Ground Motion Record Selection Using Multi-objective Optimization Algorithms: A Comparative Study

Ali Kaveh1*, Roya Mahdipour Moghanni1, Seyed Mohammad Javadi1

1 Centre of Excellence for Fundamental Studies in Structural Engineering, School of Civil Engineering, Iran University of Science and Technology, Narmak, Tehran, 16846-13114, Iran
* Corresponding author, e-mail: alikaveh@iust.ac.ir

Received: 10 May 2019, Accepted: 25 June 2019, Published online: 08 August 2019

Abstract
Performing time history dynamic analysis using site-specific ground motion records according to the increasing interest in the performance-based earthquake engineering has encouraged studies related to site-specific Ground Motion Record (GMR) selection methods. This study addresses a ground motion record selection approach based on three different multi-objective optimization algorithms including Multi-Objective Particle Swarm Optimization (MOPSO), Non-dominated Sorting Genetic Algorithm II (NSGA-II) and Pareto Envelope-based Selection Algorithm II (PESA-II). The method proposed in this paper selects records efficiently by matching dispersion and mean spectrum of the selected record set and target spectrums in a predefined period. Comparison between the results shows that NSGA II performs better than the other algorithms in the case of GMR selection.

Keywords
ground motion record selection, multi-objective particle swarm optimization, non-dominated sorting genetic algorithm II, pareto envelope-based selection algorithm II

1 Introduction
Currently, time history dynamic analysis is used for different aims such as reliability analysis of structures, designing important structures and performance evaluation of structures. The accurate estimation of structural performance can only be obtained when suitable ground motion records are used to implement time history dynamic analysis. Different researches are performed in the case of an earthquake record selection problem. For instance, Jayaram et al. [1] proposed an algorithm to select ground motions that match the target response spectrum mean and variance. They used a greedy optimization algorithm to enhance the match between the target and the sample means and variances. Some of the researchers took advantage of Genetic algorithm to selection of appropriate earthquake records. As an example, Naeim et al. [2] presented a new approach to selection of a set of recorded earthquake ground motions that match a specified site design spectrum. They performed dynamic time history analysis under the records selected by the genetic-based algorithm. Bojórquez et al. [3] developed the Spectral Shape-Based Record Selection approach to enhance the similarity between target and mean spectrum using \( N_p \) (a parameter proxy of spectral shape) and Genetic Algorithms. They suggested that using the parameter \( N_p \) results in similar target and mean spectral shape. Yaghmaei-Sabegh et al. [4] proposed a new optimal record selection with scaling framework using real-permutation and binary permutation genetic algorithms. Their result illustrated that binary-permutation GA is slightly more reliable than real-permutation GA. Kaveh and Mahdavi [5] developed a method for spectral matching of ground motions utilizing the wavelet transform and enhanced colliding bodies optimization. They used wavelet transform to decompose the original ground motions to different levels and each level was multiplied by a variable. They employed enhanced colliding bodies optimization technique to calculate the variables in order to minimize the error between the response and target spectra. Baker and Lee [6] proposed an improved algorithm for selecting ground motions. Their algorithm consists of four main steps including screening...
a ground motion database for suitable motions, statistically simulating response spectra from a target distribution, finding motions whose spectra match each statistically simulated response spectrum, and then performing an optimization to further improve the consistency of the selected motions with the target distribution. Ghafoory-Ashtiany et al. [7] selected ground motion sets in such a way that reliable prediction of the mean collapse capacity of a particular group of structures was achieved. Their method was based on the categorization of SDOF systems into groups and selection of optimal subsets of ground motions for them. Kayhan et al. [8] proposed a harmony search based earthquake record selection and scaling solution in order to obtain ground motions which are compatible with the design spectrum. Najafi and Tehranizadeh [9] provided a very simple and practical procedure for selecting ground motions in addition to compare two common scaling methods based on the uniform hazard spectrum (UHS) method and presents scale factors of the selected ground motions related with these methods. Macedo and Castro [10] described an advanced record selection and scaling utilizing Harmony search. They presented SeEQ, a fully integrated framework that implements established methods for ground motion record selection and scaling. One of the major advances of their suggested tool is the possibility to compute the Conditional Mean Spectrum (CMS) for the European territory. Moschen et al. [11] used a multi-objective algorithm to select and scale earthquake records to match a target median spectrum and target dispersion with the aid of scaling records in order to fit mean spectrum and dispersion spectrums, and obtain mean and dispersion spectrums from a selected GMR set. The objective functions attempt to minimize the difference between the target dispersion and mean spectrums, and obtain mean and dispersion spectrums from a selected GMR set.

