# Optimization of shear wave velocity (Vs) from a post-liquefaction settlement using a genetic algorithm Multi-Objective NSGA II

KAMEL GOUDJIL[1], BADREDDINE SBARTAI[2]
[1]Department of Civil Engineering, [2]Department of Civil Engineering
[1]University August 20, 1955-Skikda, [2]University of Badji Mokhtar-Annaba
[1]Road el-hadeik, 21000 Skikda, [2]Sidi Amar BP12 Annaba 23000
[1]ALGERIA, [2]ALGERIA
K_goudjil@yahoo.com, bsbartai@hotmail.fr

*Abstract:* - Heuristic methods such as neural networks (ANN), genetic algorithms (GA) and particle swarm optimization (PSO) have been widely used in the geotechnical field. Several studies have demonstrated the effectiveness of these methods in predicting and optimizing seen their capacity to address the linear or non-linear problem. Our aim through this work is to apply the principle of back analysis using genetic algorithms NSGA II coupled with a simplified method based on measures of shear wave velocity to identify the shear wave velocity (Vs) on the basis of measure settlement post-liquefaction. The results show that genetic algorithms NSGA II has successfully employed to optimize the shear wave velocity (Vs). However, we can used this method to optimize any geotechnical engineering parameter while the conditions are satisfied.

*Key words:* Genetic Algorithm (NSGA II), liquefaction, Shear Wave Velocity, Settlement.

## 1 Introduction

The shear wave velocity (Vs) is a leading geotechnical property that leads to the evaluation of the shear modulus. The shear modulus is necessary in the dynamic analysis that covers a wide range of geotechnical applications including underground constructions [1], deep foundations [2], the soil-structure interaction [3], machinery foundations [4], response of the free field [5] and the susceptibility of the soil to liquefaction [6].

Gmax can be measured in the laboratory using a resonant column device or bender elements. Laboratory testing requires very high quality, undisturbed samples. High quality sampling and testing is quite expensive and is often not possible for cohesionless soils. For that reason, various researchers have studied the relationships between Vs and penetration tests, such as the CPT, the SPT, and the Becker Penetration Test (BPT) [7]. Other researchers have used stochastic methods, such as neural networks and genetic algorithms to predict the shear wave velocity (Vs) according to geotechnical soil parameters [8] and [9].

View of the importance of the parameter *V*s (shear wave velocity of soil, and the efficiency of genetic algorithms in back analysis [10]. In this paper, we decided to apply the back analysis to identify all values of (Vs) along the soil profile from a post-liquefaction settlement using genetic algorithms. To do this, we tried to answer the following question:

- What are the values of (Vs) along the soil profile that induce an overall settlement close to the observed settlement?

Therefore, first we apply a single objective genetic algorithm. Nevertheless, with the data that we have, the genetic algorithm method failed to identify the values of (Vs) that correspond to the actual values. That is why; we have reformulated the question as follows:

- What are the values of (Vs) along the soil profile that induce an overall settlement close to settlement observed and that correspond to the actual characteristics of the soil in question?

The appropriate response to the above question is to find solutions (values (Vs)) that satisfy two objectives simultaneously. The first goal is that the induced settlement of these solutions is close to the observed settlement (0.3m) and the second goal is these solutions should be very close to actual values. This makes our problem a multi-objective problem for this reason; we opted for the multi-objective genetic algorithm NSGA II [11].

The data used in this paper is from the report of Boulanger et al [12]. The value (0.3 m) is the observed settlement occurred in the site named 'Moss landing, located in Monterey Bay, California' after Loma Prieta earthquake 1989.

## 2 Genetic algorithm NSGA II

For solving problems multi-objective optimization, many genetic algorithms have been developed [13]. Among the most significant of them are:

- The Genetic Algorithm for Vector or VEGA Evaluation,
- Niched Pareto Genetic Algorithm(NPGA) using tournament selection, based mainly on the Pareto dominance,
- The NPGAII algorithm, based on the degree of domination of an individual,
- The NSGA or Not dominated Sorting Genetic Algorithm,
- The algorithm Micro-GA, referring to algorithms with small populations,
- Finally the algorithm NSGA-II, based on a classification of individuals in several level

The latter uses a procedure based on non-dominated sorting faster than its predecessor based on non-dominance or Pareto optimal, an elitist approach that preserves population diversity and safeguard the best solutions found in previous generations of one hand, secondly, a comparison operator based on a calculation of the distance crowding. In this paper, NSGA II is that we used.

