City Research Online

City, University of London Institutional Repository

Selection of earthquake ground motions for multiple objectives using genetic algorithms

Panagiotis E. Mergos[1*], Anastasios G. Sextos[2, 3]

[1] Department of Civil Engineering, City, University of London, London EC1V 0HB, UK
[2] Department of Civil Engineering, Aristotle University, 54124 Thessaloniki, Greece
[3] Department of Civil Engineering, University of Bristol, Bristol BS8 1TR, UK

Abstract. *Existing earthquake ground motion (GM) selection methods for the seismic assessment of structural systems focus on spectral compatibility in terms of either only central values or both central values and variability. In this way, important selection criteria related to the seismology of the region, local soil conditions, strong GM intensity and duration as well as the magnitude of scale factors are considered only indirectly by setting them as constraints in the pre-processing phase in the form of permissible ranges. In this study, a novel framework for the optimum selection of earthquake GMs is presented, where the aforementioned criteria are treated explicitly as selection objectives. The framework is based on the principles of multi-objective optimization that is addressed with the aid of the Weighted Sum Method, which supports decision making both in the pre-processing and post-processing phase of the GM selection procedure. The solution of the derived equivalent single-objective optimization problem is performed by the application of a mixed-integer Genetic Algorithm and the effects of its parameters on the efficiency of the selection procedure are investigated. Application of the proposed framework shows that it is able to track GM sets that not only provide excellent spectral matching but they are also able to simultaneously consider more explicitly a set of additional criteria.*

Keywords: seismic assessment; ground motion selection and scaling; response-history analysis; multi-objective optimization; genetic algorithms

1 Introduction

The demand for controlling the earthquake-induced damage to structures necessitates the use of accurate structural analysis procedures. Clearly, if the GM was known, the most accurate procedure for determining seismic demands would be the rigorous nonlinear response-history analysis with step-by-step integration of the equation of motion in the time domain. Nonlinear response-history analysis is used as the reference method for determining seismic demands in modern design guidelines such as

---

[*] Corresponding author. Panagiotis E. Mergos, Lecturer in Structural Engineering, Department of Civil Engineering, City, University of London, EC1V 0HB, London, UK.
E-mail address: panagiotis.mergos.1@city.ac.uk, Tel. 0044 (0) 207040 8417

the *fib* Model Code 2010 [1, 2], however, key is the selection of GM records to serve as input excitations. This is because previous studies (e.g. [3, 4]) have shown that the selection of earthquake-induced GMs represents the most significant source of uncertainty in the calculation of seismic demands undermining the reliability of the, otherwise rigorous, response history analysis. Therefore, efficient methodologies for the reliable selection of GMs in seismic assessment and design of structures are essential.

Depending on the information available and the nature of the problem under investigation, synthetic, artificial, modulated or real records can be used with or without scaling [5]. Synthetic motions are generated by simulating the entire propagation from the earthquake rupture (source) to the site of interest for a scenario of given magnitude, distance and source seismological characteristics. Artificial records aim to match almost perfectly a target response spectrum through random vibration procedures [6]. The facts that there are several tools available for generating artificial GMs (e.g. [7]) and that low record-to-record variability is easily achieved explain why this procedure is acceptable, if not implicitly promoted, by most modern seismic codes. On the other hand, artificial records tend to lack characteristics of real earthquake GMs, such as the time-varying intensity and frequency content that can significantly affect the structural response. Moreover, aiming for spectral matching within a wide range of periods, particularly when the target is the code-based design spectrum, they may include unrealistic pulses, while the duration of motions is inevitably constant, effectively integral in the generation procedure. For all the above reasons, real records are often appropriately modulated via wavelet adjustment [8, 9] or other techniques to achieve the desirable spectrum compatibility without losing important features of earthquake GMs. Again some limitations also exist in this case, as the modulating procedures may drive to unrealistic strong motion duration and number of cycles leading to overestimation of cumulative damage effects on structures [5, 10]. With the rapid increase of the real GM records that are currently available online via several open databases, selection and scaling of an appropriate subset of real motions is deemed an attractive alternative as long as their variability, frequency content and cumulative damage potential is controlled. Use of real record suites is long permitted by modern seismic codes (e.g. ASCE7/SEI-2016 [11] and Eurocode 8 [12]) and is increasingly employed in practice.

The key difference in real GM selection procedures, however, is whether they:

(a) are based on preliminary *seismological and site parameters* for a given earthquake scenario that are roughly similar to those likely to cause the GM intensity of interest [13] to be followed by *matching* (in fact, exceed) a threshold percentage of the *average* spectrum of the seed motions versus a target, typically code-based, spectrum. This procedure is the standard method in modern seismic codes. In some variations that will be discussed below, GM variability is also kept under control.

(b) aim to relax the preliminary criteria and directly select GMs based on their *spectral shape*, the rational being that the latter is a more important parameter compared to the preliminary criteria, which are only imperfect proxies of the frequency content and duration of GMs and naturally, hence not well correlated to seismic demand [14-16]. In this line of thinking, as long as the spectral shape of the GMs is compatible, at least the elastic seismic demand imposed on the structure is expected to be equivalent, irrespectively of how well the preliminary criteria were satisfied.

In the first context, typically, the selection and scaling of sets of recorded accelerograms follows two basic phases. In the *pre-processing* (or, *pre-selecting* according to the ASCE/SEI-2016 [11] terminology) phase, a pool of eligible GMs is determined that satisfies a number of preliminary criteria such as adequate compliance with the regional seismicity (in terms of earthquake magnitude and source-to-site distance as defined by a deaggregation of the results of the Probabilistic Seismic Hazard Analysis, PSHA), local soil conditions (shear wave velocity of the upper 30m) at the site of interest, the tectonic environment and earthquake rupture mechanism. Other criteria may include usable frequencies and sampling of GMs, some restrictions as per the number of GMs that can be used from the same event, strong ground motion intensity and duration as well as a minimum and maximum limit for the permissible scale factors to prevent bias to structural responses by excessive scaling [17].

For all these criteria, predetermined permissible ranges are determined by the engineer. It has been reported [10, 18] and is also quite anticipated, that the application of a large number or stringent pre-selection criteria may reduce the number of eligible records and hence, hinder the spectrum compatibility process that follows.

Depending on the number *m* of the eligible GMs that satisfy the preliminary criteria set and the desirable size *n* of the GM subsets ($n \leq m$) different combinations of recorded motions are formed to be examined and prioritized. The total number of these possible combinations is given by Eq. (1). The usual design option for *n* is 7, since this is typically the minimum number required by most of the code provisions to permit the use of average response quantities as design values.

$$\binom{m}{n} = \frac{m!}{\mathrm{n}!(m-n)!} \tag{1}$$

The second (*processing*) phase of the GM selection procedure typically involves a heuristic procedure searching for sets of uniformly scaled, non-uniformly scaled or unscaled GMs that provide optimum compatibility with the pre-defined target response spectrum. The aim is to ensure that the average spectral values, of the scaled motions included in the set, will match spectral ordinates of the target spectrum, within a pre-scribed period range around the fundamental period of vibration. While assessing spectral matching of the eligible GM sets that can be formed by the *n* candidate records, additional criteria have been imposed to control GM variability in terms of spectral acceleration [18-19] and spectral displacement [20], or structural response dispersion [21], as well as perform selection using both spectral and waveform matching in addition to conducting signal processing, response spectra analysis and soil response analysis [22].

The *spectral-shape* approach on the other hand, is employing the Conditional Mean Spectrum (CMS) [23] as the target spectrum with parameter ε (epsilon). Kottke and Rathje [24] developed a semi-automated procedure that was able to select and scale GMs to fit a target acceleration response spectrum, while controlling the variability within the GM set itself. Given that CMS does not account for the aleatory variability in the response spectrum and the fact that, by definition, spectral values, only represent the peak response of a single-degree-of- freedom (SDOF) oscillator of a specific period, Bradley [25] considered conditional values of other GM properties (e.g. significant duration SD, Arias Intensity I_a etc) by introducing the generalized conditional intensity measure (GCIM), thus permitting to capture total and temporal duration and energy features of ground shaking. Jayaram *et al*. [26] explicitly considered both the response spectrum mean and variance in a Conditional Spectrum (CS) context, while Lin *et al*. [27] extended to CMS further, introducing the Exact Conditional Spectrum (ECS) that accounts for multiple causal earthquakes (as opposed to one in the standard CMS approach) and multiple GMPMs that are often considered in a PSHA computation. To overcome the ambiguity of choosing a scalar conditioning intensity measure IM, the need for amplitude scaling the GMs to the above IM and the potential bias in seismic demand estimate and the need for several sets of GMs to derive the seismic demand hazard curve (SDHC) an alternative method has been proposed by Kwong *et al*. [28] leading to a single set of unscaled GMs. Furthermore, recent studies [29-31] consider matching more than one selection objectives, which are the target response spectrum mean and variance as well as the correlation structure. This is achieved by using an objective function that is the weighted sum of the individual objective functions. Detailed reviews of the available methods for GM selection can be found in [32-34], while a comparative assessment as per the ECS and the GCIM is presented in Kwong and Chopra [35] using synthetic GMs.

All previous efforts aim at selecting a single GM set. Recently, Moschen *et al*. [36] proposed a multi-objective optimization framework with the aim to obtain a family of GM sets that simultaneously match (a) their median pseudo-acceleration spectrum with a target spectrum and (b) a target dispersion of spectral values on a predefined period range given that the same scale factor is used for each GM. As a result, a Pareto-front of selections is established that provides the best trade-offs between the two selection objectives. Next, a post-processing phase is required, where the user selects the most suitable solution of the Pareto-front based on engineering judgement and/or decision-making theory.

