Paper No. 1761. (quote when citing this article)
Eleventh World Conference on Earthquake Engineering
ISBN: 0 08 042822 3

OPTIMAL SEISMIC MICROZONATION: A GENETIC APPROACH

C. MONTOYA-DULCHE and F. VITIELLO-SBORGIA

National Center for Disaster Prevention
Delfin Madrigal 665, Mexico 04360, D.F.

ABSTRACT

The optimal seismic microzonation problem is solved using an analogy to the evolution of biological systems. With this perspective, each possible zonation is an organism which evolves, competing to preserve its genetic patrimony. In this process the winning organisms will be those which can adapt to the environment, perfecting generation after generation under natural selection and, therefore, converging to the optimal zonation. In the first section of this paper, the problem of optimal seismic microzonation is stated. In the second section a general approach using genetic algorithms philosophy and some specific techniques for evolutionary simulation are introduced. Genetic algorithms are applied to an example case of seismic microzonation in the third section and a general method to improve the search process is proposed. Finally, results are discussed.

KEYWORDS

Genetic algorithms; seismic microzonation; optimization; hazard studies; assessment

OPTIMAL SEISMIC MICROZONATION

In Earthquake Engineering, we have to balance the material, logistic and social damages, with the cost of the structures. Obviously, the use of lower seismic coefficients produces economic but weak buildings and the higher the values, the stronger are the structures, and more expensive they are. To obtain the optimal seismic coefficient, for a defined structural type, two curves need to be computed. The first one is the relation between cost and seismic coefficient (cost curve), which is unique for a specific structure. The second one is the relation between the expected economical losses and the actual seismic coefficient (damage curve) due to geographical position, which is a function of the structure location. There are several methods to estimate the optimal coefficient. One is by taking the intersection of the curves and another is getting the minimum of the addition of these curves. When calculating the optimal seismic coefficient for a large number of points, it is possible to obtain curves that join equal values (contour graphs). In the same way, using the cost function, we get curves that define the cost for a single structural system type called unitary cost function (Cu). Within this context, the seismic microzonation problem is defined as the aggregation of several administrative areas in an specific number of zones, with the same optimal seismic coefficient, for a given structural system type. Administrative areas are geographical regions in which the domain is discretized. Each administrative area i with an structural system type j have a Cu_{ij} associated. This value represents the amount of assigning an optimum seismic coefficient for a single structural system type, inside this area. The value of this function is the maximum cost present in an administrative area since it represents the maximum optimal coefficient. Theoretically, taking the maximum coefficient, the expected damage will be minimized. For each area, the number of expected structures is known and is called expected frequency F_{ij} . With these parameters it is possible to compute the cost of assigning the seismic coefficient for all the structural system types as follows

$$C_i = \sum_j F_{ij}\, Cu_{ij} \tag{1}$$

When several administrative areas are grouped in a zone K, the maximum value of the cost function of each defined structural system type is assigned, which again corresponds to the maximum seismic coefficient. In this way, the zonation cost is

$$CZ_K = \sum_{i \in K} \sum_j F_{ij}\, \mathrm{MAX}_j (Cu_{ij}) \tag{2}$$

The optimal solution for the seismic microzonation problem, for an n zones case, is obtained by aggregating the administrative areas such that the total cost CZ_T is minimum. This is given by

$$Optimum\ Cost = \mathrm{MIN}(CZ_T) = \mathrm{MIN}\left(\sum_{K=1}^{n} CZ_K\right) \tag{3}$$

In the general case, there is at least one optimal solution. However, this could have several combinations that satisfy the minimum aggregation cost. For a discretized domain, the minimum cost is achieved by defining as many zones as administrative areas. On the other hand, the maximum cost is when all administrative areas are aggregated in a single zone. Determining the number of zones is another optimizing problem but is beyond of the scope of this paper. There is not a formal mathematical solution for the latter problem. Solving by exhaustive search is an easy programming task, however, these methods are not suitable to be applied due to large computer processing time even when there is a reduced number of administrative areas.