The remainder of this article is classified in five sections, as follows. Section 2 is devoted to the description of multi-objective algorithms and Section 3 presents the target design spectrum. The framework of the study is discussed in Section 4. In Section 5, application of the multi-objective algorithms in GMR selection is studied. Finally, some concluding remarks are provided in Section 6.

2 Multi-objective optimization

Metaheuristic algorithms are becoming an important part of modern optimization and they provide good solutions to different optimization problems [14–16]. In the field of earthquake selection and scaling, metaheuristics can be used to match the mean spectrum and design spectrum. Most of the studies in this regard are performed based on single objective algorithms (e.g. [2, 17]). On the other hand, some researches were carried out using multi-objective algorithms. For instance, Mergos and Sextos [18] selected a set of ground motions using multi-objectives genetic algorithm. Georgioudakis et al. [19] selected and scaled records in order to fit mean spectrum and dispersion with target spectrum and dispersion with the aid of the Differential Evolution algorithm. A multi-objective problem can be defined by Eqs. (1) and (2).

Minimize: \( F(X) = f_1(X), f_2(X), \ldots, f_n(X), \)  
Subject to: \( R_{i}^{\text{lower}} \leq x_i \leq R_{i}^{\text{upper}}, i = 1, 2, \ldots, m, \)

where \( n \) is the number of objective functions, \( m \) is the number of variables and \( R_{i}^{\text{lower}} \) and \( R_{i}^{\text{upper}} \) are the boundaries related to variable \( i \).

2.1 MOPSO algorithm

MOPSO is one of the well-known metaheuristic algorithms suggested by Coello Coello and Lechuga [20] to solve multi-objective optimization problems. It uses the concept of Pareto dominance in PSO, and allows it to deal with multi-objective optimization problems.

PSO, developed by Kennedy and Eberhart in 1995 [21], is a population-based optimization technique inspired by the social behavior of fish and bird schooling. In PSO, a
population includes a certain number of particles. Each particle is characterized by its velocity and position. It is assumed that there are \(N\) particles in the swarm where the position and velocity of the \(i\)-th particle is represented by \((x_i = x_{i1}, x_{i2}, \ldots, x_{in})\) and \((v_i = v_{i1}, v_{i2}, \ldots, v_{in})\). Each particle \(i\) memorizes its best obtained position which is described by \((p_{besti} = p_{besti1}, p_{besti2}, \ldots, p_{bestin})\) and the best one among all \(p_{besti}\) in the swarm is recognized as globally best position and noted by \(g_{best}\). Each particle is evolved by updating its positional information by considering its and global best position, as demonstrated in Eqs. (3) and (4).

\[
v_i(t + 1) = w v_i(t) + c_1 r_1 (p_{besti} - x_i(t)) + c_2 r_2 (g_{best} - x_i(t)), \tag{3}
\]

\[
x_i(t + 1) = x_i(t) + v_i(t + 1), \tag{4}
\]

where \(i\) is the iteration number, \(w\) is the inertial weight, \(c_1\), and \(c_2\) are two learning factors related to the personal and global best particles, respectively. \(r_1\) and \(r_2\) are two random numbers generated uniformly in the range \([0, 1]\).

MOPSO is an extended form of the PSO which takes advantage of Pareto dominance concept to tackle with MOPs. According to the concept of Pareto dominance, the historical record of the best solutions found by a particle could be used to store non-dominated solutions generated in the past. The use of global attraction mechanisms combined with a historical archive of previously found non-dominated vectors would motivate convergence towards globally non-dominated solutions. Therefore, it is based on the idea of having a global repository in which every particle will deposit its experiences after each cycle. Furthermore, the updates to the repository are performed considering a geographically-based system defined in terms of the objective function values of each individual.