Notwithstanding the significant advances made with respect to GM selection procedures, there is still a level of ambiguity as per whether the spectral shape is adequate to represent fully important GM characteristics such as duration and cumulative seismic demands that can affect significantly the response of structural systems [37-38] and are solely considered in seismic assessment by nonlinear response history analysis. This is particularly true in the framework of major existing seismic codes, such as Eurocode 8 [12], where a single type (Type 1 and Type 2) target spectrum is only to be matched for all sites in high and low seismicity regions respectively.

Independently of the selection objectives and constraints, the use of optimization algorithms to facilitate the GM selection process becomes increasingly popular. In this context, the use of metaheuristic algorithms is particularly beneficial as it is more computationally efficient than exhaustive search and it drives to near global optimal solutions without the need for calculating gradients of cost or constraints functions [39]. Additionally, it can be combined with variable and different, within the same set, scale factors increasing substantially the potential for minimizing the corresponding objective functions. Naeim *et al*. [40] were the first to implement Genetic Algorithms (GA) in the selection of ground motion sets. Macedo and Castro [41] applied the Adaptive Harmony Search algorithm to select recorded GMs for code-based and CMS-based spectral matching. Stefanidou *et al*. [42] used a greedy heuristic optimisation algorithm to facilitate GM selection specifically at the bedrock level combined with one-dimensional site response analysis of a bridge-soil system. Moschen *et al*. [36] used the NSGA-II GA [43] to support the multi-objective optimization framework discussed previously. Nevertheless, the existing literature seems to be lacking sufficient information and guidance regarding the efficient use of metaheuristic optimization algorithms as well as the corresponding implications in GM selection.

The purpose of this study is to contribute a simple, yet more versatile and computationally efficient GM selection approach that is fully compatible with major design codes and, at the same time, permits optimum matching to *multiple*, appropriately *weighted*, objectives that involve spectral compatibility with a target spectrum and compliance with target values of additional selection parameters. The latter may reflect local seismicity for a given earthquake scenario, local soil conditions and expected strong GM intensity and duration at the site of interest and the preferred level of GM scaling (i.e. unscaled spectra) in the selection procedure. In this way peak earthquake demands are considered along with cumulative seismic effects for structural systems assessed with the aid of rigorous nonlinear response history analysis. It is recalled that within the existing selection methods the above objectives are typically considered in terms of permissible ranges of their parameters in the pre-processing phase. Treating them explicitly as selection objectives in the proposed study, provides GM sets with the best trade-offs between spectral matching and compliance with their target parameter values in the most computationally efficient way.

The Weighted Sum Method (WSM) is used to address the multi-objective optimization problem of this study that supports decision making both in the pre-processing and post-processing phase of the selection of GM sets. It is noted that this is the first time that this method is employed to derive not a single but a family of Pareto-optimal GM sets. Therefore, the necessary steps of the applied method are described in detail.

The optimum selections of GMs of this study are performed by the use of an efficient GA that is able to track near global optimum solutions of constrained problems with both discrete and continuous design variables. A parametric application of this algorithm is also conducted herein to demonstrate the effects of its parameters on the selection of GMs to fill a part of this gap in the existing literature. The structure of the method and demonstration of its efficiency are presented in the following sections.

2 Multi-objective optimum GM selection methodology

Similarly to [36], the methodology used herein to select GM sets is divided into three distinct phases: pre-processing; processing and post-processing. Each of the phases are described in detail below.

2.1 Preliminary criteria for GM selection and scaling (Pre-processing)

In this phase, a number of pre-selection criteria are applied in order to select a pool of eligible GM records from a strong motion database such as the European Strong-motion Database (ESD) [44], the PEER GM Database [45] and others. As discussed previously, these criteria can be of seismological nature such as the fault rupture mechanism and directivity of the seismic waves, as well as the earthquake magnitude M and source-to-site distance R. Furthermore, they can be related to the local soil conditions at the site of the structure as quantified, typically, by the shear-wave velocity $V_{s,30}$ at the uppermost 30m. Additionally, strong GM intensity criteria can be used expressed in terms of parameters that are easily available in the online databases, such as the peak ground acceleration (PGA), peak ground velocity (PGV) and displacement (PGD) and the spectral acceleration $S_a(T_1)$ at the fundamental period T_1 of the structure under examination. Another criterion of interest, particularly in light of nonlinear analysis and cumulative seismic effects, is the significant duration of GM t_s also available in the most databases. Furthermore, upper and lower limits of the scaling factors Sf values can be introduced in this phase to limit the bias of the results of structural responses due to GM scaling.

The main difference in this phase of the proposed methodology with respect to the others is that the pre-selection criteria can be significantly relaxed depending on the adopted selection objectives that will be discussed in the following. In this manner, sufficient pools of eligible GMs can be used to ensure satisfactory spectrum compatibility. Furthermore, the tedious procedure, performed manually by the user, of re-adjusting the pre-selection criteria until sufficient spectrum matching is achieved can be avoided.

In addition, a target response spectrum must be specified in the pre-processing phase. Depending on the nature of the envisaged seismic assessment, the Uniform Hazard Spectrum UHS [46], the Conditional Mean Spectrum CMS [23] or a spectrum predicted for scenario events determined by appropriate empirical response spectra attenuation relationships (e.g. [47, 48]) can be used as target spectra. Furthermore, when seismic assessment is performed in the framework of a seismic code, the smooth elastic code spectrum should be set as target spectrum.

The next step in this phase, deals with defining the selection objectives. These objectives can be related directly to the parameters described previously for determining the permissible ranges of the pre-selection criteria. Setting these objectives will be discussed in greater detail in the processing phase of the GMs selection procedure.

A final, provisional, step in the pre-processing phase is the selection of appropriate fixed values of the weights of the selection objectives to be used in the WSM. This step supports decision making in this early phase of the selection procedure. Clearly, the values of these weights should be problem dependent for three reasons: i) the importance of the various objectives depends on the subject and the goals of the seismic assessment. For example, an increased weight for the significant duration of strong GM for the seismic assessment by means of nonlinear response history analysis of a structure susceptible to cumulative damage effects and a lower weight for a structure insensitive to cumulative damage can be used; ii) the values of the weights should reflect not only the importance of the objectives but also the level of confidence for their target values. It is not reasonable to provide high weight values to objectives with highly uncertain target values; iii) the trade-offs between the various objectives depend on the available pool of records.

Hence, in the general case, to select the weight values of the various objectives in the pre-processing phase, a limited number of weight combinations should be first selected based on engineering judgement and reflecting the subject and the goals of the seismic assessment as well as the level of confidence of the corresponding target values. Next the values of the objective functions should be evaluated and the combination of weights that "works best" for the problem under investigation should be adopted. In the cases, however, where one objective can be characterised as main and the rest as secondary, as it is often the case with the spectrum compatibility and the rest selection objectives, trial selections can be made by increasing gradually from zero the weights of the secondary objectives and monitor the results of the main objective function. The increase of the weights of the secondary objectives should be terminated when the main objective function worsens rapidly and/or when it exceeds a pre-specified goal value that is acceptable for this objective.

In any case, if fixing of these weights is not possible in the pre-processing phase then the full range of different weight values has to be assumed in the WSM to derive the Pareto-fronts of the selection objectives, as discussed in the following. Based on these fronts, decision-making will then take place in the post-processing phase.

2.2 Processing

In this phase, a single set of GMs is selected in the case of single-objective optimum selection or multi-objective optimum selection with fixed weight values determined in the pre-processing phase. Alternatively, a number of different sets of GMs on the Pareto-front of the selection objectives can be derived in the case of multi-objective optimum selection with no pre-fixed weight values. These Pareto-optimal solutions represent the best feasible trade-offs between the different selection objectives. There are no other selection solutions that can reduce one objective of the Pareto-optimal solutions without increasing another one at the same time, as shown in Fig. 1 for the case of two selection objectives. In the following, the steps followed to derive optimal GM sets are described in detail.

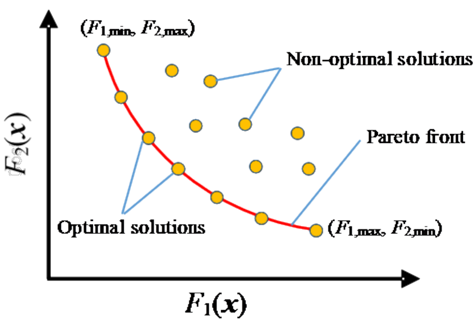


Fig. 1: Pareto-optimal solutions

Optimization problem formulation

In a general multi-objective optimization problem, the aim is to minimize a set of p objective functions $F_i(\boldsymbol{x})$ (i =1 to p) subject to q number of constraints $g_j(\boldsymbol{x}) \leq 0$ (j =1 to q). A selection is represented by the solution vector $\boldsymbol{x}$, which contains l number of independent selection variables x_k (t = 1 to l). The previous can be summarized as:

$$\begin{aligned} &\text{Minimize:} && [F_1(\boldsymbol{x}), F_2(\boldsymbol{x}), \dots, F_p(\boldsymbol{x})] \\ &\text{Subject to:} && g_j(\boldsymbol{x}) \leq 0 \ (j = 1 \text{ to } q) \\ &\text{Where:} && \boldsymbol{x} = (x_1, x_2, \dots, x_l) \end{aligned} \tag{2}$$

The WSM is the simplest and most intuitively meaningful means of solving a multi-objective optimization problem [39, 49]. In this method, the multi-objective optimization problem is solved as a single-objective optimization problem, where an equivalent single objective $F(\boldsymbol{x})$ has to be minimized. $F(\boldsymbol{x})$ is simply a weighted linear combination of all the individual objective functions $F_i(\boldsymbol{x})$. Therefore, the optimization problem is written as:

$$\begin{aligned} &\text{Minimize:} && F(\boldsymbol{x}) = \textstyle\sum_{i=1}^{p} w_i \cdot F_i(\boldsymbol{x}) \\ &\text{Subject to:} && g_j(\boldsymbol{x}) \leq 0 \ (j = 1 \text{ to } q) \\ & && \textstyle\sum_{i=1}^{p} w_i = 1 \\ &\text{Where:} && \boldsymbol{x} = (x_1, x_2, \dots, x_l) \end{aligned} \tag{3}$$

In Eq. (3), w_i are the values of the weights corresponding to the individual function objectives $F_i(\boldsymbol{x})$. In the simple case of a bi-objective optimization problem, shown in Fig. 1, Eq. (3) becomes:

$$F(\boldsymbol{x}) = w_1 \cdot F_1(\boldsymbol{x}) + w_2 \cdot F_2(\boldsymbol{x}) \tag{4}$$

As discussed, the values of the weights may be fixed at this stage based on decision making in the pre-processing phase. In this case, the optimal solution derived will be a single point on the Pareto-front in the objective solution space. Alternatively, by changing the weight values gradually from 0 to 1, a series of Pareto-optimal points is evaluated and thereby the Pareto-front is obtained.