Before describing the operating principle of the NSGA II algorithm, we must first explain the following concepts: Genetic operators, Pareto optimal solution, Sort Method non-dominated and distance approximation (Crowding Distance).

### 2.1 Genetic operators

Reproduction plays a f undamental role in the transition from one generation to another. This represents one cycle of the genetic algorithm. It is done by applying genetic operators: selection, crossover and mutation.

The selection is to choose individuals from which to create the next generation. The selection of individuals runs mostly based on their evaluation function for single-objective problems and other parameters that will be described later as Pareto rank and crowding distance for multi-objective problems. Several operators of selection exist among of these methods: roulette-wheel selection, Ranking Selection and tournament selection.

The crossover is the first step in a process of reproduction; the genes of the parents are used to form a new chromosome. Several kinds of crossover are available: Single point crossover, two point crossover, Uniform crossover and Arithmetic crossover.

The mutation is the modification of one or more genes of the individual selected to introduce variability in the population with certain probability Pm between zero and one.

### 2.2 Pareto Optimal Solutions

The definition of Pareto optimal solution results directly from the notion of dominance that means is impossible to find a solution that improves performance on a criterion without causing degradation of another criterion. That is why in the multi-objective optimization; the notion of compromise is always mentioned.

### 2.3 Method of non-dominated sorting

Individuals of the current population are sorted to form several Pareto fronts. All non-dominated population individuals receive the rank No. 1 and define the front 1. T hese individuals are removed from the population and the rest of the population is again sorted. Similarly, all non-dominated population individuals receive the rank No. 2 and define the front No. 2. The operation is repeated until that all individuals have a rank (figure 1).

### 2.4 Crowding distance

In the case where two individuals have, the same rank Deb. et al [11] conceptualized a criterion called Crowding Distance. The latter represents the average distance of each objective between the two closest points, located on either side of the same solution of the Pareto front. I distance designates this quantity. This technique maintains a good diversity throughout the population, and allows exploring a wider space of solutions (figure2).

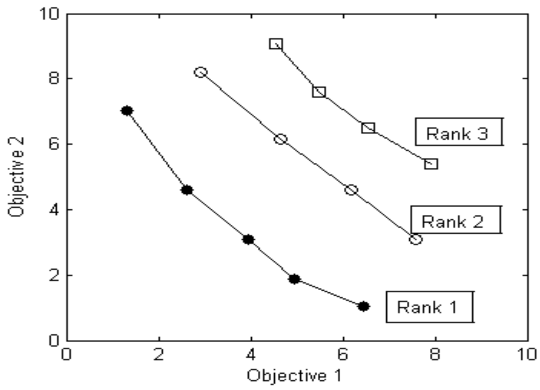


Fig.1 Illustration of non-dominated sorting method of NSGA II.

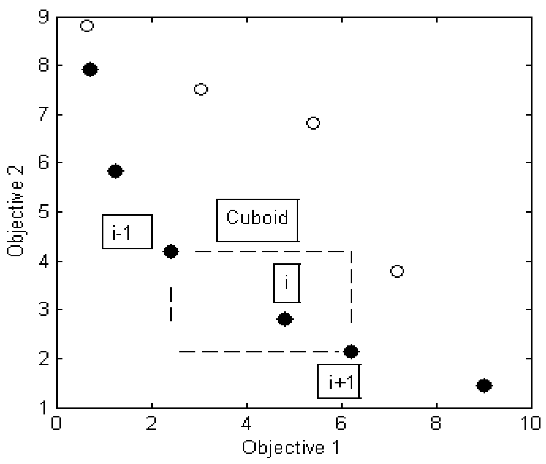


Fig.2 The crowding distance calculation

### 2.5 Working principle of NSGA II

NSGA II varies from simple genetic algorithm only in the way the selection operator works. The crossover and mutation remain as u sual. We describe hereafter the working principle of NSGA II as shown in the figure 3.