BRIEF INTRODUCTION TO GENETIC ALGORITHMS

Terminology

Genetic Algorithms is an optimization process that uses the natural selection as search instrument. It assumes a group of possible solutions or artificial organisms represented by a set of *chromosomes.* Chromosomes codify the organism characteristics by means of a symbol string which is stored in the *genes*. The offset of the genes inside the string is defined as *locus*. The genes values are represented by *alleles* taken from an *genetic code*. The group of chromosomes that characterize an artificial organism is called *genotype*. The information represented by means of the genotype must be decoded to obtain the morphology of the organism. This decoded information is called *phenotype*. The successful level of the artificial organism is defined by a fit function which uses the phenotype as arguments. In Figure 1 all these concepts are showed. Evolutionary process simulation is a group of artificial organisms (*population*) that changes in time by means of tree steps: selection, cross over and mutation.

Genetic Algorithms

Using natural selection as a search instrument, genetic algorithms is a set of techniques (Goldberg, 1989) to simulate natural processes. Over this point of view, each possible solution is represent by an artificial organism that competes to preserve its genetic patrimony. The successful level of the organisms is evaluated by a fit function that represents the object function in the optimization problem, such that the winning organism at the end of the simulation will be the best solution and perhaps the optimum. In the following section we are going to discuss several algorithms.

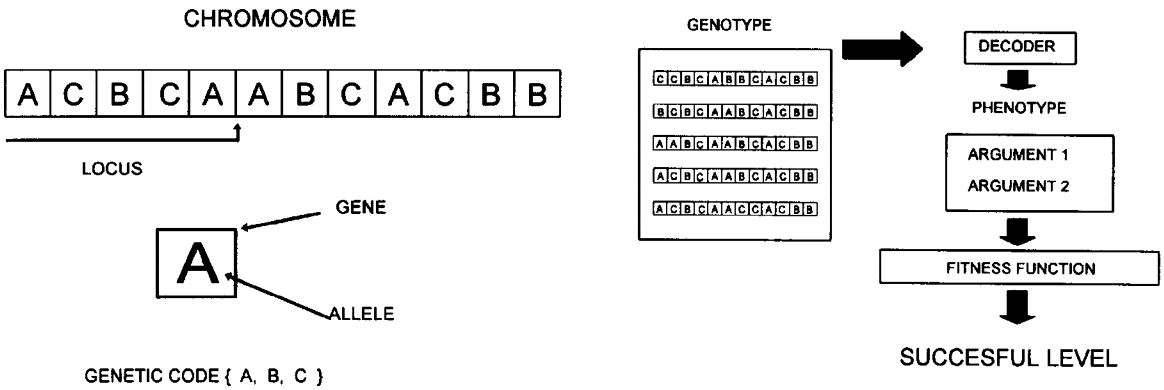

Fig. 1. Chromosome composition and successful level evaluation

Selection. As nature, stronger organism will have a best opportunity to reproduce. The strength of an organism is represented by the level of fitness function or, in terms of optimization, by the object function. In this process, the genotype of each organism is decoded to the phenotype to compute the successful level Fa_s. This value is used to estimate the probability of reproduction $\Pr_s$ as follows:

$$Pr_s = \frac{Fa_s}{\sum_{i=1}^{m} Fa_s} \tag{4}$$

Using this probability the mates for reproduction (cross over) will choose.

Cross Over. In a reproduction process there is a possibility that a couple of mates transmits its genetic patrimony. To simulate this possibility the cross over probability Pc is used. This term represents the probability that the mates' genes will be mixed in the new organisms. If there is not cross over, the new organisms will be an exact copy of the parents. On the other hand, if cross over is present, the reproduction process is simulated obtaining a random locus value and interchanging the genes using the locus as a boundary. This process is shown in Fig 2. This procedure is repeated until a new set of *m* organisms is created (*new generation*).

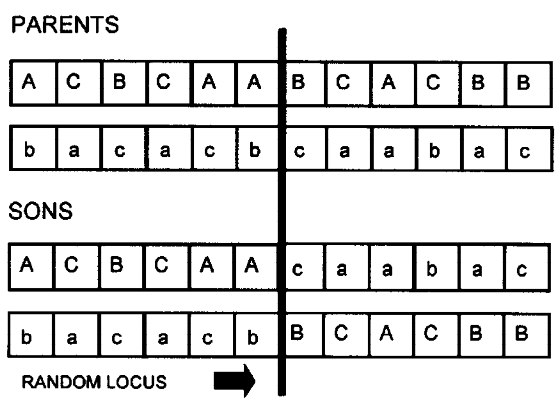

Fig. 2. Cross over

Mutation. In each generation it is possible to have mutation. Mutation is the sudden change of the content of a single allele that belongs to any chromosome of any organism.