Equation (4) is easy to implement when the individual objectives $F_i(\boldsymbol{x})$ refer to similar quantities. However, in the general case where the $F_i(\boldsymbol{x})$ (i = 1, 2) are of different nature then the selection of their corresponding weight values w_i is not straightforward. In this cases, it is recommended (e.g. [49]) to use a normalized version of the equivalent single-objective function $\bar{F}(\boldsymbol{x})$ given by:

$$\bar{F}(\boldsymbol{x}) = w_1 \cdot \bar{F}_1(\boldsymbol{x}) + w_2 \cdot \bar{F}_2(\boldsymbol{x}) \tag{5}$$

In Eq. (5), $\bar{F}_i(\boldsymbol{x})$ (i = 1, 2) are the normalized individual objective functions given by Eq. (6), where $F_{i,min}$ is the minimum value of $F_i(\boldsymbol{x})$ calculated by setting in the optimization problem of Eq. (3) that $F(\boldsymbol{x}) = F_i(\boldsymbol{x})$ and $F_{i,max}$ is the maximum value of $F_i(\boldsymbol{x})$ calculated by setting in the optimization problem of Eq. (3) that $F(\boldsymbol{x}) = F_j(\boldsymbol{x})$ with $j \neq i$ (j = 1, 2). The afore-described values are presented in Fig. 1.

$$\bar{F}_i(\boldsymbol{x}) = \frac{F_i(x) - F_{i,min}(x)}{F_{i,max}(x) - F_{i,min}(x)} \tag{6}$$

Having established $\bar{F}_i(\boldsymbol{x})$ optimal values, the corresponding $F_i(\boldsymbol{x})$ values are easily retrieved by solving Eq. (6). Similar considerations hold in the general case of more than two selection objectives.

Selection variables

The selection variables are the properties of the optimization problem that can change values during the solution procedure. In the selection of sets of GM records of size *n* examined herein, the first *n* design variables x_k (k = 1 to n) are the records serial numbers inside the pool of eligible *m* records ($1 \leq n \leq m$) and the second *n* design variables x_{k+n} (k = 1 to n) are the corresponding scale factors Sf_k (k = 1 to n). Therefore, the design vector $\boldsymbol{x}$ can be written as:

$$\boldsymbol{x} = (x_1, \dots, x_n, x_{n+1}, \dots, x_{2n}) \tag{7}$$

In this formulation, the first *n* design variables (serial numbers) are represented by integer values. The second *n* variables (scale factors) are represented by positive real numbers. It is noted that this formulation is the most general and allows for the maximum search space in the selection procedure. Clearly, different formulations of the scale factors design variables can be used depending on the selection constraints and the nature of the problem under investigation. For example, fixed scale factors can be used based on a pre-scaling procedure that matches all individual spectra to the target spectrum at a specific period. In this case, the design variables are reduced from 2*n* to *n*. Alternatively, a common variable scale factor can be used to preserve the relative intensity of the un-scaled records [36]. In the latter case, *n*+1 independent selection variables are required.

Selection constraints

The selection constraints $g_j(\boldsymbol{x})$ (j =1 to q) represent additional limitations in the selection and scaling of GMs procedure that are complementary of the selection criteria in the pre-processing phase. Typically, they represent criteria aiming at improving the reliability of spectral matching that are not addressed in the pre-processing phase. For example, Eurocode-8 (Part 1) [12] sets the following spectral matching limitations in the selection of sets of GMs:

i) The mean of the zero period acceleration values of the scaled individual response spectra should be larger than the zero period acceleration of the target code spectrum.

ii) In the range of periods between $0.2T_1$ and $2T_1$, no value of the mean 5% damping elastic spectrum, calculated from all time histories, should be less than 90% of the corresponding value of the 5% damping of the code elastic response spectrum.

Selection objectives

As discussed, typically, objectives in the selection of GMs procedure are related to metrics quantifying the quality of spectrum compatibility of the scaled GMs with the target spectrum. A very common metric that is focussing on mean estimates of structural response is the normalized root-mean-square-error δ between the scaled average spectrum of the set of GMs and the target spectrum:

$$\delta = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{Sa_{avg,sc}(T_i) - Sa_{trg}(T_i)}{Sa_{trg}(T_i)}\right)^2} \tag{8}$$

In the above equation, $Sa_{avg,sc}(T_i)$ is the spectral ordinate of the scaled average spectrum at period T_i given by Eq. (9), $Sa_{trg}(T_i)$ is the ordinate of the target spectrum at the same period and N is the number of period sample values used within the pre-defined range of periods, where matching is envisaged (e.g. between $0.2T_1$ and $2T_1$ according to [12]). In Eq. (9), $Sa_j(T_i)$ is the un-scaled ordinate of the individual spectrum *j (j = 1 to n)* at period T_i and Sf_j is the corresponding scale factor.

$$Sa_{avg,sc}(T_i) = \frac{\sum_{j=1}^{n} Sf_j \cdot Sa_j(T_i)}{n} \tag{9}$$

The main contribution of the present study is the explicit consideration of additional objectives in the selection of GMs that are not directly related to spectrum compatibility. These objectives refer to target properties values related to regional seismicity, local site conditions, strong GM intensity and duration as well as the magnitude of scale factors. These properties are linked not only to peak responses, addressed by spectral matching, but also to the energy content and the cumulative demands of earthquake GMs. Hence, they should be taken into consideration in the seismic assessment of structural systems by means of rigorous nonlinear response history analysis. Nevertheless, they are considered only as constraints, in the form of permissible ranges, in the pre-processing phase of the existing GM selection procedures.

Following the proposed approach, the selections with the optimal trade-offs between matching these additional properties target values and spectral matching are obtained. It is noted that targeting these additional properties values in the conventional approach of GM selection can be pursued only manually by gradually modifying their permissible ranges. However, the latter procedure can be tedious and not drive to optimal solutions since specific parameter values are not targeted. Furthermore, limiting significantly the permissible ranges may drive to unfeasible selections or selections with inadequate spectral compatibility.

In all cases, the additional selection objectives can be set to minimize the normalized root-mean-square-error e_Q (Eq. 10) between the individual GM properties values Q_j (j = 1 to n) and the corresponding target property value of the selection procedure Q_{trg}, where Q can be set as M, R, $V_{s,30}$, PGA, $S_a(T_1)$, t_s and any other desirable scalar property.

$$e_Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Q_i - Q_{trg}}{Q_{trg}}\right)^2} \tag{10}$$

In addition to the above, another objective is examined herein in order to reduce the bias in the seismic assessment of structures introduced by extensive spectral scaling [17]. According to this objective, the scale factors of the individual spectra should remain as close as possible to unity (i.e. un-scaled spectra). This objective can be easily expressed in the same context as the one used in Eq. (10) by taking Q_i (i = 1 to n) from Eq. (11) and setting $Q_{trg} = 1$.

$$Q_i = \max\left(Sf_i, \frac{1}{Sf_i}\right) \quad (i = 1\ to\ n) \tag{11}$$

As discussed, previous studies (e.g. [36]) consider as an additional objective to matching central spectral values the control of variability within the GMs suites by matching a constant target dispersion of pseudo spectral acceleration regardless of period. Other selection objectives may also be used to limit the variability of spectral matching. For example, the average or the maximum values of the normalized root-mean-square-errors δ_i between the scaled individual spectra and the target spectrum can be used as selection objectives. The role of the latter additional objectives will also be examined in the following of this study.

Solution algorithm

The optimization problem of Eq. (3) is solved in this study by the use of the mixed-integer GA implemented in MATLAB 2017a [50]. It is important to clarify, however, that any other mixed-integer optimization algorithm can be used in the framework of this study. Genetic Algorithms [51] belong to the class of stochastic, nature-inspired heuristic algorithms. They are based on Darwin's theory of natural selection and evolution. GAs iteratively modify populations (generations) of individuals in order to evolve toward an optimum solution. An individual $\boldsymbol{x}$ (genome) represents a candidate solution to the optimization problem. The values of the design variables x_k (k = 1 to l) forming each individual are called genes. In order to create the next population, GAs select certain individuals in the current population (parents) and use them to create individuals in the next generation (children). The GA algorithm implemented in MATLAB-R2017a treats both continuous and discrete design variables. To serve this goal, special crossover and mutation functions are used to ensure that discrete variables take values only from pre-determined discrete sets of values [52]. Furthermore, the algorithm is able to account for nonlinear constraints by using the penalty function approach [53].