Initially, a random parent population P0 is created using creation function (the function that creates the initial population) see options of Matlab toolbox. The population is sorted based on the non-domination. Each solution is assigned a fitness (or rank) equal to its non-domination level (1 is the best level, 2 is the next-best level, and so on). Thereafter, the usual tournament selection, recombination, and mutation operators are used to create an offspring population of size N inscribed by $Q_0$. Then, a combined population Rt=Pt+Qt is formed. The population Rt of size 2N is sorted according to non-domination. Since all previous and current population members are included in Rt, elitism is ensured. Now, solutions belonging to the best non-dominated set F1 are of best solutions in the combined population and must be emphasized more than any other solution in the combined population. If the size of F1 is smaller than N, we definitely choose all members of the set F1 for the new population Pt+1 The remaining members of the population Pt+1are chosen from subsequent non-dominated fronts in the order of their ranking. Therefore, solutions from the set F2 are chosen next, followed by solutions from the set F3, and so on. This procedure is continued until no more sets can be accommodated. In general, the count of solutions in all sets from F1to$F_{Last}$ would be larger than the population size. To choose exactly N population members, we sort the solutions of the last front ($F_{Last}$) using the crowded-comparison operator in descending order and choose the best solutions needed to form the new population Pt+1. This procedure is repeated until one of the Stopping criteria is reached.

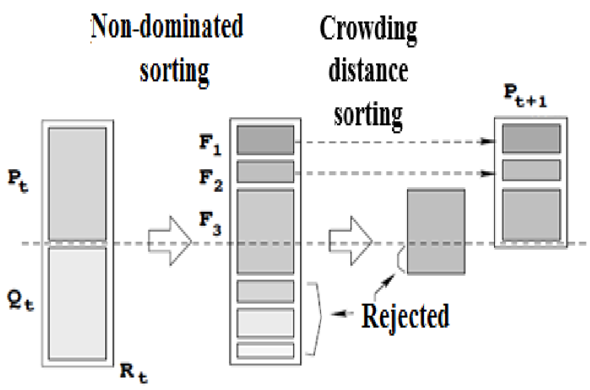


Fig.1 Schematic of the NSGA-II procedure.

## 3 Problem Formulation

As we mentioned above, our work is in two parts. In the first, we treat the problem as a si ngle optimization problem as against the second part, we treat it as a multi-objective optimization problem.

The database that we used to calculate the settlement are taken from paper of Idriss and Boulanger [12].

### 3.1 Problem of single objective optimization

In this part, the mathematical formulation of the problem is defined as follows:

$$\begin{cases} \min F_{obj}(x) \\ x^{lower} < x < x^{upper} \end{cases} \tag{1}$$

Where

$$F_{obj} = \left[\left(\left(\sum_1^n Sett_i\right) - Observed\ Sett\right)^2\right]^{0.5} \tag{2}$$

Sett= settlement which corresponds to the depth subscript (i) of the soil.
Observed Sett=settlement observed equal 0.3 m.

The settlement is evaluated by the simplified method based on shear wave velocity.
The method is describe in details in [14].

Our aim from this mathematical formulation is to find the parameter values of (Vs) that induce global settlement very close to the observed settlement. In other words, this is to answer the question: What are the values of the shear wave velocity (Vs) for an overall settlement close to the observed settlement. This makes our problem a back analysis that is why we used genetic algorithm.