The archive controller and the grid, two components are used by MOPSO. The archive controller performs as the decision-maker for the addition and removal of solutions in the archive. In every iteration, each non-dominated solution found by the initial population is compared with every solution which is not inside the repository. In the grid system, each time new solutions are added into the archive, the grid space will adapt to accommodate solutions that are outside the boundary of the current grid. The adaptive grid is an objective function space which is divided into regions. This space is formed by hypercubes and its components are equal to a number of objective functions. The flowchart of the MOPSO is shown in Fig. 1. The main steps of MOPSO algorithm are as follows:

1. Initialize the population \((x_i\) where \(i\) is the number of population).
2. Initialize the velocity for each particle \((v_i = 0)\).
3. Evaluate each particle.
4. Store non-dominated particles’ position in the repository \((Rep)\).
5. Generate hypercubes of the search space explored so far, and locate the particles using these hypercubes as a coordinate system where each particle’s coordinates are defined in line with the values of its cost functions.
6. Initialize the memory of each particle by storing its initial position as the best found position as follows: \(x_{besti} = x_i\)
7. Compute the velocity of each particle as \(v_i = w v_i + r_1 (x_{best} - x_i) + r_2 (Rep - x_i)\)
8. Compute the new position of each particle as follow: \(x_i = x_i + v_i\)
9. Maintain the particles within the search boundaries and avoid the generation of solutions out of the boundaries.
10. Evaluate each of the particles.
11. Update the contents of \(Rep\) and insert non dominated solutions into the repository and any dominated solutions in the repository are eliminated.
12. Update each particle best position by replacing it with the current best solution.
13. Repeat step 7 until termination criteria are satisfied.

### 2.2 NSGA-II Algorithm

NSGA-II is a popular non-domination based genetic algorithm for multi-objective optimization problems which is developed by Deb et al. [22]. It is one of the most efficient algorithms to solve multi-objective optimization problems. NSGA-II takes into account non-dominated sorting
concept for fitness assignments in order to all non-dominated individuals are assigned front number 1 and individuals only dominated by individuals in front number 1 are assigned front number 2 and so on. At the next step, the selection phase is conducted by tournament selection, and individual with law front number is selected. If two individuals are from the same front, the individual with the highest crowding distance is selected. In each iteration, offspring are generated using Simulated Binary Crossover (SBX) and polynomial mutation. The flowchart of the algorithm is illustrated in Fig. 2.

2.3 PESA-II

Pareto Envelope-based Selection Algorithm II (PESA-II), developed by Corne et al. [23] is a multi-objective evolutionary optimization algorithm, which utilizes the mechanism of genetic algorithm together with selection based on Pareto envelope. PESA-II uses an external archive to store the approximate Pareto solutions. Parents and mutants are selected from this external archive, based on the grids created based on the geographical distribution of archive members. PESA-II is a multi-objective genetic algorithm, which uses grids to make selections and create the next generation. PESA-II algorithm consists of two population parameter, including, $P_r$, the size of the internal population (IP) and, $P_g$, the maximum size of the archive or external population (EP). The main steps of PESA-II algorithm are as follows:

1. Generate and evaluate each chromosome of an initial population of IP, and initialize the external population to the empty set.
2. Incorporate the non-dominated members of IP into EP.
3. If a termination criterion has been reached, then stop, returning the set of chromosomes in EP as a result. Otherwise, delete the current contents of IP, and repeat the following until ($P_f$) new candidate solutions have been generated:
   - With probability, $P_C$, select two parents from EP, produce a single child via crossover, and mutate the child. With probability ($1-P_C$), select one parent and mutate it to produce a child.
4. Return to step 2.

The steps of the algorithm are shown in Fig. 3.
3 Target design spectrum

The study aims to select suitable ground motion set for a site located in Los Angeles on soil type C. The design acceleration response spectrum is obtained from ASCE 7-16 standard [24] defined as:

\[
S_{DS} \begin{cases} 
0.4 + 0.6 \frac{T}{T_0} & T < T_0 \\
S_{DS} & T_0 \leq T \leq T_s \\
\frac{S_{DS}}{T} & T_s \leq T \leq T_L \\
\frac{S_{DS}}{T^2} & T > T_L 
\end{cases},
\]

in which \(S_{DS}\) and \(S_{D0}\) denote the design spectral response acceleration parameter at short periods and the design spectral response acceleration parameter at \(T = 1\) sec, respectively. In the above equation, \(T_0\) and \(T_0\) are defined by Eqs. (6) and (7). The target spectrum is illustrated in Fig. 4.

\[
T_0 = 0.2 \frac{S_{DS}}{S_{DS}}, \\
T_s = \frac{S_{DS}}{S_{DS}}.
\]

4 The framework of the study and target spectrums

The proposed ground motion approach in this study consists of three main steps that are described as follow:

a. Preselection of ground motions according to characteristics of the considered site from a database. In this paper, 202 ground motion records are selected for the selected site from PEER ground motion database [25]. The following criteria are considered in the preselection of records:
- Moment Magnitude of 5.5 to 7.9,
- Shear velocity for soil type C (very dense soil and soft rock) based on ASCE 7-16 [24]: \(366 \leq v_s \leq 762\),
- The focal distance of 0 to 80 km as suggested in other studies ([11, 26]),
- Acceleration spectral ordinates are obtained for damping ratio of 5%.