To simulate this process, a mutation probability defined as Pm in a Bernoulli variable is used. For each gene of all chromosomes in the generation the Bernoulli variable is computed. If it is true, the allele will change to other valid value of the genetic code.

Special Techniques for an Evolutionary Simulation

Circular Cross Over. Until now it was described the cross over as the interchange of chromosomes segments. However, some times, the continuous edges interchange can decrease the search performance. This aspect can be solved using the circular cross over. Circular cross over assumes that genes are disposed in a ring, such that, the start and end genes of the chromosome are neighbors. The simulation is done by generating two random variables. The first one is the normal locus an the other is an offset. These values define the segment of the chromosome for a normal cross over. In Fig 3, theses aspects are shown.

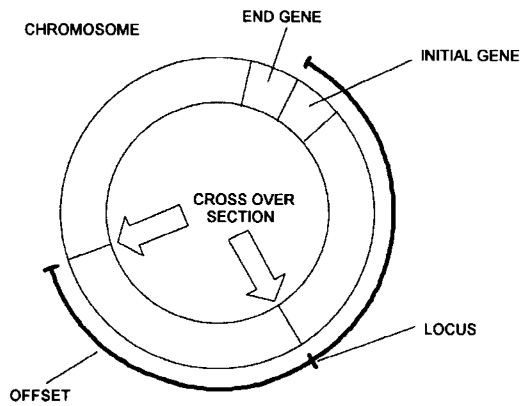

Fig. 3. Circular cross over

Scaling. At the very beginning of the evolutionary simulation, strong organisms, i.e. high fitness-function values, can conduce to a false optimum. On the other hand, when simulation achieves a high number of iterations, organisms presents very similar fitness function values which produce lower convergence performance. To avoid this problem the scaling method is used. There are several scaling techniques. For this research we applied the simplest; linear scaling. This method assumes that for each generation, the fitness function value Fa_s is modified by means of

$$Fe_s = \alpha \, Fa_s + \beta \tag{5}$$

Coefficients α and β must be selected satisfying two conditions. The first one is that the average values before and after of the transformation will be equal ($Fa_v = Fe_v$), insuring in next generation, that the number of average organisms will be one. The second condition is fixed using the Cm parameter which represents the number of sons of the best organism of the generation. More details may be found elsewhere (Goldberg, 1989).

Minimization of the Problem. Genetic Algorithms search solutions for the highest fit function values. However, in this case, a minimization is required. To compute the minimum value using a maximization technique, the system must be changed (duality). Defining Fa_{ls} as the maximum possible value of the fitness function and Fa_r as the minimized solution, the following equation represents the fit function applied for this research

$$Fa_r = Fa_{ls} - Fa \tag{6}$$

GENETIC ALGORITHMS IMPLEMENTATION TO OPTIMAL SEISMIC MICROZONATION

Genetic Model

Genetic implementation of this problem assumes that each possible zonation corresponds to an artificial organism with a single chromosome (*haploid*). The length of the chromosome is equal to the number of administrative areas such that the locus will be equivalent to the index of each one and the gene will represent to the area itself. In this way, the allele of each gene will define the zone in which an administrative area would be located. In this model the decodification between genotype and phenotype is virtual, i.e. genotype is equal to phenotype. This similitude is just for computers. Conceptually, decodification processes change the meaning of the genotype (string of integers) to phenotype (possible microzonation). For this paper we took a small example of 100 administrative areas disposed as follows:

6.3 4.6 71 21	5.8 4.0 15 93	5.4 3.0 82 69	4.9 3.4 47 33	4.4 3.8 42 39	4.0 3.9 37 76	3.7 4.3 82 86	3.3 4.2 9 67	2.9 3.7 16 62	2.4 3.0 71 81
6.9 5.7 37 59	6.5 5.0 52 92	6.0 4.6 86 86	5.5 5.5 74 59	5.0 6.0 84 52	4.5 6.1 6 60	4.0 5.8 80 22	3.7 5.4 5 56	3.3 4.7 67 36	3.0 4.0 75 70
7.5 6.2 87 14	6.9 6.0 82 69	6.6 7.5 98 7	6.1 9.2 34 99	5.7 7.9 93 1	5.0 7.9 2 15	4.6 7.9 23 73	4.1 7.4 27 40	3.7 6.2 57 5	3.4 5.4 55 92
8.2 6.2 85 82	7.8 7.0 24 3	7.2 8.2 29 65	6.7 8.8 70 2	6.2 10.0 0 32	5.7 10.0 27 2	5.1 10.6 50 10	4.6 10.8 37 16	4.2 10.0 50 36	3.8 7.9 6 96
9.0 6.2 73 95	8.5 6.8 33 50	7.9 7.9 15 9	7.3 6.0 16 46	6.8 10.5 91 86	6.3 10.6 95 48	5.7 12.1 58 17	5.1 12.3 30 15	4.6 12.1 20 59	4.2 9.4 56 56
10.0 6.2 72 15	9.2 6.2 47 43	8.6 7.3 55 39	7.9 7.9 8 94	7.4 10.4 13 74	6.8 10.7 78 3	6.2 12.1 93 66	5.7 12.2 4 26	5.1 12.1 29 25	4.7 9.3 93 67
11.0 6.1 99 55	10.3 6.3 44 52	9.5 6.8 62 69	8.8 7.0 57 98	8.0 9.3 49 25	7.4 9.8 98 22	6.8 10.0 34 19	6.2 10.0 36 88	5.7 11.6 15 55	5.1 9.3 57 79
12.0 5.7 65 25	11.2 5.47 96 74	10.4 6.3 88 63	9.6 7.0 35 93	8.8 6.4 38 6	8.0 8.3 73 18	7.4 8.4 24 40	6.8 8.5 98 26	6.2 9.5 85 82	5.7 8.3 91 67
13.0 5.0 60 25	12.1 5.3 38 8	11.3 5.7 92 4	10.3 6.7 5 56	9.5 6.4 64 3	8.8 6.3 52 4	8.0 6.2 98 43	7.4 5.6 0 11	6.8 5.6 81 50	6.1 5.4 12 62
14.0 4.5 20 19	13.0 5.0 5 90	12.1 5.5 53 4	11.0 6.2 63 64	9.9 6.4 78 71	9.3 6.4 13 71	8.7 7.4 18 27	7.9 6.2 81 56	7.2 5.0 42 25	6.6 4.6 99 12

Fig. 4. Administrative areas with its associated parameters

Each cell of Fig. 4 represents an administrative area. The first two numbers are the cost function values and the others are the expected frequency for two structural system types, respectively. In Fig. 5 the contour graph of these cost functions are plotted. For the evolutionary simulation, parameters were defined as follows

PARAMETER	VALUE
Population	100 Organisms
Cross over probability	Pc=0.75
Mutation probability	Pm=0.001
Constant for the best generation son	Cm=1.8

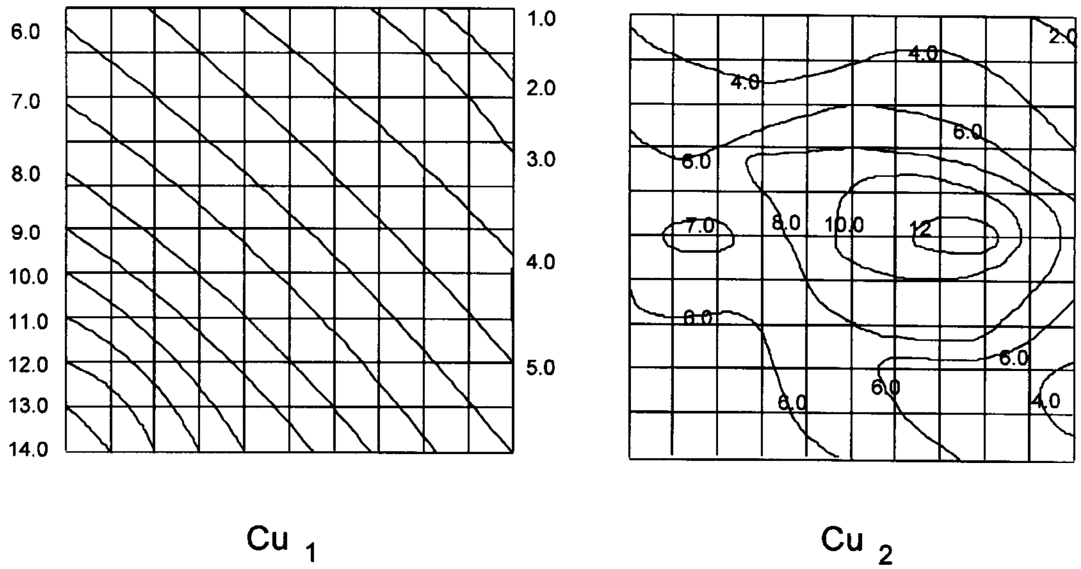

Fig. 5. Contour graph for the cost function of each structural type

For this research, the number of zones in which the areas will be grouped is 4. Using the data mentioned, the maximum cost to incorporate all the areas in a zone is 129,859 an the minimum 69,094 units. Finally, making a complete analogy between genetic and seismic microzonation Table 1 is shown

