the total sample space (1269 records) for the earthquake of 1985

Crossover rate	Mutation rate	Average warning time
90%	1%	89
	3%	89
	5%	84
	10%	90
70%	1%	85
	3%	90
	5%	89
	10%	90
50%	1%	89
	3%	85
	5%	90
	10%	85

algorithm could alert the population with more time before the seismic wave reaches México City.

To conclude the experimental results' section, figure 7 shows the distribution of the best proposed stations to alert with the maximum time if an earthquake of 7 ° Richter would occur. Then, table 13 shows the settings of the genetic algorithm for this result.

6. CONCLUSIONS AND FURTHER WORK

We can say that under the conditions described above, the genetic algorithm provided a better solution for the location of the seismic stations. Through several experiments, we observed that the time to alert the population of Mexico City was over 50% higher than the current configuration of SAS.

Although this work was focused on maximizing the number of epicenters located in a coverage area of the stations, we observed that the genetic algorithms could be considered as a useful tool in planning and building seismic sensor

Table 10: Evaluation of the training phase using the 10% of the total sample space (253 records) for the earthquake of 1985

Crossover rate	Mutation rate	Average warning time
90%	1%	89
	3%	84
	5%	84
	10%	84
70%	1%	90
	3%	83
	5%	89
	10%	83
50%	1%	86
	3%	83
	5%	89
	10%	84

Table 11: Evaluation for greedy search algorithm

All earthquakes		Earthquakes>5.0 °		Earthquake
Maximum	Average	Maximum	Average	1985
90	27,83	90	64,9	76,05

Table 12: Final result summary in terms of the warning time

Best	All earthquakes		Earthquakes>5.0 °		Earthquake
Config.	Max.	Avg.	Max.	Avg.	1985
SAS	80	34	80	65.23	57
G. A.	99	31,43	99	68,94	89
Greedy	90	27,83	90	64,9	76,05

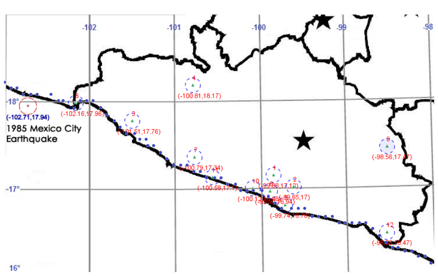

Figure 7: The best configuration proposed by the genetic algorithm

Table 13: Genetic algorithm's settings for the best configuration

% of sample for training	Crossover rate	Mutation rate
100%	100%	5%

station networks. This approach could be applied in other states of Mexico or other countries.

Since configuring the locations of seismic-sense stations is a basically numerical problem, some state-of-the-art numerical optimizers, like Differential Evolution (DE) or Covariance Matrix Adaptation Evolution Strategy (CMA-ES) will be applied and compared against the conventional genetic algorithm. Finally, we will choose the best strategy to solve the problem in question.

Further work might focus on locating the seismic stations considering as the objective function to maximize warning times instead of the number of epicenters around of a station and verify if it gives better results.

In addition, other characteristics like the variation and reflections of waves over strata in earth could be considered during the simulation process. And other georeferenced data like soil properties could be analyzed to show more real and maybe better results.

7. REFERENCES

[1] M. Chávez and K. B. Olsen. Modeling of strong ground motions observed for the 09/10/95, mw=8, colima-jalisco (méxico) earthquake. In *3rd ACES Workshop Proceedings*, pages 149–156, 2002.

[2] H. Ferrer-Toledo, M. Cárdenas-Soto, and F. Chávez-García. Regional path effects on seismic wave propagation in central mexico. *Bulletin of the Seismological Society of America*, 93(3):973–985, 2003.

[3] J. D. Frez Cardenas, F. A. Nava Pichardo, and J. G. Acosta Chang. Source rupture plane determination from directivity doppler effect for small earthquakes recorded by local networks. *Bulletin of the Seismological Society of America*, 100(1):289–297, 2010.

[4] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[5] A. Iglesias and S. Singh. Comportamiento sistemático de bajas aceleraciones provocadas por sismos de trinchera en méxico. In *Reunión Anual 2002 de la Unión Geofísica Mexicana*, page 386, 2002.

[6] M. A. Iglesias. *Trigonometría esférica: Teoría y problemas resueltos*. UPV, Universidad del País Vasco, 2004.

[7] V. Kostoglodov, K. Larson, A. Lowry, W. Hutton, S. Singh, J. Santiago, and S. Franco. Reasoning about naming systems. *Geophysycal Research Letters*, 30(15):1807, 2003.

[8] M. Mitchell. *An introduction to genetic algorithms*. MIT Press, Cambridge, Massachusetts, 1996.

[9] S. S. Nacional. Catalogo: Sismos de 1998 al 2005. `www.ssn.unam.mx/website/jsp/catalogo2.jsp`.

[10] F. A. Nava Pichardo. Fourier spectral analysis for unevenly spaced, average value, data. *Computers & Geosciences*, 36(7), 2010.

[11] M. A. Santoyo, S. K. Singh, and M. O. Schroeder. Evidencia estadística de interacción entre grandes sismos de subducción en la región sismogénica del pacífico mexicano. In *Reunión Anual 2002 de la Unión Geofísica Mexicana*, 2002.

[12] Q. Zhang and C. Wang. Using genetic algorithm to optimize artificial neural network: A case study on earthquake prediction. In *Second International Conference on Genetic and Evolutionary Computing*, pages 128–131, 2008.