2.3 Post-processing

This last phase of the selection procedure is required only when multi-objective optimum selection is applied without fixing the values of the selection weights in the pre-processing phase. In this case, a number of Pareto-optimal selections is returned by the algorithm and it is up to the engineer to design the most suitable to the goals of the seismic assessment. Elements of engineering judgement and decision theory can be used in the stage. It is generally considered (e.g. [54]) that spectral matching is more important than parameters related to the seismology of the region, local site conditions, strong GM intensity and duration. Therefore, it could be reasonable, depending always on the nature of the problem under investigation, that the Pareto-optimal solutions that provide good spectral matching without yielding high $e_Q$values to be preferred. Furthermore, goal programming concepts can be directly applied in this phase, where solutions are chosen to satisfy a pre-determined goal for one objective and yield the optimum feasible outcomes for the rest of the objectives under consideration. For example, the Pareto-optimal solution with the minimum $e_Q$value could be selected among all solutions with $\delta \leq \delta_{lim}$, where δ_{lim} is a satisfactory limit value for δ (e.g. $\delta_{lim} = 0.05$).
At this point, it is important to highlight the advantages of the proposed Pareto-front of the error metrics approach with respect to a more manual solution, where the user visually inspects the optimal GM sets obtained by the different weight combinations and selects the one that finds more satisfactory. The proposed approach offers quantified information related to the achievement of the specified selection targets. This quantified information facilitates automation and objectivity of the selection procedure. In this manner, the amount of effort required by the user is drastically reduced. Furthermore, it is not required by the user to make subjective judgements regarding the level of achievement of the various selection objectives.

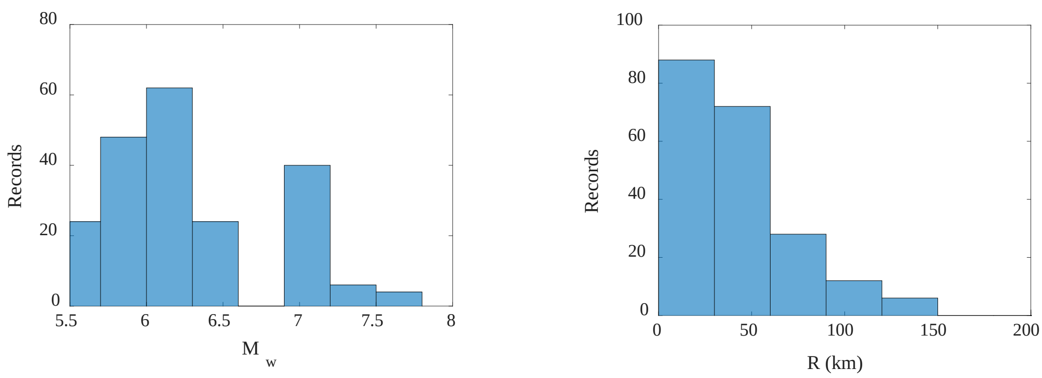
3 Ground motion selection applications

3.1 Introduction

To illustrate the applicability and efficiency of the proposed methodology, GM sets will be selected according to Eurocode 8 – Part 1 [12] for the seismic assessment of a structure of ordinary importance with fundamental period $T_1 = 0.75$s resting on stiff soil profile with $V_{s,30} = 600$ m/s that is classified as ground type B according to the specifications of EC8- Part 1. The structure is assumed to be located in a region of high seismicity that is dominated by an earthquake scenario of moment magnitude $M_w = 6.2$ at an epicentral distance of R = 20 km. The motions will be selected for the 10% probability of exceedance in 50 years seismic hazard level with anticipated PGA = 0.24g.

In the pre-processing phase, a pool of eligible GMs has to be derived from a GM database. In this study, all GMs are taken from the European Strong-motion Database (ESD) [44]. Next, appropriate pre-selection criteria need to be applied to derive GMs compatible with the conditions of the seismic assessment under investigation. Herein, a number a broad range criteria are used to filter the eligible GMs. These are: i) $M_w > 5.5$; ii) $R \geq 10$km; iii) Ground type B; iv) Both horizontal GM components have PGA $\geq$ 0.02g. The first criterion is compatible with the definition of the zones of high seismicity in EC8- Part 1and the last criterion is used to avoid excessive scale factors. The second and third criteria are supposed to be reflecting the soil and seismology conditions of the problem. Applying these filters in the ESD returns 184 GMs with two horizontal components, therefore 368 different GM records. From these motions, only the 104 (208 horizontal records) are recorded on soil profiles with known $V_{s,30}$. The latter records are used herein as the basic pool of eligible GM records because $V_{s,30}$ will be used as selection objective in the following and the corresponding values are required. Fig. 2 presents histograms with the basic properties of the eligible records of the GM database.

As a target spectrum, the Type-1 (high seismicity) spectrum of EC8 – Part 1 is used in this study as prescribed for soil class B, PGA = 0.24g, importance factor $\gamma_I = 1$ (structures of ordinary significance) and 5% viscous damping. It is important to note herein that, in the selection of GMs according to the earthquake scenario examined in this study, it is generally more appropriate to use a target spectrum determined specifically for this scenario by appropriate empirical response spectra attenuation relationships or by PHSA. However, because different GM selections will be examined later in this study for various scenarios and objectives, it was decided to use always the same target spectrum for common reference.



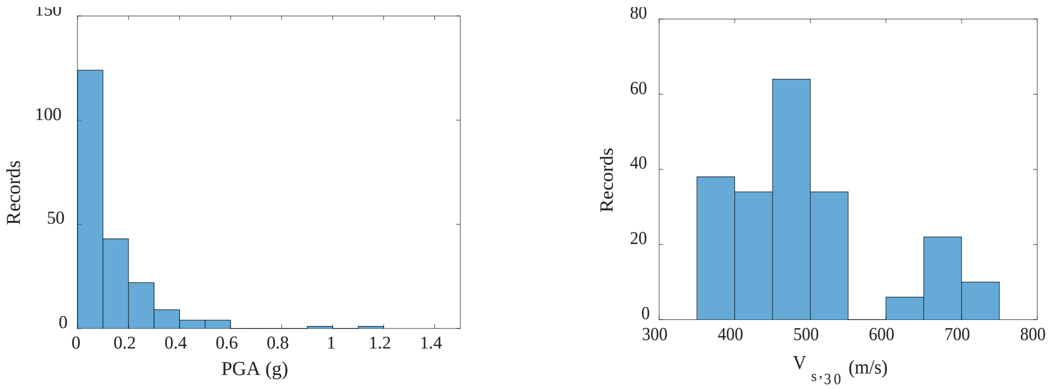


Fig. 2: Histograms of eligible GM basic properties

In the following, the results of the GM selections will be presented for different objectives. First, single-objective selections will be presented to act as a reference and verify the efficiency of the applied numerical procedure. To this end, applications of the GA algorithm with different parameters are examined and compared to provide a further insight, with respect to the existing literature, into the applicability and efficiency of the use of metaheuristic algorithms to the selection of GMs and its implications. Next, a number of different multi-objective optimum selections are examined and critical comments are made with regards to the derived optimal solutions.

3.2 Single-objective selections

In this section, selection of sets of 7 GMs are examined for a single objective, which is to minimize the error δ between the scaled average spectrum of the set of GMs and the target spectrum as given by Eq. (8). EC8-Part 1 [12] selection constraints are applied as discussed in §2.2. In addition, it is set that, after scaling, the zero period acceleration values of the individual response spectra, should be between 0.5 and 2.0 times the corresponding value of the target spectrum to limit excessive variability of the scaled individual spectra. Furthermore, it is specified that both horizontal components of the same GM cannot be part of the same GMs set for uniaxial ground motion excitation. Lastly, it is decided that the individual scale factors cannot be higher than 10 and lower than 0.1.

The optimum selections are performed with the use of the mixed-integer GA solution algorithm discussed in §2.2. In total, 14 selection variables are used. This counts for 7 integer variables for the serial numbers of the GMs and 7 continuous for the scale factors, assumed variable and independent herein. A population size of 100 individuals with 5 elite individuals is used. Each GA run is terminated after 3000 generations. In total, 100 independent runs of the GA are performed. Due to the stochastic nature of the GA algorithm, each run may provide a different selection set and corresponding δ value.

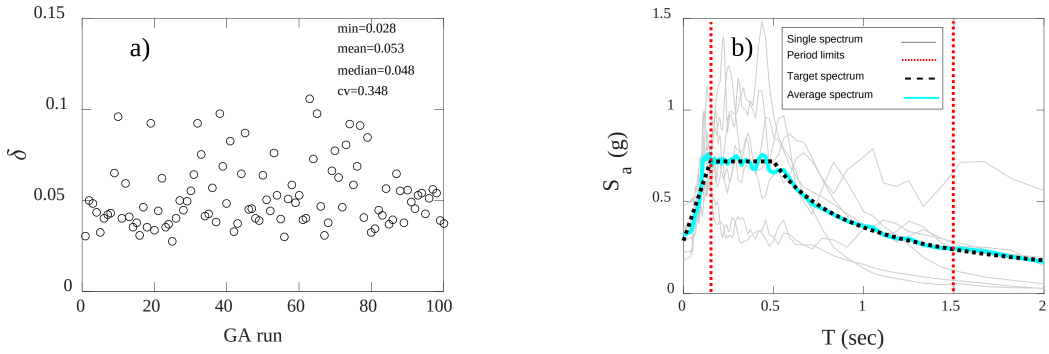


Fig. 3: Single-objective GM selections using the GA algorithms. a) δ values obtained from different GA runs; b) Spectra of the GM set with the minimum δ value

Figure 3a presents the δ values obtained from the 100 independent GA runs. A considerable variation of these values is observed with a coefficient of variation of 0.35. The median value is 0.053 and the minimum 0.028. The latter value was obtained at the 25th run of the algorithm. Therefore, in general, a sufficient number of independent GA runs is required to get the best solution. However, in this example, it may be observed that even from the 1st run a δ value very close to the minimum is obtained that could be used to select an alternative set of GMs. Furthermore, Fig. 3b shows the spectra of the set of GMs with the minimum δ value and how they compare with the target spectrum. It is evident that excellent matching, inside the specified period limits, of the average and the target spectrum is achieved. High variability of the individual spectra is observed. This is the case because EC8 does not control the level of variability inside the GM sets. The issue of the variability of GM spectra is addressed separately in a following section of this study.

