A hybrid multi-objective algorithm to predict the characteristics of soil profiles from seismic ground motion records

Zamila Harichane [a,*], Mourad A. Khellafi [b], Amina Sadouki [a]

[a] *Geomaterials Laboratory, Hassiba Benbouali University of Chlef, 02000 Chlef, Algeria*
[b] *Department of Civil Engineering, University of Djelfa, BP 3117 Djelfa, Algeria*

ABSTRACT

The underlying goal of this study is to present an efficient algorithm to identify soil parameters such as thicknesses, shear wave velocities, damping and others parameters of subsurface layers, and site amplification characteristics (natural frequencies, peak amplitudes) from a given pair of seismic records. It is a hybrid procedure combining the stochastic genetic algorithms (GAs) optimization method, to find a point close to the global optimum in the global search phase, and a gradient based local determinist method (Levenberg-Marquardt: LM), to refine the solution. To improve the performance of the global search phase, a multi-objective optimization algorithm is used to minimize the errors between some characteristics of the theoretical amplification function and the experimental one of vertical array records. The weighted sum method which combines the weighted objectives into a single objective function is used to solve the optimization problem. The efficiency of the present algorithm is proven by several examples. Results show that the scheme works well and the curve fitting was always satisfying. Also, the proposed procedure leads to good approximations, requiring a lower computational effort, yet with good rates of convergence. Moreover, neither the growing number of parameters nor the vastness of the search space reduces the efficiency of the algorithm in predicting the characteristics of soil profiles and site amplification commonly required in seismic risk mitigation.

ARTICLE INFO

Article history:
Received 30 June 2017
Revised 27 August 2017
Accepted 13 September 2017

Keywords:
Soil parameter
Amplification function
Genetic algorithm
Seismic record
HMO

1. Introduction

Earthquakes still difficult to be predicted and an important challenge for seismic risk mitigation is thus to develop methodologies to take advantage of ground motion records to extract the information that may help in earthquake hazard reduction. Engineering analysis of strong motion records in previous destructive earthquakes demonstrated that site amplification and associated damage to structures were caused by local site conditions (Ince, 2011; Govindaraju and Bhattacharya, 2012; Tallett-William et al., 2016; Khellafi et al., 2013). Local site amplification, response analysis, and earthquake hazard analysis play one of the most critical roles in seismic studies and geotechnical earthquake engineering (Li, 2014; Guellil et al., 2017). During an earthquake shaking, the bedrock input excitation motion is amplified on the soil surface when propagating in the soil deposit due to natural variability of soil properties. So, the knowledge of the soil properties is essential to understand ground motion modifications due to soil conditions. These soil properties are conventionally obtained by in-situ or laboratory tests which are costly. Alternatively, the inverse analysis can be useful to identify the physical characteristics of a soil medium from the calculated site amplification (Şafak, 1997).

Site amplification is correlated to soil thicknesses, shear wave velocities, material damping and soil densities (Harichane et al., 2005; Kokusho et al., 2005). The identification of site parameters that influence site amplification was studied by several researchers (Rathje and Navidi, 2013; Khellafi et al., 2016).

* *Corresponding author*. Tel.: +213-027-776398 ; Fax: +213-027-721794 ; E-mail address: zharichane@gmail.com (Z. Harichane)
ISSN: 2149-8024 / DOI: https://doi.org/10.20528/cjsmec.2017.09.015

In practice, most natural soil deposits have mechanical properties variables with depth and should be taken into account for reliable dynamic analysis. This goal is reached usually by dividing the soil profile into homogeneous layers, horizontally stratified, with constant properties within each layer, but variable from one layer to another. The SHAKE computer program (SHAKE 2000), for example, incorporates this solution and still widely used by the geotechnical community.

The algorithm given in this paper provide a convenient tool to predict some characteristics of soil profiles and site amplification, more efficiently than an existing one (Harichane et al., 2005, 2012). The soil profiles are idealized as deterministic and/or random multilayered media (Sadouki et al., 2012; Djilali Berkane et al., 2014). Likewise, an approximate manner by assuming an equivalent single layer (Şafak, 1995) may be used to understand site amplification effects. The goal of this study is to present and use a hybrid multi-objective (HMO) algorithm for inverting experimental data to find the best combination of thicknesses of soil layers and their corresponding shear wave velocities and other parameters. The hybrid technique is obtained by combining the stochastic optimization method of Genetic Algorithms (GAs) with the deterministic method of Levenberg-Marquardt (LM). The main interest in using a hybrid technique relies on the fact that, in spite of being a fast technique, the results provided by the LM method strongly depend on the initial guess. For an unsuitable choice of these values the result of the LM method may even diverge (Harichane et al., 2005, 2012). Furthermore, the GAs method usually yields to a relatively large residual error in the identification or to a prohibitively large computational cost. The multi-objective optimization algorithm will provide a set of non-dominant designs where a further improvement for one objective will be at the expense of another. Theses combinations, however, provide results with greater performances, better accuracy and low computational cost, in order to contribute to earthquake risk mitigation by offering a set of some characteristics of soil profiles and site amplification.

2. Identification Procedure

Most of the real world optimization problems are naturally multi-objective; they usually have two or more objective functions which must be satisfied at the same time and possibly are in conflict each other. In order to simplify the solution, many of these multi-objective problems tend to be modeled as mono-objective problems (Deb, 2001). The inverse problem considered here is formulated as a hybrid identification procedure using, firstly, a global multi-objective optimization method via Genetic GAs with only a few iterations to find the best combination of the sought parameters. The optimization problem is formulated as the minimization of the error between: (i) the peaks characteristics (frequencies and amplitudes) of the theoretical amplification functions and the empirical (or measured) ones and, (ii) the shear wave velocity values of the bottom layer and the bedrock. Next, the Levenberg-Marquardt method is used to refine the solutions. This is outlined by the elaboration of a numerical program in FORTRAN language where its algorithm is schematized in Fig. 1.

```
Begin
    INITIALIZE random size population with random individuals
    INITIALIZE probabilities of selection, crossover and mutation
    while no improvement for very long time (2V) do
        EVALUATE individuals, weighting factors and fitness
        SELECT parents randomly from the population
        CROSSOVER parents to obtain offspring
        MUTATE new population
        UPDATE probabilities of selection, crossover and mutation
            if Best fitness improved then
                GROW population size 1
            else if no improvement for long time (V) then
                GROW population size 2
            else
                SHRINK population size
            end if
    end while
        USE best solution as initial guess for LM
    while desired precision is reached do
        EVALUATE Jacobian matrix and increment parameters
        UPDATE parameters
    end while
end
```

Fig. 1. The overall of the proposed procedure.

2.1. Objective functions

The identification of the sought soil properties consists to minimize the differences between model and measured amplification functions. The first objective function is the sum of squared differences between theoretical and empirical curves expressed by the objective function:

$$e_1 = \frac{\sum_{j=1}^{N_p} \int_0^{\omega_{max}} |Y_{ei}(\omega) - Y_{di}(\{\gamma\},\omega)|^2 d\omega}{\sum_{j=1}^{N_p} \int_0^{\omega_{max}} |Y_{ei}(\omega)|^2 d\omega}, \tag{1}$$

where Y_{di} is the theoretical amplification function given by the model and Y_{ei} is the measured amplification function at the N_P considered points couples. ω_{max} is the maximum circular frequency defining the measured function and $\{\gamma\}$ is the parameters vector to be identified. The second and third objective functions e_2 and e_3, respectively, consist to minimize the differences between the N_k peaks characteristics errors, i.e. the gap between the natural frequencies $\omega_j^{(di)}$ of the theoretical amplification function and the corresponding values of the measured function $\omega_j^{(ei)}$, and the gap between peaks amplitudes $Y_j^{(di)}$ of the theoretical function and the corresponding values of the measured function $Y_j^{(ei)}$, respectively. The fourth (e_4) objective function minimizes the gap between the shear wave velocities at the base of deposit and at the bedrock. These three objective functions are, respectively:

$$e_2 = \sum_{j=1}^{N_k} \left|\omega_j^{e_i} - \omega_j^{d_i}\right|,$$

$$e_3 = \sum_{j=1}^{N_k} \left|Y_j^{e_i} - Y_j^{d_i}\right|,$$

$$e_3 = |v_H - v_r|. \tag{2}$$

The global objective function is the weighted sum of the preceding expressions:

$$\chi^2(\{\gamma\},\omega) = \sum_{j=1}^{4} w_j e_j, \tag{3}$$

where w_j are the normalized weights and Σw_j=1.