Table 1 Parameters used in genetic algorithm.

| Number of variables | 1 |
|---|---|
| Nbr. of Population | 200 |
| Generation | 100 |
| Selection elitism | 2 |
| Crossover | 0.8 |
| Mutation | 0.2 |

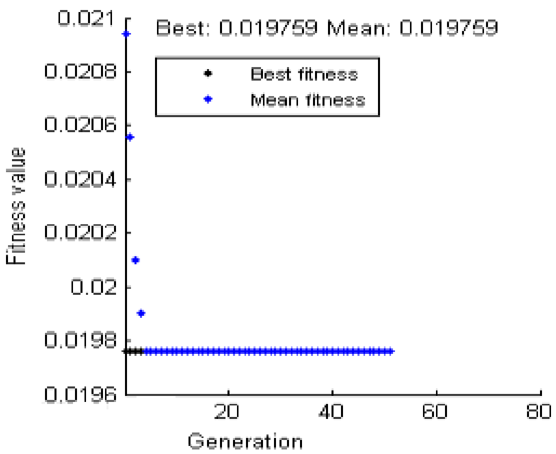


Fig.2 Variation in the best function value in each generation versus iteration number.

Figure 4 s hows that the algorithm converges from the first generation to the best fitness values but the comparative table between the target values and optimized indicate that the error is high despite the global settlement optimized is very close to the actual settlement. This is due to the fact that we have given only two bounds for all points of the soil profile that are LB and UB (100m / s) and (300 m /s)within this range, we can find several different combinations that give all a global settlement very close to the global settlement observed.

In order to highlight this hypothesis, we conducted several scenarios by changing the lower and upper bounds, the results are summarized in Table 3.

Table 3 confirms what we have advanced. This result does not mean anything about the effectiveness of genetic algorithms. Because in every studies of optimization used genetic algorithm such as those we have cited in the introduction the databases with which the researchers worked are couples values (i.e. that each value optimize has a target value). Whereas in this article we want to find the values of all the points from a single point which is the observed settlement. It is for this reason that we have reformulated our question as follows:

What are the values of the shear wave velocity (Vs) of all points along the soil profile giving a total settlement close the settlement observed and each value of shear wave velocity correspond to the actual characteristics of the soil?

Table 2 Comparison between real and optimized values of (*V*s) obtained through genetic algorithm.

| Depth(m) | real values (m/s) | Optimized values (m/s) | Error |
|---|---|---|---|
| 1.8 | 146,61 | 108,7363 | 37,8737 |
| 2.6 | 105,46 | 112,9863 | 7,5263 |
| 3.4 | 120,25 | 114,9863 | 5,2637 |
| 4.1 | 130,03 | 114,9863 | 15,0437 |
| 4.9 | 134,29 | 114,9863 | 19,3037 |
| 5.6 | 151,59 | 114,9863 | 36,6037 |
| 6.4 | 183,09 | 114,9863 | 68,1037 |
| 7.2 | 188,43 | 114,9863 | 73,4437 |
| 7.9 | 192,53 | 114,9863 | 77,5437 |
| 9.4 | 183,12 | 114,9863 | 68,1337 |
| 10.2 | 169,56 | 114,9863 | 54,5737 |
| 11 | 172,31 | 114,9863 | 57,3237 |
| Global Settlement | 0,3156 | 0,438 | 0,1224 |

The answer to this question is to find solutions that satisfy the two objectives cited below simultaneously:

- The total settlement of all points is close to the observed settlement.
- Each value of (Vs) along soil profile correspond to actual characteristics of the soil.

This makes our problem a multi-objective optimization problem.

Table 3 Different combinations set of (Vs)

| Depth (m) | Variant1 | Variant 2 | Variant 3 |
|---|---|---|---|
| 1.8 | 106.7064 | 103.0859 | 105.9093 |
| 2.6 | 106.7064 | 103.0859 | 113.3854 |
| 3.4 | 106.7064 | 103.0859 | 113.3854 |
| 4.1 | 106.7064 | 120.7662 | 113.3854 |
| 4.9 | 106.7064 | 120.7662 | 113.3854 |
| 5.6 | 106.7064 | 120.7662 | 113.3854 |
| 6.4 | 106.7064 | 120.7662 | 113.3854 |
| 7.2 | 106.7064 | 120.7662 | 113.3854 |
| 7.9 | 106.7064 | 120.7662 | 113.3854 |
| 9.4 | 106.7064 | 120.7662 | 113.3854 |
| 10.2 | 106.7064 | 120.7662 | 113.3854 |
| 11 | 106.7064 | 141.8853 | 113.3854 |
| Global Settlement | 0.438 | 0.438 | 0.438 |