One of the advantages of the preselection of GMRs is reducing the number of possible combinations of the GMRs. The number of possible combinations of records can be defined by Eq. (8).

\[
\binom{k}{n} = \frac{k!}{n!(k-n)!},
\]

where \(n\) is the total number of GMRs (202 in this study) and \(k\) in the number of record set where in this study is assumed 7.

b. Performing the optimization process in order to select ground motion sets and scale factors. In this step, multi-objective algorithms are implemented to select record sets that minimize the considered objective functions.

c. In this step, the final solution should be selected from Pareto optimal solutions for as much as several solutions lie in the repository.

According to ASCE 7-16, it is recommended to perform nonlinear dynamic time history analysis by a set of 7 ground motions. Thus, the problem can be defined as the selection of 7 ground motion set out of sample space and an appropriate scale factor (SF) for each of them. The sample space can be shown by an \(m \times n\) matrix where \(m\) is the number of discreet periods and \(n\) is the number of ground motions.

\[
S_A = \begin{bmatrix} S_{a11} & S_{a12} & S_{a13} & \cdots & S_{a1n} \\
S_{a21} & S_{a22} & S_{a23} & \cdots & S_{a2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
S_{an1} & S_{an2} & S_{an3} & \cdots & S_{ann} \end{bmatrix} = [S_A, S_A, S_A, \ldots, S_A],
\]

where \(S_A\) denotes as the spectral acceleration associated with the \(i\)-th GMR. It is assumed that the spectral acceleration is distributed log normally, therefore the geometric mean of the \(i\)-th point of the period range can be computed as

\[
\text{mean}_i = \exp \left( \frac{1}{n} \sum_{j=1}^{n} (SF_j \ast \ln(S_{aj})) \right).
\]

In the above equation, \(SF\) represents the scale factor of the \(j\)-th GMR. The earthquakes obtained by only matching the mean spectrum with the target spectrum underestimates the response variability; thus, provide no insight regarding the dispersion around the mean. In order to
consider the record to record variability, it is important to consider minimizing the difference between target variance and variance of the selected records in predefined period set. Therefore, it is important to fit the target dispersion and dispersion around the mean. The dispersion around the mean spectrum in the period range is given by

$$\sigma_j = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} \left( \ln(S_{ij} \ast SF_j) - E\left( \ln(SA_j \ast SF_j) \right) \right)^2}.$$  

(11)

For each period, Eq. (11) leads to a constant standard deviation of the natural logarithm of the elements of the random vector, SA, over the predefined period. In order to find appropriate record set, period range should be defined to cover all the periods in which structural response is sensitive. In this regard, ASCE 7-05 has specified the period range 0.2 $T_1$ to 1.5 $T_1$ as an effective period range for mid-rise buildings. Therefore, in this study, the period range is assumed to be 0.2 to 1.5 sec based on the period of the considered structure.

5 Application of the multi-objective algorithms in GMR selection

After preselection phase of ground motions, the considered optimization algorithm should be applied to obtain optimal solutions. The first objective function of the problem is to minimize the difference between target mean spectrum and mean of the ground motion set over the selected period range. The second objective is to minimize the dispersion around the mean and target dispersion. The target dispersion is constant and assumed to be equal to 0.8 in all periods as recommended by different studies (e.g. [27] and [28]). The multi-objective function is defined by Eq. (12).

$$\left\{ \begin{array}{l}
F_1 = \text{minimize} \sqrt{\frac{1}{m} \sum_{i=1}^{m} (S_{ij} - \text{mean})^2} \\
F_2 = \text{minimize} \frac{1}{m} \sum_{i=1}^{m} (\sigma_i - \sigma_{\text{ref}})^2
\end{array} \right. ,$$  

(12)

where $\sigma_i$ refers to target dispersion and $S_{ij}$ is the value of the target spectrum at period $i$.