2.2. Levenberg-Marquardt method

The optimization problem expressed in Eq. (3) will be solved by applying a systematic algorithm. To minimize the error function (Eq. 3), its first derivate with respect to each of the unknown parameters has to be calculated as:

$$\frac{\partial \chi_1^2}{\partial \gamma_j} = \frac{\partial(\boldsymbol{X}^T\boldsymbol{X})}{\partial \gamma_j} = 0. \tag{4}$$

The vector $\boldsymbol{X}$ is then expanded into Taylor series and only the first order terms are retained. To improve the convergence of the solution of the resulting system of equations, a damping parameter λ is added to the Levenberg-Marquardt algorithm (Marquardt, 1963). The deterministic Levenberg-Marquardt (LM) method consists in constructing an iterative procedure which starts with an initial guess $\boldsymbol{A}^0$ and at the $(k+1)^{th}$ iteration, the new estimate is given by:

$$\{\gamma\}^{k+1} = \{\gamma\}^k + \Delta\{\gamma\}^k, \quad k = 0,1,2,\ldots, \tag{5}$$

with the variation $\boldsymbol{\Delta}\{\gamma\}^k$ being computed from:

$$\Delta\{\gamma\}^k = -[(J^T)^k J^k + \lambda^k I]^{-1}(J^T)^k X^k, \tag{6}$$

where **I** is the identity matrix and **J** the Jacobian matrix.

The iterative procedure is continued until some convergence criterion is satisfied. Being deterministic and based on the Jacobian matrix that depends on the gradients, the LM method presents a fast convergence but with the risk to stop in a relative minimum. Furthermore, as mentioned before, if the initial guess is a reasonable one (as the one obtained by the GAs method) then the LM is likely to achieve the desired minimum with few iterations (Walter and Pronzato, 1997).

As this paper deals with an inverse analysis to identify the sought parameters by using down-hole array records, the acceleration recorded at the bedrock and at the ground surface are needed to calculate the objective function for the optimization problem. Fourier amplitude spectra between recorded and calculated accelerations will be smoothed such that the error functions converge.

2.3. Genetic algorithms method

Since there is several number of extrema in the error function due to the use of seismic data, it seems to be not effective to apply a gradient method. Then GAs may be effective in such problem (Goldberg, 1989). Unlike traditional methods, a GA searches a global optimal solution and does not need to calculate the gradient of the objective function and thereby makes GA a highly promising tool (Kao et al., 2010; Pezeshk and Zarrabi, 2005; Koh and Perry, 2007; Sato et al., 2013). The GA is well suited to solve large combinational design problems.

The creation of the initial population is a blind random search for the solution in the large space of possible solutions. GAs are powerful tools for locating an optimal model by rapidly exploring model space. They make use of a stochastic search through model space employing a transition probability rule to improve the solution (Rodríguez-Zúñiga et al., 1997). Then, once a population of models (i.e. parent) is created, the GA scheme evaluates and selects individuals for reproduction, generates new individuals by mutation, crossovers, and direct reproduction, and finally creates new generation in all iterations (Koza, 1992).

3. Measured and Model Functions

3.1. Measured functions

In the present study two measured function are used. The first one is the standard spectral ratio technique (SSR) which computes the ratio of the Fourier transform of the signal at the measuring point to the same quantity at the reference point as:

$$Y_{ei}^{(sr)}(f) = \frac{F_s(f)}{F_{sr}(f)}, \tag{7}$$

where f is the frequency, $F_s(f)$ and $F_{sr}(f)$ are the Fourier transforms (F) of the ground surface and the reference rock records (generally the acceleration data), respectively. A smoothing technique is used to reduce the effects of noise which is always present in the records.

An alternative to spectral ratios, is the cross spectral ratio (CSR) (Şafak, 1997). This technique computes the amplification (measured) function as the ratio:

$$Y_{ei}^{(cs)}(f) = \frac{F_{rs}(f)}{F_{rr}(f)}, \tag{8}$$

where $S_{rr}(f)$ is the power spectral density of the reference rock record, and $S_{rs}(f)$ is the cross-power-spectral density between the reference rock and the ground surface records (Fig. 2).

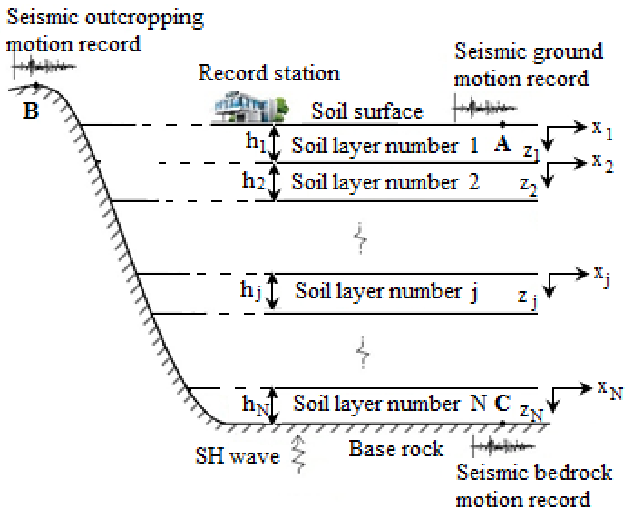


Fig. 2. Ground response nomenclature of a soil deposit overlying bedrock.

3.2. Model functions

Several theoretical models related to the 1-D SH wave propagation in a linear viscoelastic soil deposit to model the soil behavior during earthquakes exist. Some of these model functions needed to achieve the identification procedure are used in the present study.

3.2.1. Amplification function of an equivalent single layer over bedrock

This model was proposed by Şafak (1995). It harmonizes the amplification function with that of an equivalent single homogeneous layer over bedrock as:

$$Y_{di}^{(es)}(f) = \frac{(1+r)^{-\pi f \tau / Q}}{\left[1+2r\cos(4\pi f \tau)e^{-2\pi f \tau / Q}+r^2 e^{-4\pi f \tau / Q}\right]^{1/2}} . \tag{9}$$

In Eq. (9), r represents the reflection coefficient at the interface between the soil layer and the bedrock for the up-going waves. τ is the one-way travel time of waves in the layer and Q, which is called a quality factor, represents the damping in the soil. The unknown parameters governing this model are r, τ, and Q and will be identified using a given pair of soil- and rock-site records in conjunction with the identification procedure, in frequency domain.

3.2.2. Amplification function of a multilayers soil deposit

This function was widely used to study site amplification and site response according to the 1-D SH waves propagating in homogeneous as well as inhomogeneous layered deposits (Wolf, 1985; Harichane et al., 2005; Sadouki et al., 2012; Khellafi et al., 2016; Guellil et al., 2017). The soil medium is idealized as a layered soil deposit consisting of N horizontal homogeneous isotropic viscoelastic layers (Fig. 2) overlying a bedrock. More details on the evaluation of this amplification function may be found in the cited references and in the Appendix A.