### 3.2 Problem of multi-objective optimization

A multi-objective optimization problem, it is a process that minimize two or more functions. i.e., to find solutions that satisfies all functions simultaneously. Depending on the objectives cited above, the number of functions to minimize in this article is two. The first function is that of single-objective problem and the second function may be any proven reliable empirical formula. Several correlations between Vs and commonly measured geotechnical properties (such as SPT and CPT penetration resistance, and undrained shear strength) are developed [7].

The fact that the database used contain the values of SPT-$N_{60}$, therefore we chose the formula of Hasancebi and Ulusay [15] defined by equation (3).

$$V_s = 104{,}79 N^{0{,}26} \tag{3}$$

As a result, the mathematical formulation of the multi-objective optimization problem (POMO) is:

$$\begin{cases} \min of \{F_1, F_2\} \\ x = \{x_1, x_2, \dots x_n\} \\ x^{lower} < x < x^{upper} \end{cases} \tag{4}$$

Where

- $F_1$, $F_2$ are the two objective functions and are defined by equations 5, 6.
- -X represent the vector of non-dominated solutions.
- $X^{lower}$, $X^{upper}$ represent respectively the lower and upper bounds of ($V_s$).

$$F_{obj1} = \left[\left(\left(\textstyle\sum_1^N Sett_i\right) - 0.3\right)^2\right]^{0.5} \tag{5}$$

$$F_{obj2} = [(x_i - 104.79 N_{60}^{0.26})^2]^{0.5} \tag{6}$$

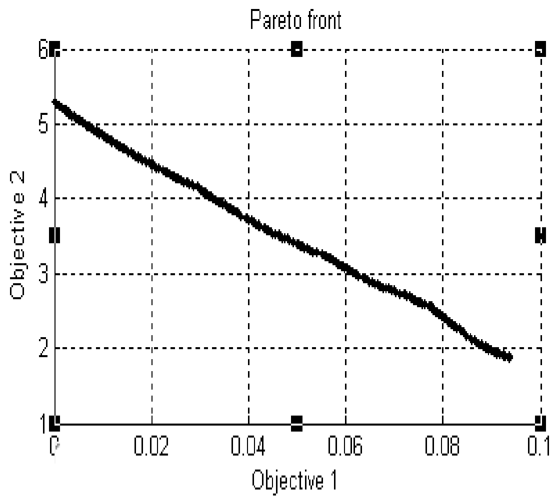


Fig.5 Pareto-optimal solution set of two objectives

Figure 5 shows the Pareto front, as against the figures 6, 7 and 8 shows how genetic operators of NSGA II evolve the population towards the Pareto front. In figure6, individuals are a bit far from the Pareto front but from the fifth generation individuals start to converge to the optimal solutions and at the 15th generation (Figure 8), they become increasingly close to the front.

To verify the performance of NSGA II, the criteria defined by equations 20 and 21 were calculated. $R^2$ (absolute fraction of variance) equal to 0.9999, it is very close to 1 and RMSE (root-mean squared error) equal to 4.029, it is not as good as $R^2$ but it is also little. Table 4, which shows the correlation between the optimized and target values, is in agreement with the performance criterion, which proves the effectiveness of NSGA II in the field of optimization.

$$R^2 = 1 - \left[\frac{\sum_1^n (Set_{optimised} - Set_{actual})^2}{(\sum_1^N Set_{actual})^2}\right] \quad (7)$$

$$RMSE = \left[\frac{\sum_1^n (Set_{Optimised} - Set_{actual})^2}{n}\right]^{0.5} \quad (8)$$