In order to solve the optimization problem, each member of the population is represented by a vector as shown in Fig. 5. As shown in Fig. 5, appropriate records and scale factors for each of them should be selected in optimization phase. But, in the study performed by Moschen et al. [11], the same scale factor has been considered for the selected ground motion set. The total number of 202 GMRs are selected from PEER database, and the three mentioned multi-objective algorithms are used in order to select a suitable set of GMRs. Fig. 6 shows the Pareto fronts obtained from each algorithm and Fig. 7 represents the Pareto optimal solution obtained from three algorithms. The spectra of the selected GMRs are plotted in Fig. 8. As it can be seen from this figure, the dispersion spectrum related to GMRs selected by NSGA II algorithm has the best match in comparison with the other two dispersion spectrums. In order to compare the performance of the three considered multi-objective algorithms, the ten best members of the Pareto-front solutions from the result of each algorithm are selected. The error between the mean spectrum of each Pareto-front member and mean target spectrum are calculated and ten members with lower maximum errors in the predefined period set are selected. The amount of errors is plotted in Fig. 9 which shows that the errors related to NSGA II algorithm are lower than those of the others and

<table>
<thead>
<tr>
<th>Earthquake name</th>
<th>Station</th>
<th>Magnitude</th>
<th>Scale Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Livermore-01</td>
<td>Fremont - Mission San Jose</td>
<td>5.8</td>
<td>4.46</td>
</tr>
<tr>
<td>Big Bear-01</td>
<td>Wrightwood - Nielson Ranch</td>
<td>6.46</td>
<td>4.67</td>
</tr>
<tr>
<td>Landers</td>
<td>Fort Irwin</td>
<td>7.28</td>
<td>3.93</td>
</tr>
<tr>
<td>Landers</td>
<td>Barstow</td>
<td>7.28</td>
<td>5.42</td>
</tr>
<tr>
<td>Landers</td>
<td>Silent Valley - Poppet Flat</td>
<td>7.28</td>
<td>3.53</td>
</tr>
<tr>
<td>Livermore-01</td>
<td>APEEL 3E Hayward CSUH</td>
<td>5.8</td>
<td>3.91</td>
</tr>
<tr>
<td>Mammoth Lakes-02</td>
<td>Convict Creek</td>
<td>5.69</td>
<td>1.31</td>
</tr>
</tbody>
</table>

Table 1 Selected GMRs in a Pareto-front member obtained from NSGA II algorithm
MOPSO gives the higher value of errors. The spectral spectrums of ten best members and the mean spectrum of each member are plotted in Fig. 10. From this figure, it is clear that NSGA II has a better performance in comparison to other algorithms and GMRs selected by NSGA II algorithm have higher matching with target spectrum. For instance, one of the Pareto-members obtained from NSGA II algorithm is shown in Table 1. It is worth mentioning that different objective functions can be employed in order to select appropriate records. For instance, the direction of records can be considered in selection of the records in order to horizontal and vertical components of ground motions be compatible with target spectrum.

6 Conclusions
In this paper, an optimization based record selection approach is proposed. The optimization problem is solved using three different multi-objective algorithms including NSGA II, MOPSO, and PESA II. The goal of the study is to select appropriate ground motion set with their scale factors in order to match a target spectrum and dispersion of spectral values in a predefined period set. A predefined record to record variability is considered in selection of earthquake set by taking into account second objective. The 202 GMRs are selected from PEER database based on the preselection criteria including magnitude, shear velocity, focal distance, and damping ratio. The problem is solved by three multi-objective optimization algorithms and the results are compared. This investigation shows that NSGA II has better performance than the other algorithms, and the GMRs selected by NSGA II has a better match with target spectral acceleration and target dispersion in the period set. In selection of earthquakes by multi-objective optimization approach, a set of suitable solutions exists in Pareto-optimal solutions. Each solution is a ground motion record set with the same quality in terms of the fitness score because there is no designation as to which objective, matching the median or the dispersion, is more important. One of these solutions should be selected in the post processing phase.

Conflict of interest
On behalf of all authors, the corresponding author states that there is no conflict of interest.
Fig. 7 The left column stands for all GMRs in Pareto-front, and the right column stands for the mean of the GMRs in every member of the Pareto-front: NSGA II (a); MOPSO (b), and PESA II (c)
Fig. 8 The dispersion spectrums: NSGA II (a); PESA II (b), and MOPSO (c)

Fig. 9 The errors of the ten members of Pareto-front with minimum errors
Fig. 10 The left column stands for all GMRs in 10 Pareto-front members having minimum errors and the right column stands for the mean of the GMRs in every 10 selected members of the Pareto-front: NSGA II (a); MOPSO (b), and PESA II.

References

https://doi.org/10.1193/1.3608002

https://doi.org/10.1193/1.1719028