3.2.3. Normalized cross spectral density function of random media

This model matches, in a simple stochastic manner; the inherent variability of soil properties. Djilali Berkane et al. (2014) used this model to investigate the contribution of soil layers stochasticity in spatial variation of seismic response spectra. They obtained free response spectra by random vibration theory starting from the cross spectral density of the total ground surface displacement due to incident random waves impinging the bedrock layer, given by Zerva and Harrada (1997). The normalized cross spectral density function of the random media is given by:

$$S(\omega) = [\omega_0^4 + (2\beta + 4\xi_0^4 - 2)\omega_0^2\omega^2 + (\beta-1)^2\omega^4] \cdot |T^{dl}(\omega)|^2 + 4\beta^2\omega_0^4\omega^2\sigma_{\omega\omega}^2 \cdot |T^{dl}(\omega)|^4 , \tag{10}$$

where β is the participant factor, ω_0 the natural frequency, ξ_0 the damping ratio and $\sigma_{\omega\omega}$ the standard deviation of the natural frequency of the random media. $T^{dl}(\omega)$ is the harmonic amplification function of a single degree of freedom oscillator with frequency ω_0 and damping ξ_0:

$$|T^{dl}(f)|^2 = \frac{1+4\xi_0^2(f/f_0)^2}{\left[1-(f/f_0)^2\right]^2+4\xi_0^2(f/f_0)^2} . \tag{11}$$

4. Results and Discussions

4.1. Comparison between measured and model functions

Before using the inverse analysis, the efficiency of the model functions to match the measured ones is studied by comparing different measured amplification functions (MAFs) and theoretical ones (TAFs) with experimental data recorded within the Garner Valley Downhole Array (GVDA) using actual surface to depth input data. The Garner Valley (California) downhole array was performed to improve the understanding of the effects of a shallow soil site on ground motion. The characteristics of the soil profile of the Garner Valley site upper the 22 *m* are given in Refs (Pecker and Mohammadioun, 1991; Steidl et al., 1996). A small-magnitude (M_L=2.5) motion recorded from a large collected motions during the period 1989-1991 was selected for computational purposes. Recorded accelerograms (horizontal components) obtained at this site are the same used by (Harichane et al., 2005). All empirical techniques use 4096 points FFT (Fast Fourier Transform) and corresponding curves are smoothed by using a triangular window of 0.5 Hz width.

The TAF is computed according to Harichane et al. (2005) using the properties given in Refs (Pecker and Mohammadioun, 1991; Steidl et al., 1996). The resulting plots of amplification functions for frequencies interval between 0 and 20 Hz are presented in Fig. 3. For clarity, Table 1 summarizes the natural frequencies (*f*) and the corresponding peak amplitudes (*PA*) that characterize the different functions. Examination of Fig. 3 shows a good agreement in terms of natural frequencies. However, it can be seen that the TAFs amplitudes are greater than the MAFs ones due the low damping values (1-4%). Also, the SSR technique gives greater amplification than the CSR but still lower than theoretical ones; though the natural frequencies (obtained with any technique) are generally in concordance.

Therefore, the identification procedure of soil parameters using curve fitting between any MAFs and TAFs methods is possible. In order to ensure a sufficient accuracy and reduce the computational cost when using the GAs method, the normalized weights in Eq. (3) should be chosen so that the global objective function is minimized by focusing the weight (w_2) of the natural frequencies error (e_2) and penalizing the weight (w_3) of the amplitude peak errors (e_3). We should also not take into account the gap error between peak amplitudes associated with the fundamental frequency in the calculation of the error e_3 because of the large deviations observed between MAFs and TAFs peak amplitudes at this frequency.

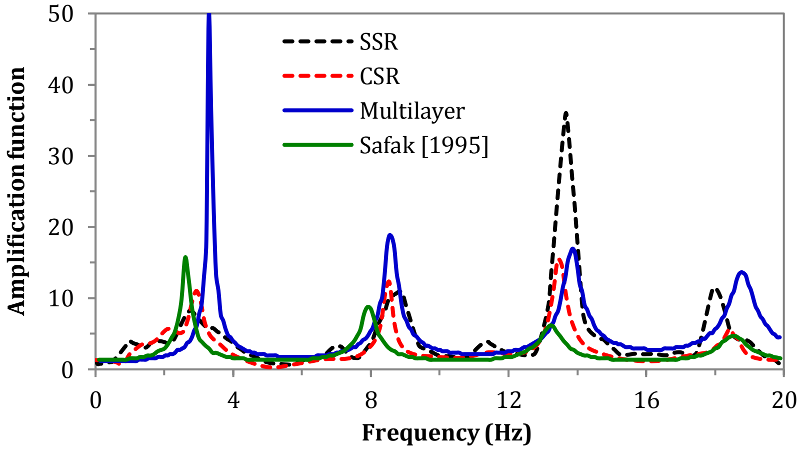


Fig. 3. Comparison of different MAFs and TAFs between free surface and 22 m depth at GVDA site.

Table 1. Comparison of natural frequencies and corresponding peak amplitudes (PA) of different MAFs and TAFs between free surface and 22 m depth at GVDA site.

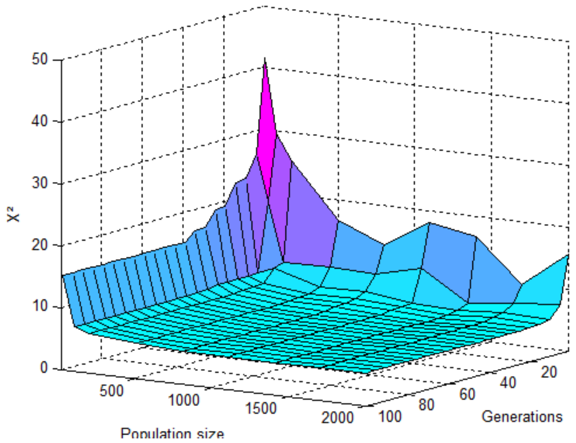
Mode	1		2		3		4	
Frequency /Peak Amp	*f* (Hz)	PA	*f* (Hz)	PA	*f* (Hz)	PA	*f* (Hz)	PA
SSR	2.8	8.3	8.9	11.0	13.7	36.0	18.0	11.5
CSR	3.1	11.0	8.7	12.2	13.5	15.3	18.3	5.9
Multilayer formulae	3.4	49.9	8.6	19.5	13.9	16.6	18.8	13.7
Şafak (1995) formulae	2.7	15.6	8.7	8.7	13.3	6.0	18.6	4.5

4.2. Identification of soil profiles characteristics

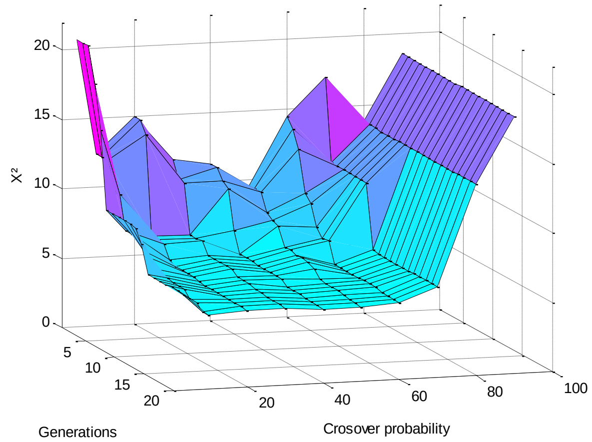
The effectiveness of the proposed algorithm is illustrated through several examples where its stability in adapting itself to match a given target is shown. In the first example, the present algorithm is applied to identify the parameters of an equivalent single layer corresponding to the best match of the theoretical and measured amplification functions with the GVDA site data. In the other cases, the validity of the present scheme is demonstrated with numerical applications.