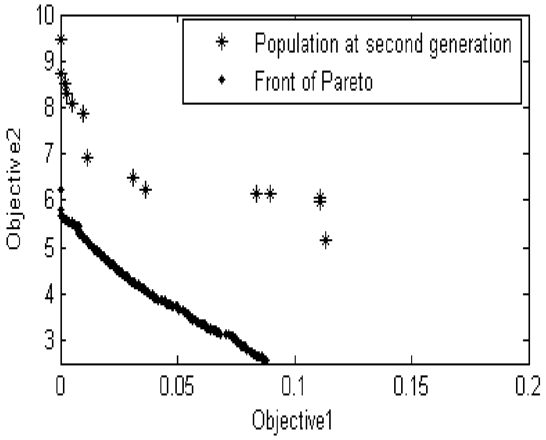


Fig.6 Evolution of population at 2nd generation toward Front of Pareto

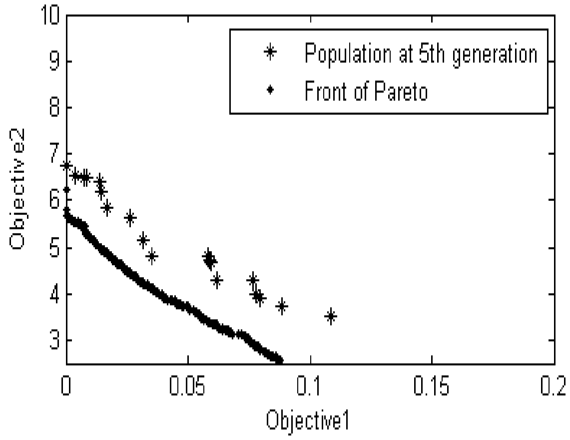


Fig.7 Evolution of population at 5[th] generation toward Front of Pareto

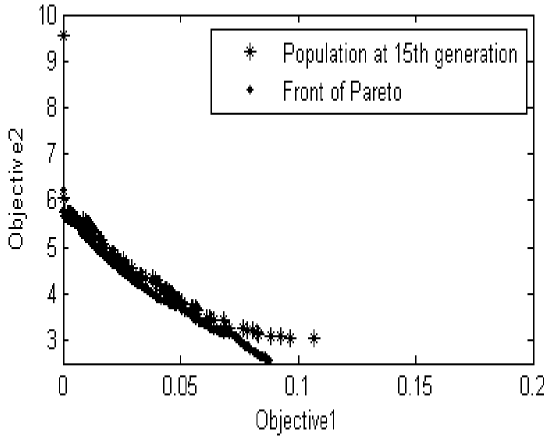


Fig.8 Evolution of population at 15[th] generation toward Front of Pareto

Table 4 Comparison between real and optimized values obtained through NSGA II of (Vs)

| Depth(m) | real values (m/s) | Optimized values (m/s) | Error |
|---|---|---|---|
| 1.8 | 146,61 | 134,48 | 12,13 |
| 2.6 | 105,46 | 112,76 | 7,3 |
| 3.4 | 120,25 | 128,34 | 8,09 |
| 4.1 | 130,03 | 140,61 | 10,58 |
| 4.9 | 134,29 | 141,17 | 6,88 |
| 5.6 | 151,59 | 165,44 | 13,85 |
| 6.4 | 183,09 | 179,23 | 3,86 |
| 7.2 | 188,43 | 187,88 | 0,55 |
| 7.9 | 192,53 | 177,43 | 15,1 |
| 9.4 | 183,12 | 173,82 | 9,3 |
| 10.2 | 169,56 | 158,65 | 10,91 |
| 11 | 172,31 | 163,5 | 8,81 |
| Global Settlement | 0,3156 | 0,3252 | 0,0096 |

## 4 Conclusion

Through this work, we have shown the effectiveness of the genetic algorithm multi-objective NSGA II in optimization parameter (*Vs*) which is a key parameter in soil dynamics. In the first part, we tried to use a simple genetic algorithm implemented in Matlab under the ga function, but unfortunately the problem data were insufficient because we do not have data on settlement at each point of the soil profile and global settlement only failed to find values close to the actual values (Vs).

For this reason that we have reformulated our problem to a multi-objective optimization problem.