Table 1. Analogy between genetic and seismic microzonation

GENETIC	SEISMIC MICROZONATION
Organism	Zonation
Genotype (haploid) Number of genes 100	Number of administrative areas
Gene	Administrative area
Locus	Administrative area index
Allele	Assigned zone
Genetic code	0,1,2,3 (zones)
Phenotype	Zonation Map
Fitness function	$129859 - \sum_{K=1}^{4} \sum_{\substack{i=1 \\ i \in K}}^{100} \sum_{j=1}^{2} F_{ij}\, \mathrm{MAX}_j(Cu_{ij})$

Initial Population

In Genetic Algorithms, the initial population is generally a random set of chromosomes. However, when the first generation is defined using some procedure which gives an start point to the search, the performance of the algorithms can be dramatically improved. This paper proposes the method to define the initial population. Analyzing the microzonation problem, it is clear that the lower number of zones leaves to the least possible combinations. Based on the minimum number the zones, two, the initial population will be defined.

In order to compare the random population with the initial definition, all the parameters for the evolutionary process simulation remain constant.

IMPROVING THE GENETIC ALGORITHMS

Our method consists in generating an initial population using the combination of binary zonations that contains the patterns or schemes of the optimal solution. Using a binary genetic code, it is possible to compute, quick and easily, a good solution for each unitary cost function separately and for the global problem joining the cost functions. In the example prepared, the patterns obtained are displayed in Fig 6.

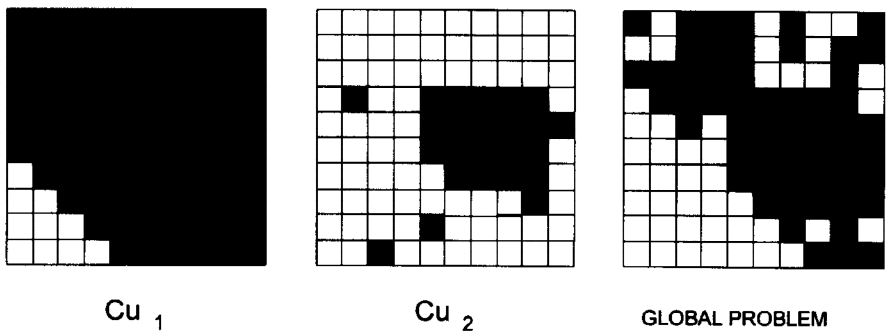

Fig. 6. Binary patterns computed using genetic algorithms

The next step is to introduce the effect of the number of zones. In this way, it is necessary to compute all the possible combinations for each binary solution using the total genetic code, i.e. replace the black and white cells by the combinations {0-1, 1-0, 0-2, 2-0, 3-0, 0-3, 1-2, 2-1, 1-3, 3-1, 2-3, 3-2}. With this procedure, genotypes will be generated that in the cross over algorithm will produce new patterns which contain the contribution of each cost function to the optimum. In the example of this paper, the number of initial phenotypes is 36. The 64 remaining are assigned randomly. These remaining phenotypes will not compete successfully against the binaries patterns. However, it will preserve the random element and will increase the number of organisms of the population. Obviously, at the second generation, all the patterns will be mixed an will be competitively appropriated for the evolutionary simulation.

RESULTS

In Fig 7, results for two simulations, using the same random seed, are presented. The curve above was obtained assigning a complete random population. The second curve represents the initial population. Using different random seeds similar results were obtained. The number of generations was 10,000 which would seems excessive (one million of organisms per simulation), however, the number of possible zonations is 1×10^{60}, approximately. On the other hand, the processing time was 12 minutes using an SUN SPARCStation 10. Even more, the asymptotic behavior of the convergence shows that practically in 1/10 of the evolutionary process, a very good value is achieved. Comparing the curves, it is clear the effect of the initial population. The genetic search performance is increased dramatically due to early identification of the patterns that presents the optimal configuration. The cost of microzonation found for the simulations are 90,063 y 85,736 units for the random and predefined populations, respectively.