4.2.1. Identification of characteristics of an equivalent single layer

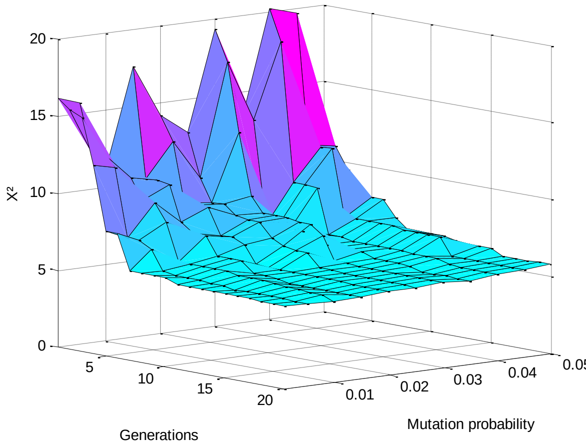
As defined by Eq. (9), the design variables for this problem consist of three unknowns which are *r*, *Q* and τ. Their identification is performed on the search space including all possible values: 0-1 for *r*, 0-1*s* for τ and 1-100 for *Q*. The soil amplification function between free surface and 22 m depth is computed by the SSR and the CSR methods. The procedure has been tested for effectiveness for the GA method with several population sizes, number of maximum generations, probability of crossover and mutation. With both methods (GA and LM), the objective functions used are given by Eq. (3). The normalized weights to be considered with the GA method will be determined to obtain the best results according to the considered models. Since the Şafak (1995) model does not permit to calculate the shear wave velocity at the base of the layer, the normalized weight $w_4 = 0$; therefore only two parameters, w_1 and w_2 are used because the third one is $w_3 = 1 - w_1 - w_2$. For these cases, a graphical shape of the objective function (Fig. 4) is established. Thereby the smoothness properties of the surface can be examined as well as the location of the optimum solution. Only two optimization parameters are involved, which means that the objective function surface is illustrated in a three dimensional space. It is obviously very difficult to inspect the function graphically if more than two optimization parameters are required.



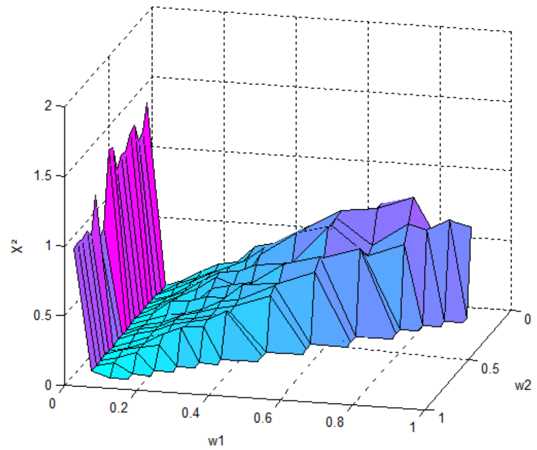
(a) Population size and number of generations.



(b) Crossover probability and number of generations.



(c) Mutation probability and number of generations.



(d) Objective function space versus w_1 and w_2.

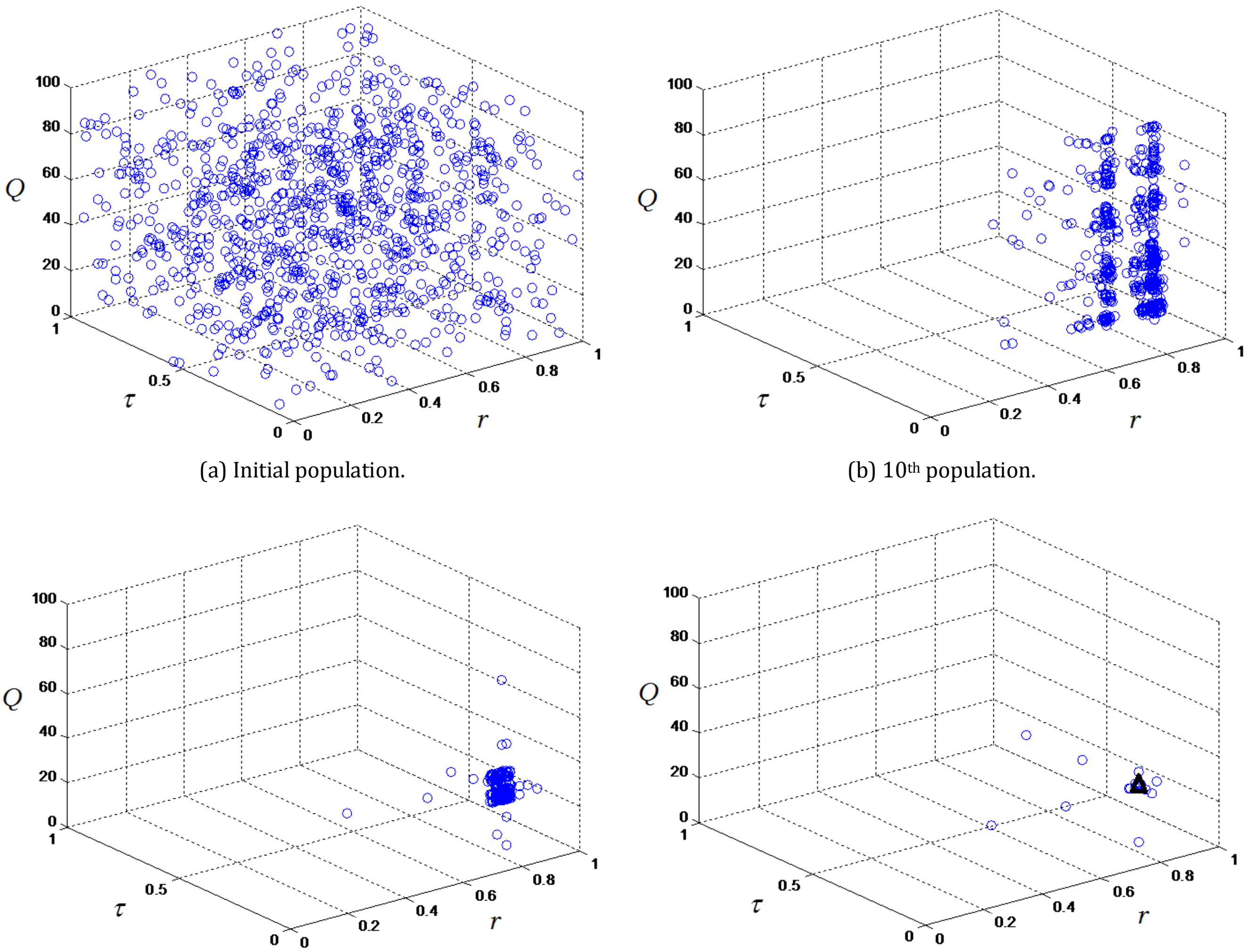
Fig. 4. Convergence history of the GA method versus.

The obtained results indicate that simultaneous increases of the population size from 20 to 2000 and the number of maximum generations from 5 to 100 have considerably improved the performance of the algorithm (Fig. 4a). A population size of 800 founds the solution in 58 GA generations and 7 LM iterations, while a population size of 1400 founds the solution in only 16 GA generations and 5 LM iterations. The crossover operation was sensitive in perturbing the objective space (Fig. 4b). As the crossover increased from 20% to 60%, the solution was found in only 16 generations. But, when the crossover rate was outside this interval, the objective function moved away from the minimum values. Consequently, the closest solution corresponding to a crossover value of 40% was chosen for the optimization process. The mutation operation was not as sensitive in perturbing the objective space when the number of generations is greater than 16. So, there was no particular difference observed when the mutation operator is increased from 0.1% to 5% (Fig. 4c). Nevertheless, a value of 0.1% provides the best solution as compared to the others. However, when the number of generations decreases below 14, the objective function increases in a remarkable way with the decrease of the generations number. The normalized weights are sensitive in perturbing the objective space (Fig. 4d). The results show that variations are more pronounced with respect to parameter w_1 with minimums values observed at w_1 = 5%, and are fairly constant with respect to the parameter w_2. The corresponding weights which led to the best result are: w_2 = 60% and $w_3 = 35\%$. Table 2 summarizes the optimal values for the GA parameters that were used for the optimization. During the first generation of the GA, the variables of the population were initiated randomly. While, for the further generations the variables were modified using the genetic operators of crossover and mutation. At the end, a few numbers of iterations with LM (2 to 9) was sufficient to refine the solutions. It is noticed that results can be improved mainly by increasing the population size or the number of generations. Generally, the optimization process works satisfactorily and led towards an optimal solution.