In the second part, we used a multi-objective NSGA type II implemented in Matlab under the function gamulti. The advantage of the latter is to find the Pareto front representing all non-dominated solutions that satisfy multiple functions simultaneously. It is this advantage that has allowed us to optimize (Vs) from a si ngle data (post-liquefaction settlement). The results obtained in this paper confirm the effectiveness of NSGA II in the field of optimization. It is recommended this method for problems whose data are insufficient to calculate geotechnical works. It is also a non-expensive way to identify physical and / or mechanical parameters soil without having to go through tests whichever its type.

*References:*


[1] F. I. Makdisi and H. B. Seed, A Simplified Procedure for Estimating Dam and Embankment Earthquake-Induced Deformations, *Journal of the Geotechnical Engineering Division*, 1978, Vol. 104, Issue 7, Pg. 849-867

[2] Randolph, M. F. Randolph, PIGLET: A Computer Program for the Analysis and Design of Pile Groups Under General Loading Conditions, Soil Report TR91, CUED/D, Cambridge University, Cambridge, England1980.

[3] J. Lysmer, T. Udaka, C. F. Tsai and H. B. Seed, H. B. 1975, F LUSH A Computer Program for Approximate 3-D Analysis at Soil Structure Interaction Problems, Report N°. UCB/EERC-75/30, Earthquake *Engineering Research Center*, University of California, Berkeley, California 1975.

[4] F. E. Richart, J. R. Hall and R.D. Woods, *Vibrations of Soils and Foundations*, Prentice-Hall International Series in Theoretical and Applied Mechanics 1970.

[5] J. C. Chen, J. Lysmer and H. B. Seed, Analysis of Local Variations in Free Field Seismic Ground Motions, *Earthquake Engineering Research Center*, Report No. UCB/EERC-81/03, Berkeley, California 1981.

[6] R.D. Andrus and, K.H.II. Stokoe, Liquefaction resistance of soils from shear-wave velocity, *Journal of Geotechnical and Geoenvironmental Engineering*, ASCE126:11, 1015-1025, 2000.

[7] B. R. Wair, J. T. DeJong and T. Shantz, Guidelines for Estimation of Shear Wave Velocity Profiles, *Pacific Earthquake Engineering Research Center*, Report December 2012.

[8] B. Derras and A. Bekkouche, Estimation de la vitesse maximale du sol par la méthode des réseaux de neurones artificiels. *Journées Nationales de Géotechnique et de Géologie de l'Ingénieur,* Grenoble 7-9 juillet 2010

[9] H. M. Abasi, A. E slami and P. T. Shourijeh, Shear Wave Velocity by Polynomial Neural Networks and Genetic Algorithms Based on Geotechnical Soil Properties' *Arabian Journal Science Engineering* , April 2013.

[10] S. Levasseur, Y. Malécot, M. Boulon, and E. Flavigny, Soil parameter identification using a genetic algorithm, *International Journal for Numerical and Analytical Methods in Geomechanics,* 32:189_213, 2008.

[11] K. Deb, S. Agrawal, A. Pratap, *and* T. Meyarivan, A Fast *and* Elitist Multi objective Genetic Algorithm*:* NSGA-II, *Journal IEEE Transactions on Evolutionary Computation*, Volume 6 Issue *2*, April 2002. Page 182-197.

[12] M. Idriss and R. W. Boulanger, *Soil Liquefaction during Earthquake,* Earthquake Engineering Research Institute, EERI Publication MNO-12, 2008.

[13] J.D. S chaffer, *Multiple Objective Optimization with Vector Evaluated Genetic Algorithms*, Proceedings of the First International Conference on Genetic Algorithms, Pg. 93-100, 1985.

[14] F.Yi, Procedure to evaluate liquefaction induced settlement based on shear wave velocity. *Ninth US National and tent^h Canadian Conference on Earthquake Engineering,* 2010.

[15] N. Hasancebi and R. Ulusay, Empirical correlations between shear wave velocity and penetration resistance for ground shaking assessments, *Bull. Eng. Geology and the Environment*, 66:203–213, 2007.