The selection of the parameters of the GA algorithm is typically based on experience from previous applications to similar optimization problems. However, as discussed, this experience seems to be lacking in the case of the selection of GMs. To fill part of this gap, a parametric study is conducted in the following to illustrate the influence of these parameters on the computational performance of the algorithm.

Fig. 4a presents the variation of δ values, in the form of box plots, as obtained from 100 independent runs of the same GA algorithm with population sizes of 50, 100 and 150 individuals. The box plots show the minimum, maximum and median (red line) values obtained. Inside the boxes, the 25th to 75th percentile solutions are contained. It can be seen that the 100 individuals population size provides the best solutions in terms of both minimum and median values. This effectively shows also that increasing the population size does not necessarily drive to better solutions. This observation is also consistent with the recommendation for the population size of this algorithm provided in [50] according to which the population size should be taken from Eq. (12), where *nvars* is the number of selection variables (i.e. *nvars* = 14). It is worth noting, however, that the differences obtained from the different population sizes are generally not very high.

$$population\ size = \min(\max(10 \cdot nvars, 40), 100) \tag{12}$$

Furthermore, Fig. 4b presents the variations of the δ values from 100 independent GA runs as a function of the number of elite individuals selected in a population of 100 individuals. It can be seen that the 5 and 10 elite candidate solutions provide similar solutions. However, the use of only 2 elite individuals drives to significant increase of both the minimum and median δ values. Again, these observations are consistent with [50], where the number of elite individuals is recommended to be 5% of the population size.

Moreover, Fig. 4c illustrates the δ values obtained from 100 GA runs with population size of 100 individuals and 5 elite individuals as a function of the number of maximum generations. As expected, the median and the scatter of the results decrease as the number of iterations increases. It is interesting to note, however, that after the 3000 maximum generations limit the improvement of the results is only marginal and therefore the further increase of iterations may be deemed as unnecessary as it increases considerably the computational cost. It is interesting to note here that the default maximum generations in [50] are 100·*nvars* (i.e. 1400), which seems that could introduce a considerable scatter in the optimum solutions.

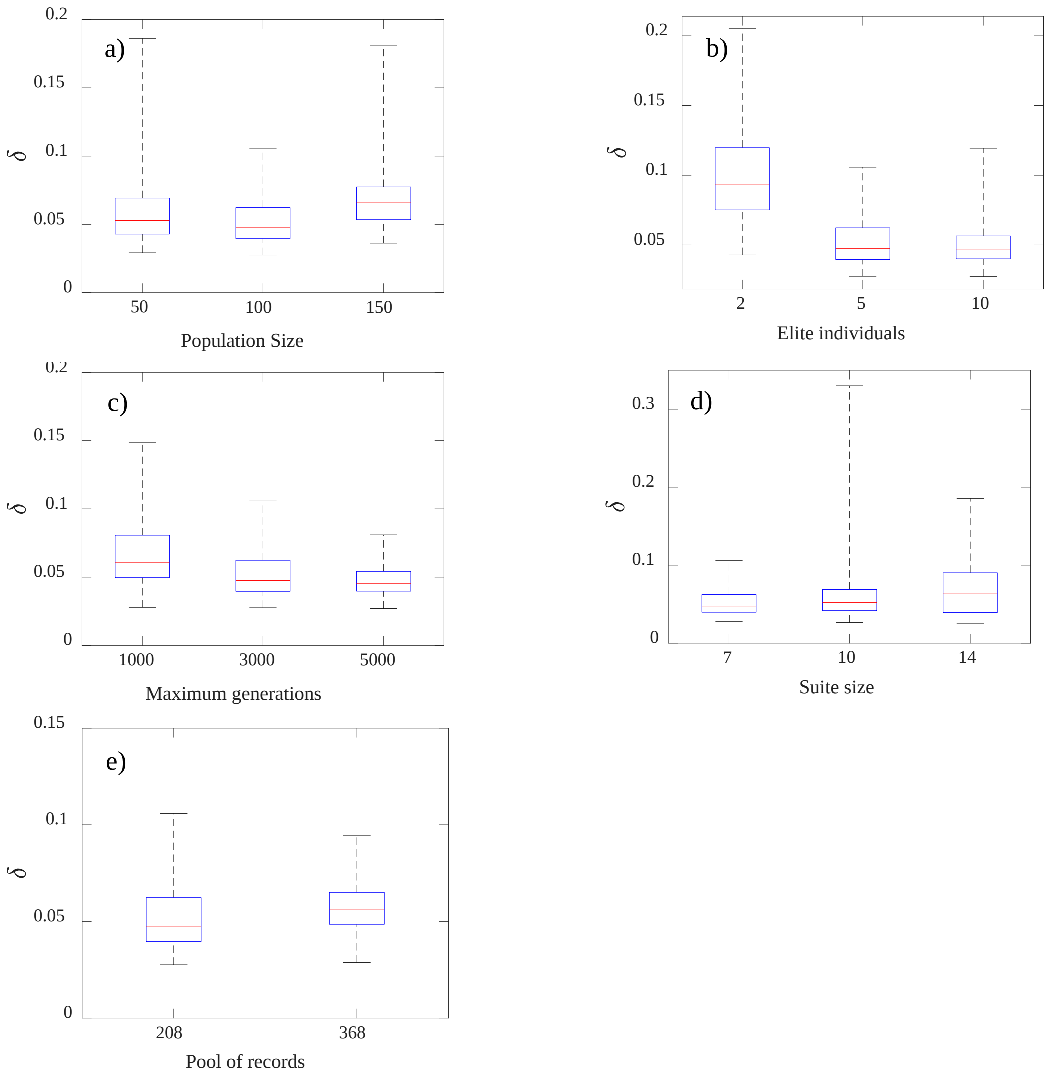


Fig. 4: Variation of δ values obtained from single-objective GM selections as a function of the: a) population size; b) number of elite individuals; c) maximum generations; d) suite size; e) pool of records.

In addition, Fig. 4d presents the δ values obtained from 100 GA runs when sets of 7, 10 and 14 GMs are pursued. This means that 14, 20 and 28 selection variables (GMs serial numbers and scale factors) are used respectively in the solution algorithm. In all cases, a population size of 100 individuals with 5 elite individuals is used. The analysis is terminated after 3000 generations. It can be observed that in all cases a similar minimum δ value is obtained. This is positive because it means that the GA algorithm manages to find high quality solutions for larger sets of GMs despite the corresponding sharp increase of the search space. However, the scatter of the optimum solutions generally increases as the set size increases.

Finally, Fig. 4e compares the obtained δ values as derived for selected sets of 7 GMs and using the same GA parameters as in Fig. 4d from two different pools of eligible GM records. The first pool of 208 records is the same as the one used in the previous examples herein and the second is the pool of 368 records that includes also records that have been recorded on soil profiles with unspecified $V_{s,30}$ as explained in §3.1. It can be seen that the same minimum value is obtained in both cases. However, the median δ value slightly increases as the pool of eligible records increases due to the corresponding increase of the available search space. It is noted that the increase of the search space can lead to an increase of the number of generations required for the GA to converge to the optimum solution and this may increase considerably the computational cost for very large pools of records. On the other hand, more candidate solutions are available and the GA may reach a solution of similar quality to a smaller database prior to reaching convergence. Hence, the direct comparison of the computational costs required for reaching the same value of the objective function from a small and a large database is not straightforward. In any case, it is reminded herein that the proposed approach can duly work in parallel with the application of stricter selection criteria in the pre-processing phase that can reduce the size of the pools of records when GA is found to be less efficient.

It is important to highlight at this point that the choice of the GA search scheme depends strongly on the goals and the computational budget of the GM selection procedure. For example, if it is important that the global optimum selections are obtained then a high number of maximum generations and independent GA runs should be used. However, for more practical applications where the users simply need to get solutions of good quality and the computational budget is limited, the number of generations and independent GA runs can be drastically reduced with rather minor effects on the quality of the results as shown in Figs 3 and 4.

3.3 Multiple-objective selections

In this section, selections of sets of 7 GMs for more than one objectives are examined. Again, the optimum selections are performed with the use of the mixed-integer GA solution algorithm with 14 independent selection variables. A population size of 100 individuals with 5 elite individuals is used. Each analysis is terminated after 3000 generations. For each combination of weights of the different objective functions, 100 independent GA runs are conducted and the solution yielding the minimum $F(\boldsymbol{x})$ function is selected to represent the corresponding Pareto-optimal point. This approach is mainly followed herein in order to increase the likelihood of obtaining the global optimal solutions for each combination of weights and thereby better demonstrate the effect of the weight values on the different optimal solutions. Alternatively, a more versatile approach for the derivation of the Pareto-fronts can be used, where the optimal solutions of all the GA runs for all the combinations of weights are examined and the Pareto-optimal points are then identified based on their definition (i.e. the points that are not dominated by other points in the objective functions space). Due to the stochastic nature of the GA algorithm, the latter approach can discover Pareto-optimal points that may be missed by the former procedure. This also means that the latter approach could potentially be applied with a lesser number of GA runs, with respect to the former, leading to a reduction of the computational cost. In the following, selections of GM sets for multiple objectives are presented. First, solutions for two and then for more than two selection objectives are examined.