Table 2. Optimal parameters chosen for GA from sensitivity analysis.

Parameter	Population size	Number of generations	Crossover probability (%)	Mutation probability (%)	Normalized weights (%)			
					w_1	w_2	w_3	w_4
Optimal value	1400	20	40	0.1	5	60	35	0

Fig. 5 shows the evolution of populations for successive generations during the GA optimization phase for the identification of the parameters of the model. The associated evolution of the objective function versus the population generation number is plotted in Fig. 6. This figure shows that the optimization procedure permits to obtain approximate solutions within a margin of error of 10% after 16 generations only. The identification results are compared in Table 3 with actual ones. A good agreement is observed between actual and identified values of r and τ. Nevertheless, the identified values of Q are slightly different from the actual ones within a margin of error of 20%.



(a) Initial population.

(b) 10th population.

(c) 17th population.

(d) 30th population.

Fig. 5. Evolution of population in the GA method for the identification of parameters r, Q and τ. Points denoted **o** are individuals in the search space; point denoted Δ is the final solution.

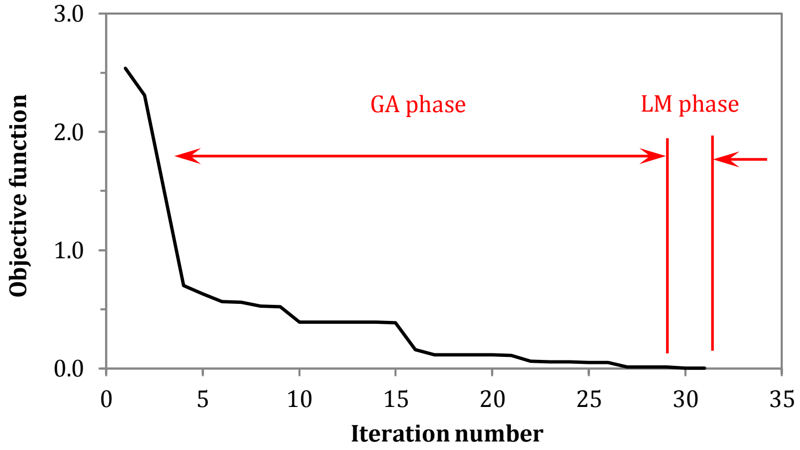


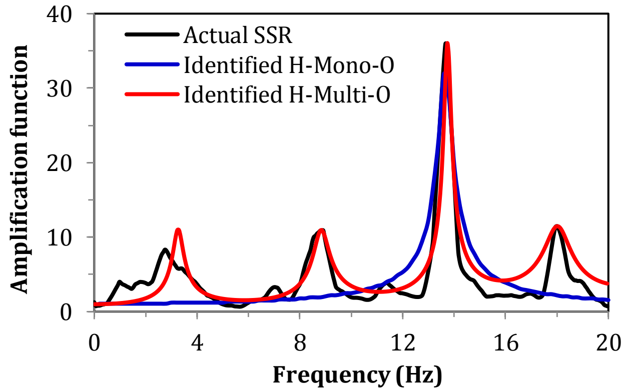
Fig. 6. Evolutions of the objective function versus iteration number.

Table 3. Comparison between identified and actual parameters values for equivalent single layer at the GVDA site.

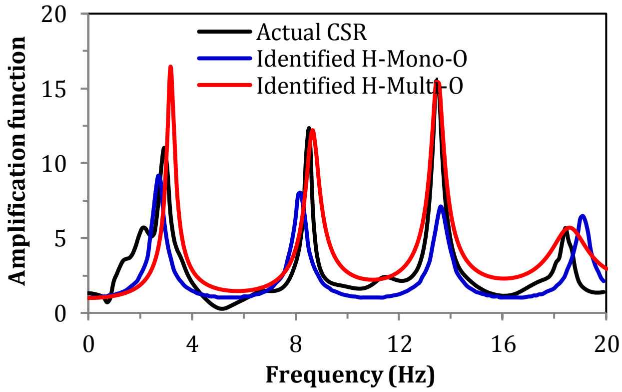
Actual values		Identification with SSR		Identification with CSR	
		Mono-objective optimization	Multi-objective optimization	Mono-objective optimization	Multi-objective optimization
r	0.8	1.0	0.9	0.8	0.9
τ	0.1	0.01	0.1	0.1	0.1
Q	32.4	65.7	37.5	40.0	39.1
Objective function		0.2	0.5	0.3	0.4

To show the effectiveness of the proposed algorithm, the identification results for, firstly, a hybrid mono-objective (H-Momo-O) optimization method and, secondly, with a hybrid multi-objective (H-Multi-O) one are compared in Fig. 7 and Table 3. Fig. 7a shows that the H-Mono-O procedure has been unsatisfactory. Even after several trials with more population sizes and generations it was not possible to closely approach the optimal solution although the error with the one objective function is less than that given by the H-Multi-O.

These results mean that during the optimization process, the GAs method searches the minimum error without taking into account the amplification function shapes. With this approach, the obtained error is less than that brought by the H-Multi-O but the curve fitting is bad; which is the advantage brought by the H-Multi-O method to the identification process. With this method, the optimization is guided by an objective function which takes into account the shape of the amplification function by minimizing in addition, the gap between the resonance frequencies and the corresponding amplitudes. Fig. 7b shows that the H-Mono-O identification is better compared to the previous case, but remains less than the H-Multi-O identification. This is due to two factors: the first one is that the CSR amplification function gives peaks amplitudes smaller than those obtained with the SSR method and do not change abruptly with frequency like with the third resonant peak. The second one is that the function identified with the H-Mono-O agrees better with the shape of the amplification function. Consequently, the H-Multi-O identification seems to work best to cover imperfections that may be encountered with the amplification function shapes.



(a) Identification using the SSR method.



(b) Identification using the CSR method.

Fig. 7. Comparison between empirical and identified amplification functions at GVDA site.

4.2.2. Identification of characteristics of a multilayers soil profile

This application consists in the validation of the present algorithm with a numerical example. It consists to identify properties values of an actual nuclear power plant site as specified in Table 4.The parameters identification is assumed to be performed on a search space including all possible values: 0-50 m for h, 1000-2500 kg/m³ for ρ, 100-1500 m/s for v_s, 0-10% for ξ. The parameters of the GA optimization are the same as in the precedent example.

Table 4. Characteristics of an actual site of a nuclear power plant (Wolf, 1985).

Depth (m)	Shear wave velocity (m/s)	Mass density (kg/m³)	Damping ratio (%)
0 – 5	200	2000	7
5 – 10	250	2000	6
10 – 20	350	2000	5
20 – 30	500	2200	5
30 – 40	800	2200	5
40 – 50	1000	2400	4
(bedrock)	1500	2500	2

a) Identification of all parameters

As was done previously, to evaluate the representativeness of the identified solution, different optimizations are performed in the same search space. The only difference concerns the initial GA population that is each time randomly selected in the search space. Results show that the solution to this problem is not unique. The identification process leads then to a very large range of possible values for each parameter. However, the proposed procedure works well and the curve fit is very good, but with very high computational cost. When the layers number N is considered unknown, the number of the sought parameters is variable and equal to $4N+1$. The identification is done in several successive steps, where within each step the procedure is run with the value of N being constant and varies from 1 to 10. The path from one step (n) to the next ($n+1$) is controlled by a transition condition which minimizes the difference between the average shear wave velocities of the n^{th} layer already identified and of the $(n+1)^{th}$ layer to be identified of the soil deposit:

$$e_5 = |\bar{V}_s^{n+1} - \bar{V}_s^n| \quad , \quad \bar{V}_s^n = \frac{\sum_{j=1}^{n} h_j}{\sum_{j=1}^{n} \frac{h_j}{v_{sj}}} \quad . \tag{12}$$

In this case, a very good curve fit from five layers soil profile is obtained. Fig. 8 shows the obtained minimum error function versus the number of layers. It should be noted that the curve fitting is very good for 3 layers as well as for 6 layers soil profiles. It can be deduced from these tests that the solution is not unique. Nonetheless, despite the vastness of the search space when the number of sought parameters is important it is not needed to highly increase the population size or the number of GA generations of the optimization methods in order to reach the objective. In spite of that, the proposed hybrid multi-objective algorithm is still working fine regardless the difficulties met in this kind of optimization problems.