CONCLUSIONS

1. Genetic Algorithms offer a robust, easy and low-time process method to obtain a successful result for the seismic microzonation problem.

Montoya O

2. The use of binary patterns to determine the initial population for an evolutionary simulation presents a dramatic improve of the genetic search.
3 Applying these techniques it is possible to obtain the optimum number of zones needed to offer the best economic, social and political solution.

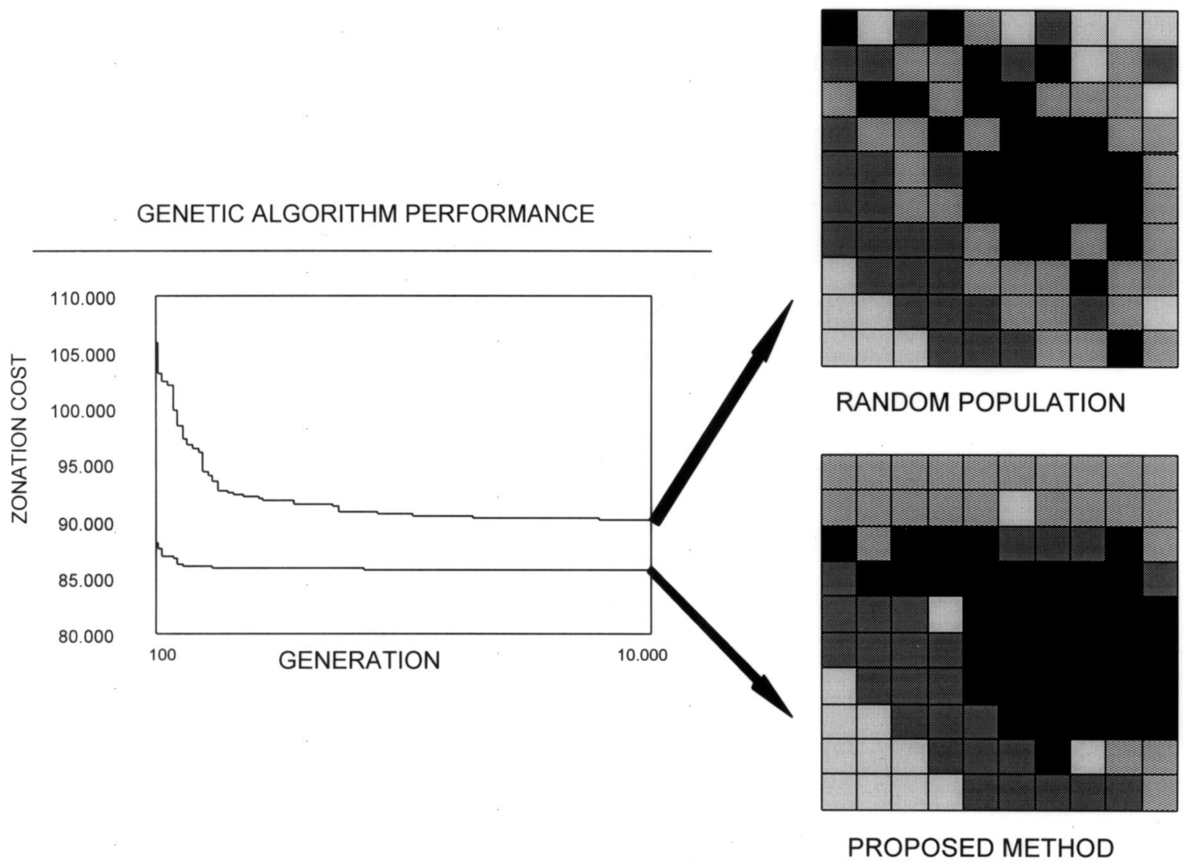

Fig. 7. Comparison between random and predefined population

RECOGNIZANCE

In memorial to Emilio Rosenblueth.

ACKNOWLEDGMENTS

We are grateful to Francisco Sánchez-Sesma, Mario Ordaz, Jorge Arboleda and Roberto Durán for its constructive suggestions. In the same way to Sergio Alcocer for the final review of this manuscript.

REFERENCE

Goldberg, D. (1989). Genetic Algorithms in Search, Optimization & Machine Learning. Addison-Wesley.