3.3.1 Selections for two objectives

Herein, bi-objective optimum selections are examined by applying the WSM, where $\bar{F}(\boldsymbol{x})$ of Eq. (5) has to be minimized. In all cases, δ is used as the first selection objective function $F_1(\boldsymbol{x})$ to represent the quality of spectral matching. In addition, a great variety of second objectives $F_2(\boldsymbol{x})$ is applied to derive the corresponding Pareto-fronts. Eq. (5) is applied for 5 different values of w_1 = 0, 0.25, 0.50, 0.75 and 1.00, whereas w_2 is determined as $w_2 = 1\text{-}w_1$. In this manner, 5 Pareto-front optimal points are derived that can be also used to approximate the entire Pareto-front, if necessary. In the following, these optimal points are numbered according to their w_1 values in ascending order. In this way, the 1st point (w_1 = 0) corresponds to the single-objective optimization for minimum $F_2(\boldsymbol{x})$ and the 5th point to the solution for minimum $F_1(\boldsymbol{x})$. Clearly, as w_1 increases more emphasis is placed on δ with respect to the alternative selection objectives. In the following, the results obtained for the different $F_2(\boldsymbol{x})$ are presented.

Seismological parameters M_w and R

It has been established that the magnitude M_w exerts a considerable influence on structural response quantities both in terms of amplitude and duration [32]. Therefore, it is generally treated as an important parameter in the selection process. The aim here is to present the Pareto-front optimal solutions in terms of δ versus e_M, where e_M is the error of the magnitudes of the selected individual GMs $M_{w,i}$ with respect to the target value $M_{w,trg}$ = 6.2 specified by the earthquake scenario under examination.

Fig. 5a illustrates the derived Pareto-optimal points. In this figure, as well as in all similar figures presented later, points are numbered (1 to 5) according to their w_1 values as discussed. As expected, point 5 exhibits the best solution in terms of spectral matching but with a high error in magnitudes and point 1 is the most representative in terms of earthquake magnitudes but with rather poor spectral matching ($\delta \approx 0.19$). In between, points 2, 3 and 4 all constitute rather attractive solutions. For example, point 4 imperceptibly increases the minimum δ value of point 5 while reduces e_M by 70%. Furthermore, points 2 and 3 also represent acceptable solutions in terms of spectral matching ($\delta \approx 0.05$) with small errors in magnitudes. At this point, it becomes more evident the main advantage of the formation of Pareto-front solutions in the selection of GM sets. It allows for informed selection-making based on the trade-offs between the various objectives.

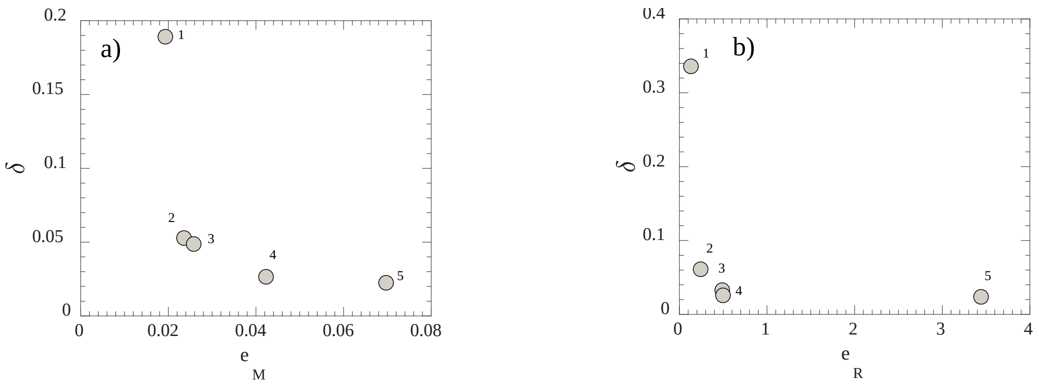


Fig. 5: Pareto-front optimal solutions: a) δ versus e_M; b) δ versus e_R

Fig. 6 presents the spectral matching attained by the GM sets of points 1 and 2. It is shown that point 2 exhibits much better matching of the target and mean spectrum, with almost the same e_M, when compared to point 1. Therefore, it would generally be preferable to point 1. In addition, the quality of spectral matching of point 2 is not significantly worse than the single-objective selection for minimum δ of Fig. 3b. Therefore, it represents an attractive solution since it combines good spectral matching with earthquake magnitudes representative of regional seismicity.

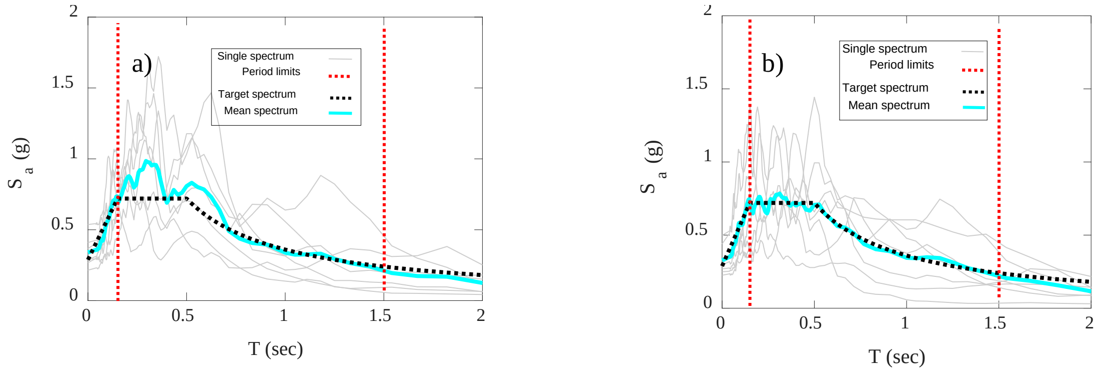


Fig. 6: Spectral matching of different δ versus e_M Pareto-optimal solutions: a) point 1; b) point 2

It is generally accepted [32] that structural responses are not significantly correlated with the source-to-site distance *R*. Nevertheless, *R* is still considered as a secondary criterion in the selection procedure. Fig. 5b presents the δ versus e_R, Pareto-front optimal solutions, where e_R is the error in distances for the target distance R_{trg} = 20 km. This Pareto-front could be used, for example, in the case of a known fault in the seismic region with variable potential earthquake magnitude. It can be seen in Fig. 5b, that points 3 and 4 greatly reduce the error in distances whereas they only marginally influence spectral matching. Therefore, even though *R* is not strongly related to structural response, points 3 and 4 could be used instead of point 1 since they are more representative of the regional seismicity and they provide equivalent spectral matching.

Soil profile parameter $V_{s,30}$

It is known that the geotechnical profile affects both the amplitude and duration of strong GMs as well as their computed response spectra [32]. Typically, the soil profile is considered in the selection procedure by site classification. In the problem investigated herein, the soil is classified according to EC8 – Part 1 as Type B with a rather broad range of $V_{s,30}$ values between 360 m/s to 800 m/s. However, if the actual $V_{s,30}$ value at the site of the structure is known (assumed 600 m/s herein), then it is worthwhile to investigate δ versus $e_{Vs,30}$ Pareto-optimal solutions by setting the known shear wave velocity as the target value. Fig. 7a presents the obtained optimal solutions. It is evident that, apart from point 5, points 3 and 4 are good quality solutions since they contain motions recorded in more similar profiles to the site of interest and they only slightly reduce spectral matching with respect to point 5. Fig. 7b presents the satisfactory spectral matching attained by the GM set of point 4 Pareto-optimal solution.

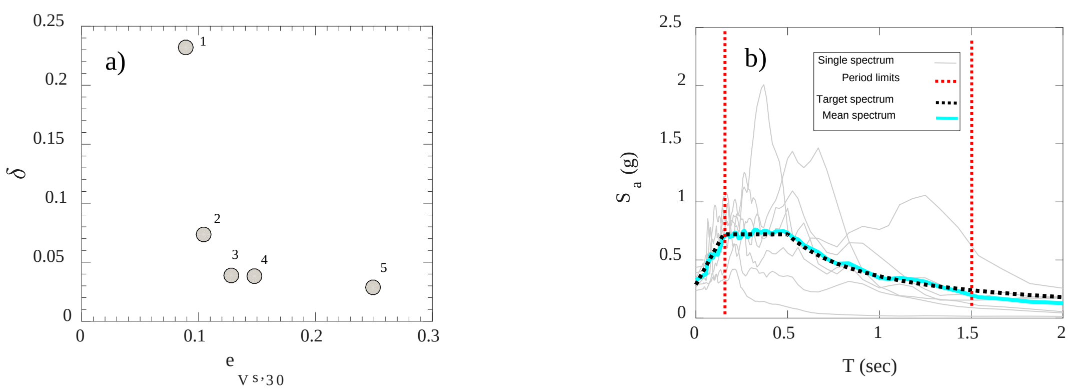


Fig. 7: a) δ versus $e_{Vs,30}$ Pareto-optimal solutions; b) spectral matching of point 4 Pareto-optimal solution

Strong motion intensity and duration parameters

Strong motion intensity parameters (widely known as intensity measures IM) are strongly related to structural response quantities. Therefore, selection of GMs based on IM is an efficient way to increase reliability of the seismic assessment procedure. The most typical examples of IM are the PGA and $S_a(T_1)$. In this section, for reasons of brevity, only PGA is addressed although it is recognised that $S_a(T_1)$ is more closely related to structural response. To include PGA in the selection procedure, one of the two objectives can be to minimize the error e_{PGA} of the individual GMs with respect to the target PGA = 0.24g. Fig. 8a shows the obtained δ versus e_{PGA} Pareto-optimal solutions. It is clear in this figure that the solution for minimum e_{PGA} yields very poor spectral matching. On the other hand, points 4 and 5 have similar and satisfactory δ values and could both be considered in the final selection.