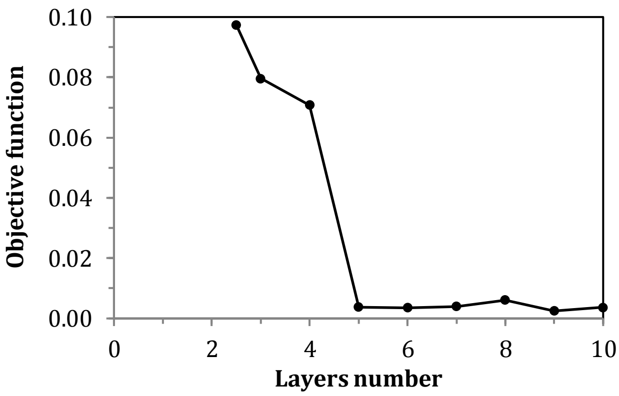


Fig. 8. Minimum objective function versus layer number.

b) Identification of reduced number of parameters

In the identification process, the GA parameters and the search space are the same as in the precedent case and the layers number is considered known (N=6). First, the process is repeated for the identification of one parameter per layer, then, the process is repeated for the identifications of two parameters per layer and finally three parameters per layer.

For the first case, the number of identified parameters is equal to 6. The proposed procedure has been accomplished successfully for the identification of h_j, v_{sj} and ξ_j, separately. The values of these parameters were identified exactly with a very good agreement between the identified and the actual curves (Fig. 9). However, the procedure has failed to identify exactly the values of ρ_j. This is because mass densities for two successive layers are dependent in the expression of q_j (Appendix A) and should not be explored independently. For the second case, the number of identified parameters is equal to 12. Herein, the proposed procedure has identified exactly the values of h_j, v_{sj} and ξ except in the case where h_j and v_{sj} were identified together. In this last case, h_j and v_{sj} are dependent in the ratios h_j/v_{sj} and v_{sj}/v_{sj+1} that prevents their identification (in expression Λ in Eq. (A.3) in the Appendix A).

For the third case, the number of identified parameters is equal to 18. The identification has been disappointing and it was not possible to closely approach the known solution. Different solutions consisting of different combinations of the sought parameters were obtained by running the proposed inversion process that resulted in a very good agreement between the identified and the actual curves (Fig. 9).

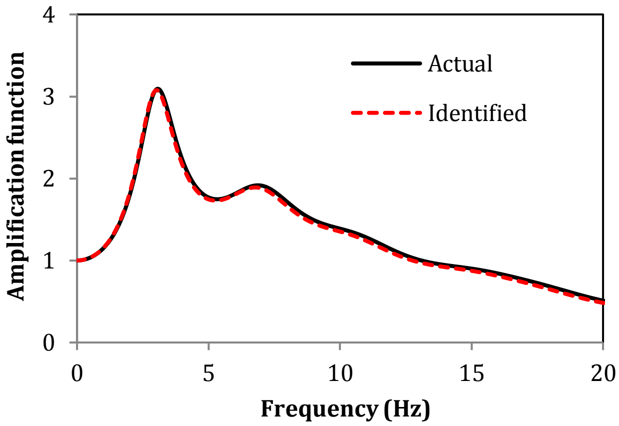


Fig. 9. Comparison between actual and identified amplification functions for a nuclear power plant.

Consequently, the proposed procedure has failed to well identify the parameters of a multilayers soil deposit especially when the number of unknown parameters is larger than three parameters per layer. Analysis of results has shown that the parameters are dependent each other and cannot be identified independently, which explains the existence of different solutions. The expansion of the search space has not decreased the performances of the proposed procedure. Indeed, the process converges constantly to a final solution without increasing the population size or the number of generations.

4.2.3. Identification of characteristics of a randomly inhomogeneous layer via the normalized cross spectral density

This application is dedicated to identify the parameter ξ_0, ω_0, β and $\sigma_{\omega\omega}$ of a random inhomogeneous layer modelling a soil deposit as defined by Eq. (10). The minimization is done by using the normalized cross spectral density function. The identification procedure is performed on the following search space: 0-100 rad/s for ω_0, 0-10% for ξ_0, 0-10 for $\sigma_{\omega\omega}$, 0-10 for β. The actual characteristics of the random layer are specified in Table 5 (Djilali Berkane et al., 2014) and the parameters of the GA optimization are the same as in the first example.

Table 5. Comparison between actual and identified parameter values of a randomly inhomogeneous layer via the normalized cross spectral density.

Parameters	Actual values	Hybrid multi-objective procedure	
		GA only	Final
ω_0 (rad/s)	3.05	3.16	3.05
ξ_0 (%)	5.0	6.0	5.0
$\sigma_{\omega\omega}$	0.12	0.01	0.12
β	1.0	0.5	1.0

Initially, the identification has been very disappointing and it was not possible to closely approach neither the known solutions nor the actual curve with sufficient accuracy. To improve the identification accuracy and efficiency, therefore, we have conducted several tests by changing the parameters of the above optimization. Analysis shows that efficiency increases significantly with higher values of normalized weight w_1 without modifying as much the GA's parameters. The normalized weights that provide the required solutions with low computational cost are: $w_1 = 80\%$ $w_2 = 10\%$ $w_3 = 10\%$ $w_4 = 0\%$.

An important number of iterations with LM (25 to 45) are needed to refine the solutions. Fig. 10 shows an example of the curve fitted to the actual one.

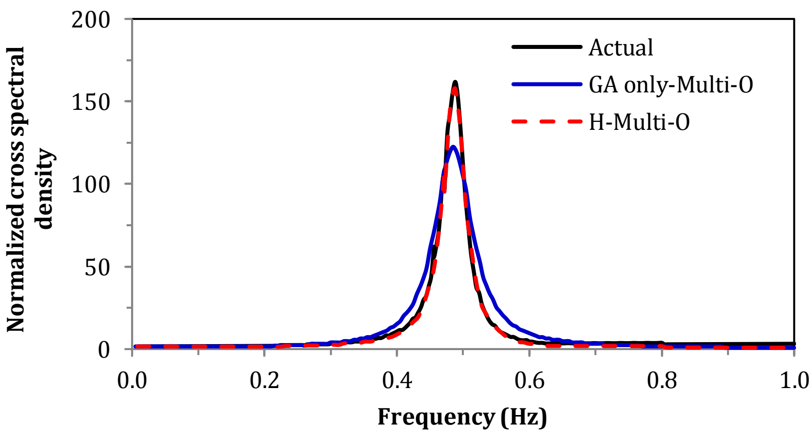
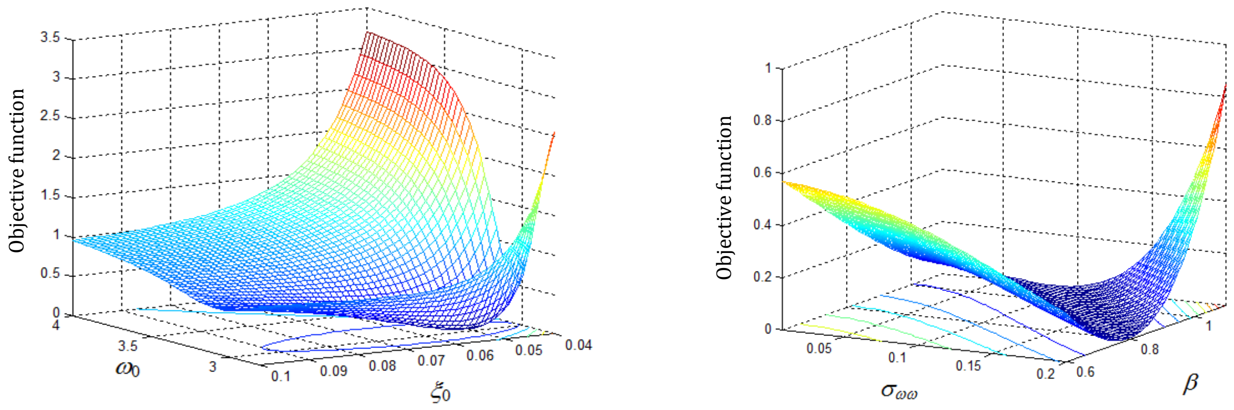


Fig. 10. Comparison between actual and identified normalized cross spectral densities for a nuclear power plant.

To explain these changes, the topology of the objective function is plotted versus two optimization variables (Fig. 11) as if more than two optimization variables are required it is obviously very difficult to inspect the function graphically.

In Fig. 11a, it can be seen that the objective function surface has a unique and well-defined minimum but in very small space close to this minimum. Thus, by comparing this space to the predefined search space, it is clear that the identification is not easy. Moreover, it can be observed in Fig. 11b that the objective function surface describes a flat and long valley in the search space in which the minimum is not well defined. Though, it is known that the narrowness of the valley greatly decreases the probability of finding the solution with evolutionary programming algorithm (Dunning, 1998), which explains the necessity of a large number of gradient iterations to reach the local minimum.



(a) The objective function versus ω_0 and ξ_0.

(b) The objective function versus β and $\sigma_{\omega\omega}$.

Fig. 11. Topology of the objective function close to the sought minimum.

Furthermore, when increasing the number of generations in the GA optimization, a few numbers of gradient iterations is needed to refine the solutions. For example, with 50 generations in the GA optimization, 4 to 9 of LM iterations are sufficient to achieve the planed goal (Fig. 12). But increasing the population size in the GA optimization does not give improvement.

5. Conclusions

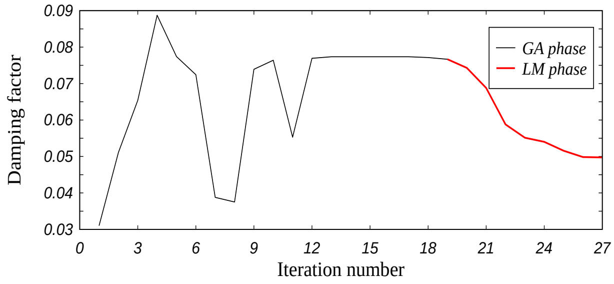
The present study concerned the identification of soil profile characteristics such as layer's thicknesses, shear wave velocity, mass density, damping, and some site amplification characteristics (natural frequencies, peak amplitudes) from a pair of records of priori known of soil constitutive equation from typical earthquake records by inverse analysis. These inverse problems are solved by minimizing the errors between the theoretical and empirical amplification functions of the subsurface layers of vertical array records. A hybrid multi-objective (HMO) optimization algorithm combining the genetic algorithms (GAs) and the traditional based gradient method (LM) is presented. Several examples were carried out in order to test the efficiency of the algorithm.

The gradient method seems to be robust and efficient only if the shape of the objective function is relatively smooth or if the initial guess is very close to the solution. While the genetic algorithm enables the convergence of a set of solutions close to the best one even with a flat or noisy error function. In order to ensure convergence of the genetic algorithm method with low calculation cost, the population is guided towards the preferred solutions with four objective functions. It is shown that the (HMO) method GA-LM leads to good approximations for the global minimum, requiring a lower computational effort, yet with good rates of convergence. Nevertheless, the normalized weights should be chosen so that the global objective function is minimized by focusing the weight of the natural frequencies error and penalizing the weight of the amplitude peaks errors, and do not take into account the gap error between peak amplitudes associated with the fundamental frequency in the calculation of the objective function.

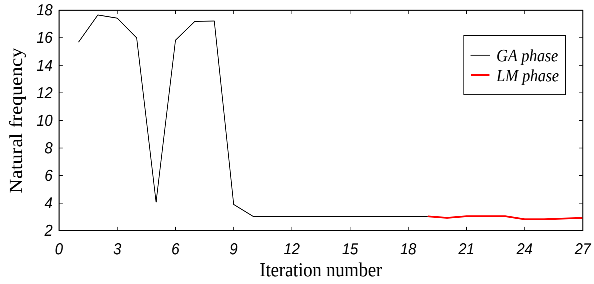
The hybrid identification scheme is validated by using experimental data recorded within the Garner Valley Downhole Array (GVDA) and numerical applications for the other cases. The identified amplification function is best fitted with the CSR than with the SSR curves.

Nevertheless, despite the observed differences in the SSR and CSR peak amplitudes, a good agreement was observed between actual and identified parameters within a margin of error of 20%. Moreover, this algorithm seems to be very efficient as long as the optimization is guided by an objective function which takes into account the form of the amplification function. For numerical examples, results show that the present algorithm works well and the curve fit was always very good. However, neither the growing number of parameters nor the vastness of the search space tends to reduce the efficiency or the robustness of the present algorithm. Nevertheless, exact identification of parameters of different models is possible only if these parameters are independents. So, if some parameters are known, the remainder can be identified exactly.

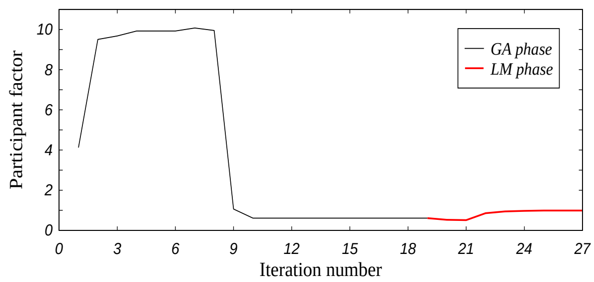
This proposed algorithm is very operative in estimating the characteristics of soil profiles and site amplification required in earthquake hazard analyses, with a much reduced cost compared to the laboratory or in-situ conventional methods, taking advantage of existing seismic ground motion records. It may be easily extended to extract soil parameters of two soil profiles from only free field acceleration records. On the other hand, to be efficiently applied to strong ground motions, an adequate model function in the framework of nonlinear analysis has to be used.



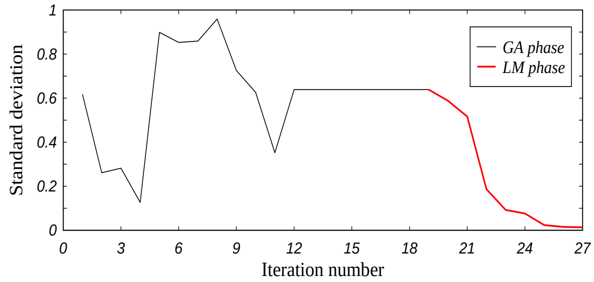
(a) Convergence history of the identified damping factor.



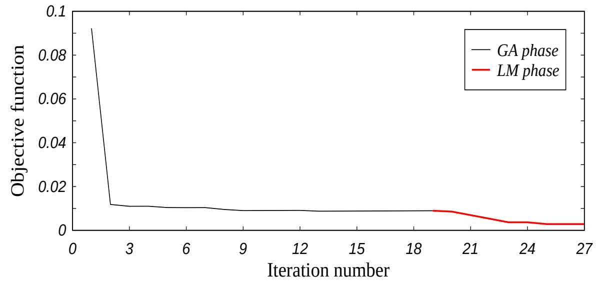
(b) Convergence history of the identified natural frequency.



(c) Convergence history of the identified participant factor.



(d) Convergence history of the identified standard deviation.