Strong motion duration may influence significantly structural response especially of structural systems that exhibit rather poor hysteretic response such as timber and masonry structures [37] as well as non-ductile reinforced concrete structures [38]. A variety of metrics exist in literature to quantify strong motion duration [55]. Perhaps, the simplest metric is the so called bracketed duration defined as the total time elapsed between the first and last excursion of a specified level of acceleration α_o. In this section, the duration $t_{s,0.02}$ is used, which is based on α_o = 0.02 g. The anticipated $t_{s,0.02}$ is directly estimated for the earthquake scenario under investigation (M_w = 6.2 and R = 20 km) by using attenuation relationships such as the one proposed by Koutrakis *et al*. [56] for the case of Greece. Using this relationship, it is found that the expected $t_{s,0.02}$ is approximately 9.5 s. To consider the anticipated duration in the selection of GMs, it is set as one of the two selection objectives to minimize the error e_{ts} of the 0.02 g bracketed duration of the individual GMs with respect to the target duration $t_{s,trg}$ = 9.5 s. Fig. 8b presents the δ versus e_{ts} Pareto-optimal solutions. It can be seen that the point 1 leads to high δ values ($\delta \approx 0.29$) whereas point 5 drives to relatively high error in duration. Therefore, points 2 - 4 could be also considered to be used for the purposes of seismic assessment.

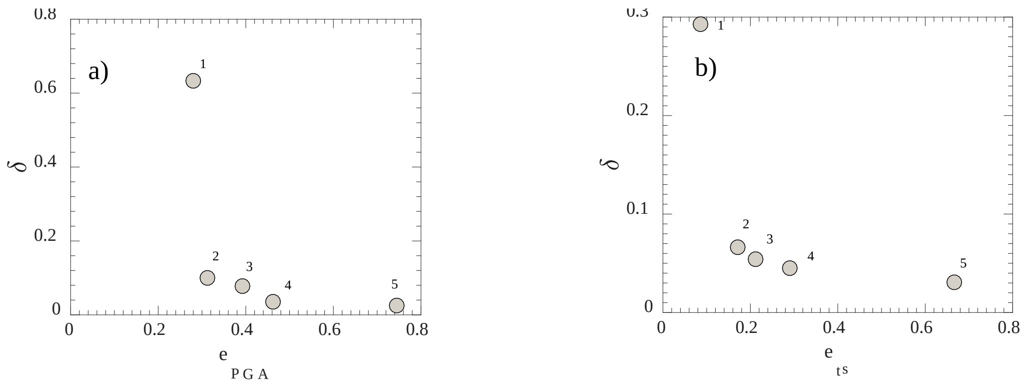


Fig. 8: Pareto-front optimal solutions: a) δ versus e_{PGA}; b) δ versus e_{ts}

Scale factors

It has been shown (e.g. [17]) that the use of excessive scale factors in the selection of GMs may introduce a bias to structural responses. On the other hand, the use of only un-scaled GMs may drive to poor spectral matching and in some cases not feasible selection of GMs sets [18]. To avoid these unfortunate cases, a bi-objective selection approach is proposed herein, where one of the objectives is set that the scale factors of the individual GMs are close to unity (i.e. un-scaled motions). As discussed in §2.2, this is easily achieved in the proposed framework of this study by minimizing the errors of the scale factors e_{Sf} with respect to unity. Fig. 9a presents the obtained δ versus e_{Sf} Pareto-optimal solutions. It is interesting to note in this figure that point 1 has zero e_{Sf}. This is expected since it simply means that the algorithm found an un-scaled set of GMs. However, this point is characterized by rather poor spectral matching ($\delta \approx 0.15$). Therefore, it will be more appropriate to use one of the other Pareto-points that have significantly smaller and generally satisfactory δ values ($\delta < 0.05$). It is also noteworthy that e_{Sf} can be significantly reduced from 5.8 (point 5) to 1.9 (point 2), leading to a mean scale factor from 6.2 to 2.9, for very small increase of δ. Fig. 9b presents spectral matching of the set of GMs of point 2. It is evident that matching of the mean and target spectrum is satisfactory. It is worth noting, however, the high variability of the GM spectra of the selected set of GMs. The issue of the variability of GM spectra is addressed in the following section.

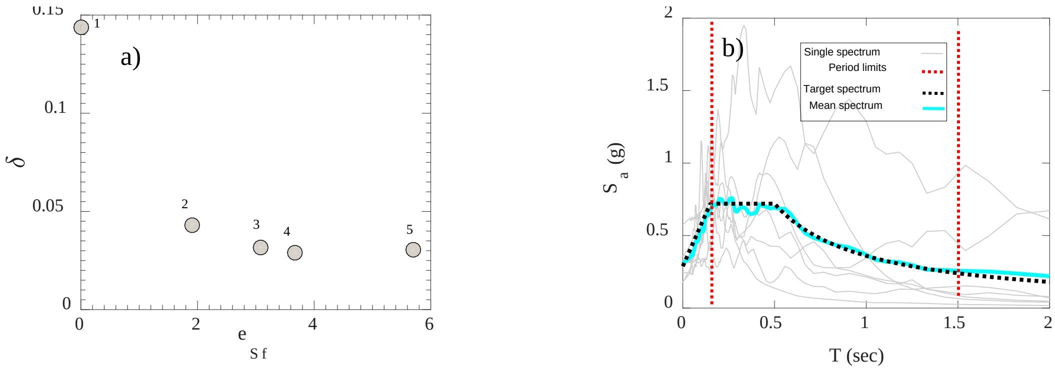


Fig. 9: a) δ versus e_{Sf} Pareto-optimal solutions; b) spectral matching of point 2 Pareto-optimal solution

Spectral matching variability parameters

As discussed, in addition to δ that represents spectral matching in terms of central (mean) values, the reduction of the average and maximum δ_i values of the individual spectra could be considered as well to limit the variability of the sets of GMs. To serve this goal, Pareto-optimal solutions are determined in terms of both δ versus mean(δ_i) and δ versus max(δ_i) and presented in Fig. 10. On purpose, the same scale is used on both axes in these figures to demonstrate the extent of which mean(δ_i) and especially max(δ_i) can be reduced with only slight increase of δ. Fig. 11, shows spectral matching attained by minimizing mean(δ_i) and max(δ_i) (points 1 in Fig. 10) that can be compared with spectral matching obtained when minimizing δ as presented in Fig. 3b. It is evident that the former solutions are characterised by significantly lesser variability (especially the mean(δ_i) solution) and at the same time very good matching of the average and target spectrum. Therefore, they could be considered in the final selection of the post-processing phase. Of course, this occurs only in the case study under examination and cannot be further generalised. In the general case, it will be more rational to use an intermediate Pareto-front point rather than the single-objective solutions. It is worth noting at this point that a target variability value in terms of mean(δ_i) and max(δ_i) can also be set herein and the corresponding point of the Pareto-front be used in the selection procedure in order to control appropriately the variability of GM spectra.

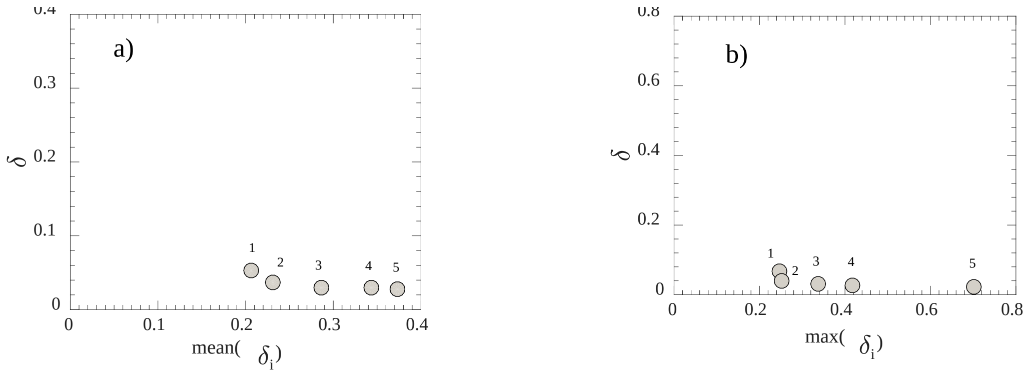


Fig. 10: Pareto-front optimal solutions: a) δ versus mean(δ_i); b) δ versus max(δ_i)

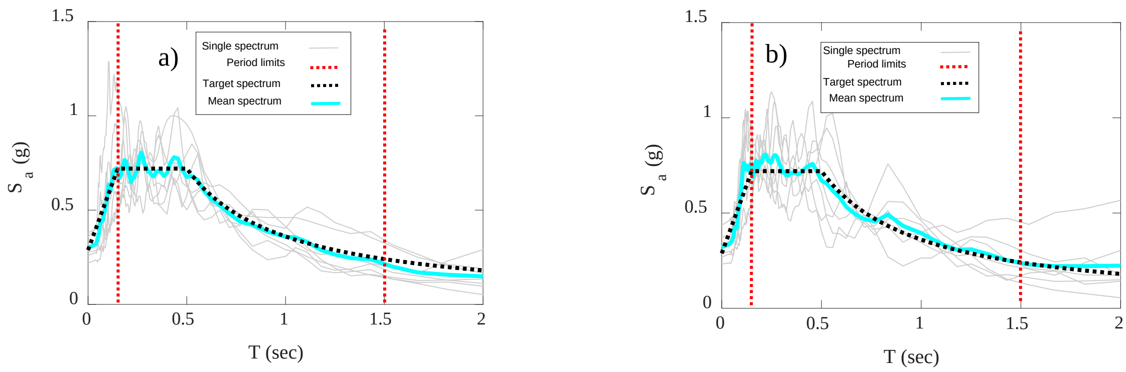


Fig. 11: Spectral matching of optimal GM sets for a) mean(δ_i); b) max(δ_i)