(e) Convergence history of objective function.

Fig. 12. Convergence history of the optimized parameters.

Appendix A.

The TAFs for a harmonic motion with frequency, $\omega=2\pi f$, are given by (Harichane et al., 2005).

$$T_1^{hm}(\omega) = \frac{2^{N+1}}{\sum \Lambda_{ml}} \qquad m, l = 1,2\,, \tag{A.1}$$

and

$$T_2^{hm}(\omega) = \frac{2^{N}}{\sum \Lambda_{1m}} \qquad m = 1,2\,, \tag{A.2}$$

where $T_1^{hm}(\omega)$ and $T_2^{hm}(\omega)$ are, respectively, the theoretical soil-to-bedrock and soil-to-rock outcropping TAFs, respectively, and Λ_{ml} are the components of the 2×2 matrix Λ which is:

$$\Lambda = \prod_{j=N}^{1} \begin{bmatrix} (1+q_j)e^{ik_jh_j} & (1-q_j)e^{-ik_jh_j} \\ (1-q_j)e^{ik_jh_j} & (1+q_j)e^{-ik_jh_j} \end{bmatrix}, \tag{A.3}$$

where $k_j = \omega / v_{Sj}$ is the wave number and

$$q_j = \rho_j v_{Sj} / \rho_{j+1} v_{Sj+1}\,. \tag{A.4}$$

The impedance ratio at the interface between layers j and j+1, v_{Sj} ($v_{Sj} = \sqrt{G_j/\rho_j}$) is the shear wave velocity, h_j is the layer thickness and i ($i^2 = -1$) the complex number. For a damped case, the formally treated equations still hold, by replacing shear modulus G_j by $G_j\left(1 + 2i\xi_j\right)$.

The soil properties (shear modulus G_j, mass density ρ_j and damping ratio ξ_j) are constants within each layer j but different among layers.

References

Deb K (2001). Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, Inc. New York, NY, USA.

Djilali Berkane H, Khellafi MA, Harichane Z, Kouici W (2014). Stochastic study of the spatial variation of ground response spectrum. *Electronic Journal of Geotechnical Engineering*, 19, 4345-4361.

Dunning T (1998). Recorded Step Directional Mutation for Faster Convergence. In Evolutionary Programming VII. Springer Berlin Heidelberg, Germany.

Guellil ME, Harichane Z, Djilali Berkane H, Sadouki A (2017). Soil and structure uncertainty effects on the soil-foundation-structure dynamic response. *Earthquakes and Structures*, 12(2), 153-163.

Goldberg DE (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA.

Govindaraju L, Bhattacharya S (2012). Site-specific earthquake response study for hazard assessment in Kolkata city, India. *Natural Hazards,* 61(3), 943–965.

Harichane Z, Afra H, Elachachi SM (2005). An identification procedure of soil profile characteristics from two free field accelerometer records. *Soil Dynamics and Earthquake Engineering*, 25(6), 431-438.

Harichane Z, Afra H, Bahar R (2012). Experimental validation of an identification procedure of soil profile characteristics from free field acceleration records, *International Journal of Geotechnical Earthquake Engineering*, 3(1), 1-17.

Ince GC (2011). The relationship between the performance of soil conditions and damage following an earthquake: a case study in Istanbul, Turkey. *Natural Hazards and Earth System Sciences,* 11(6), 1745-1758.

Jones AL, Kramer SL, Arduino P (2002). Estimation of uncertainty in geotechnical properties for performance-based earthquake engineering. PEER Report 16, Pacific Earthquake Engineering Research Center, University of California, Berkeley, USA.

Kao CY, Chung JK, Yeh YT (2010). A comparative study of the least squares method and the genetic algorithm in deducing peak ground acceleration attenuation relationships. *Terrestrial, Atmospheric and Oceanic Sciences*, 21(6), 869-878.

Khellafi MA, Harichane Z, Hamid A, Sadouki A (2013). A case study of accelerometric records analysis of May 21st, 2003, Boumerdes (Algeria) Earthquake. *International Journal of Geotechnical Earthquake Engineering,* 4(2), 34-52.

Khellafi MA, Harichane Z, Hamid A, Erken A (2016). Prediction of parameters of soil stratums and earthen dams from free field acceleration records. *International Journal of Geotechnical Earthquake Engineering,* 7(1), 33-56.

Koh CG, Perry MJ (2007). Structural damage quantification by system identification. *Journal of Earthquake and Tsunami*, 1(3), 211-231.

Kokusho T, Aoyagi T, Wakynam A (2005). In situ soil-specific nonlinear properties back-calculated from vertically array records during 1995 Kobe earthquake. *Journal of Geotechnical and Geoenvironmental Engineering,* 131(12), 1509-1521.

Koza JR (1992). Genetic Programming, On the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, MA, USA.

Li Z (2014). Uncertainty of soil properties in earthquake ground-motion site response analyses. Tenth U.S. National Conference on Earthquake Engineering Frontiers of Earthquake Engineering, Anchorage, Alaska, USA, July.

Marquardt DW (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11, 431-441.

Mercado V, El-Sekelly W, Zeghal M, Abdoun T (2015). Identification of soil dynamic properties through an optimization analysis. *Computers and Geoterchnics*, 65, 175–186

Pecker A, Mohammadioun B (1991). Downhole instrumentation for the evaluation of non-linear soil response on ground surface motion. *Proceedings of the 11th Structural Mechanics in Reactor Technology Conference*, Tokyo, Japan, August.

Pezeshk S, Zarrabi M (2005). A new inversion procedure for spectral analysis of surface waves using a genetic algorithm. *Bulletin of the Seismological Society of America*, 95(5), 1801-1808.

Rathje EM, Navidi S (2013). Identification of site parameters that improve predictions of site amplification, PEER Report 18; University of California, Berkeley, USA.

Rodríguez-Zúñiga JL, Ortiz-Alemán C, Padilla G (1997). Application of genetic algorithms to constrain shallow elastic parameters using in situ measurements. *Soil Dynamics and Earthquake Engineering*, 16(3), 223-234.

Rokonuzzaman MD, Sakai T (2010). Calibration of the parameters for a hardening-softening constitutive model using genetic algorithms. *Computers and Geotechnics*, 37(4), 573-579.

Sadouki A, Harichane Z, Chehat A (2012). Response of a randomly inhomogeneous layered media to harmonic excitations. *Soil Dynamics and Earthquake Engineering*, 36, 84-95.

Sato T, Yoshida I, Adachi Y (2013). Optimization of seismic sensor locations along highway links. *Journal of Earthquake and Tsunami*, 7(2), 11p.

SHAKE (2000). A computer program for conducting equivalent-linear seismic response analyses for horizontally layered soil deposits, a modified PC version of the original SHAKE program published in 1972 by Schnabel, Lysmer and Seed (modifications made by Idriss IM, Sum JI). EERI, University of California, Berkley, USA.

Steidl JH, Tumarkin AG, Archuleta RJ (1996). What is a reference site?. *Bulletin of the Seismological Society of America*, 86(6), 1733-1748.

Şafak E (1995). Discrete-time analysis of seismic site amplification. *Journal of Engineering Mechanics*, 121(7), 801-809.

Tallett-Williams S, Gosh B, Wilkinson S, Fenton C, Burton P, Whitworth M, Datla S, Franco G, Trieu A, Dejong M, Novellis V, White T, Lloyd T (2016). Site amplification in the Kathmandu valley during the 2015 M7.6 Gorkha, Nepal earthquake. *Bulletin of Earthquake Engineering*, 14, 3301–3315.

Walter E, Pronzato L (1997). Identification of Parametric Model from Experimental Data. Springer-Verlag, London, UK.

Wolf JP (1985). Dynamic Soil-structure Interaction. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA.

Zerva A, Harada T (1997). Effect of surface layer stochasticity on seismic ground motion coherence and strain estimates. *Soil Dynamics and Earthquake Engineering*, 16(7), 445-457.