3.3.2 Selections for more than two objectives

In this section, selection of GMs for more than two objectives is investigated. Initially, selection for two specific earthquake scenarios is examined. Next, a full 3-D Pareto front is developed. To the best of the authors' knowledge, this is the first study developing a 3-D Pareto front for the purposes of GM selection. In the first earthquake scenario (δ, M_w, R), selection takes place for adequate spectral matching and to be consistent with an earthquake magnitude M_w = 6.2 and epicentral distance R = 20 km. This selection can be facilitated by using a normalized version of Eq. (3), similar to Eq. (4), and setting: $F_1(\boldsymbol{x}) = \delta$, $F_2(\boldsymbol{x}) = e_M$ and $F_3(\boldsymbol{x}) = e_R$. For simplicity, it is assumed herein that the weights of the selection objectives are specified in the pre-processing phase of the selection procedure and that they are equal to w_1 = 0.50, w_2 = 0.25 and w_3 = 0.25. These weight values are used herein in order to achieve, for illustration purposes, equal and maximum reduction of e_M and e_R while keeping spectral compatibility as the main selection objective. The following objective function values are established for these weights: δ = 0.045, e_M = 0.032 and e_R = 0.627. If these are compared with the respective values of the best single-objective solution shown in Fig. 3b (δ = 0.028, e_M = 0.062 and e_R = 1.178), it is concluded that the multi-objective selection reduces significantly the errors in magnitude and distance while the quality of spectral matching remains still satisfactory. This is clear also in Fig. 12a that presents the (M_w, R) properties of the two selected GM sets. It is evident that the multi-objective selection approaches closer the target point with (M_w, R) = (6.2, 20 km).

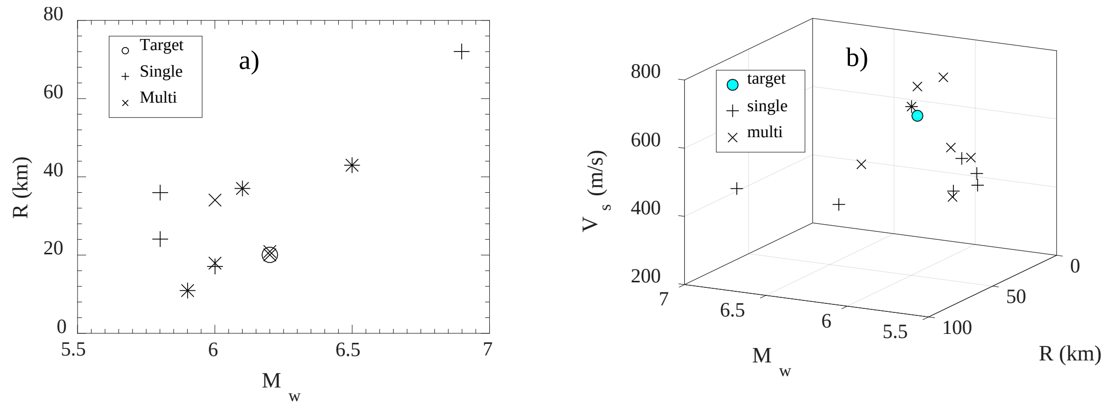


Fig. 12: Properties of GMs of the selected GM sets for: a) (δ, M_w, R); b) (δ, M_w, R, $V_{s,30}$) multi-objective scenarios

The second multi-objective scenario includes the first scenario described in the previous of this section but also sets a target value for the $V_{s,30}$ of the soil profile at the site of the structure ($V_{s,30}$ = 600 m/s). Therefore, it is a (δ, M_w, R, $V_{s,30}$) multi-objective scenario. This scenario is pursued by simply adding a fourth objective in Eq. (4), which is the error $e_{Vs,30}$. It is decided in the pre-processing phase that the weights of the objectives considered in the selection process are: w_1 = 0.5, w_2 = 0.5/3, w_3 = 0.5/3 and w_4 = 0.5/3. The following objective function values are established for these weights: δ = 0.039, e_M = 0.038, e_R = 0.37 and $e_{Vs,30}$ = 0.23. Compared with the respective values of the best single-objective solution shown in Fig. 3b (δ = 0.028, e_M = 0.062, e_R = 1.178 and $e_{Vs,30}$ = 0.28), it is concluded that the multi-objective selection reduces significantly the errors in magnitude and distance and considerably in the soil velocity while the quality of spectral matching remains still satisfactory. This is illustrated also in Fig. 12b that presents the (M_w, R, $V_{s,30}$) properties of the selected single- and multi-objective GM sets. It is evident that the multi-objective selection approaches closer the target point with (M_w, R, $V_{s,30}$) = (6.2, 20 km, 600 m/s).

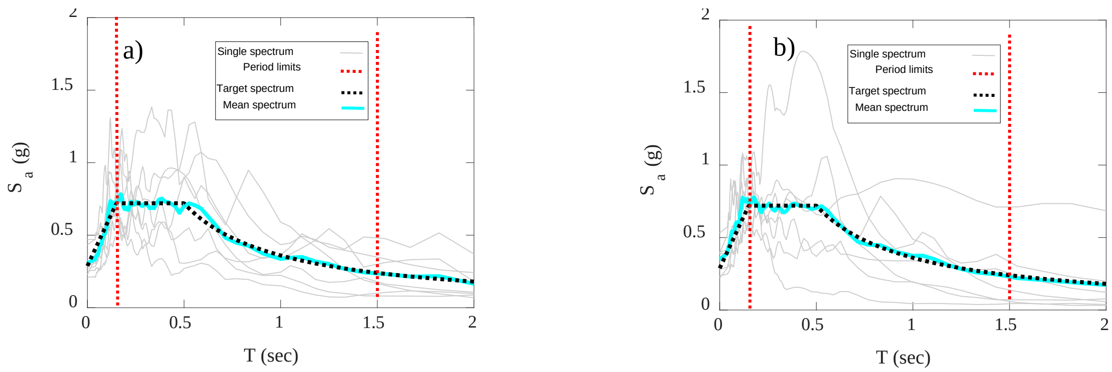


Fig. 13: Spectral matching of optimal GM sets for a) (δ, M_w, R); b) (δ, M_w, R, $V_{s,30}$) multi-objective scenarios

Furthermore, in the following, a 3-D Pareto front is developed. Spectral matching, earthquake magnitudes and spectral matching variability are considered as selection objectives by setting $F_1(\boldsymbol{x}) = \delta$, $F_2(\boldsymbol{x}) = e_M$ and $F_3(\boldsymbol{x}) = mean(\delta_i)$. In total, 15 weight values combinations are used with w_i = 0, 0.25, 0.5, 0.75 and 1.0 (i = 1, 2, 3) while satisfying $w_1 + w_2 + w_3$ = 1.0. To limit the computational cost, 30 GA runs (instead of 100) are conducted for each combination of weights and the Pareto-fronts are derived using the second, more versatile, approach described in the introduction of §3.3 (i.e. examining as candidate Pareto-optimal points the optimal points of all GA runs of all weights combinations). Fig. 14a shows the obtained Pareto-front in the form of a bubble chart, where the Pareto-optimal points are presented in the $mean(\delta_i) - e_M$ plane with circles the width of which is proportional to their δ values. Using the more versatile approach, 25 Pareto optimal points were identified. In this figure, the point with the minimum δ value ($\delta \approx 0.029$) is shown, which is almost equal to the single-objective value. It is observed, however, that this optimal point is characterised by rather high e_M and mean(δ_i) values. To minimize the variability of the GM records, points 1 and 2 can also be considered that both have minimum mean(δ_i) values and adequately small δ values ($\delta \approx 0.04$, mean(δ_i) ≈ 0.20). However, point 1 has significantly smaller e_M value than point 2. Therefore, it better represents regional seismicity and it is preferable than point 2. Fig. 14b shows spectral matching achieved by point 1. It is evident that both the matching of the mean spectrum as well as the variability of GM records are satisfactory.

The afore-described selection procedure demonstrates the benefits of considering explicitly more objectives in the GM selection procedure. It is important to clarify, however, that the use of a significant number of selection objectives increases considerably the computational cost required for the development of the Pareto-fronts as well as the effort required by the user to post-process the results. In these cases, it may be preferable to consider the objectives of secondary importance in the form of pre-selection criteria close to the corresponding target values in the pre-processing phase of the selection procedure.

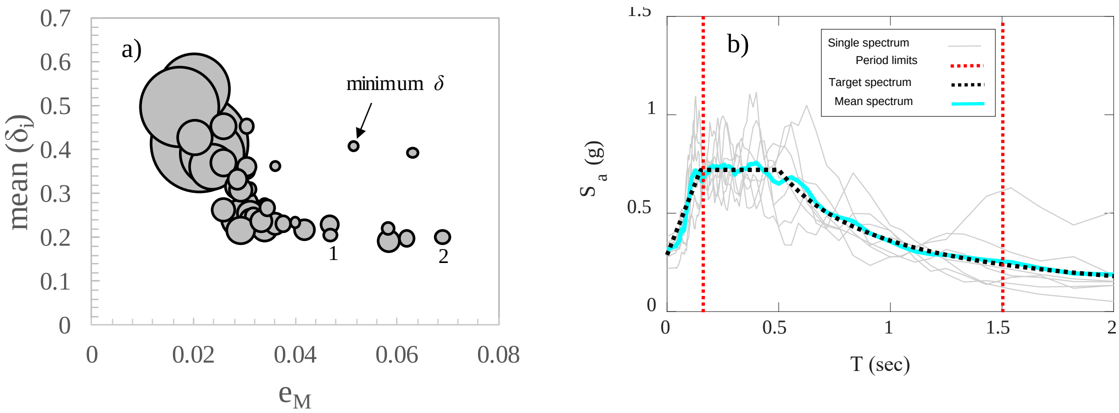


Fig. 14: a) 3D Pareto-front in the form of bubble-chart; b) Spectral matching of point 1 Pareto-optimal solution

4 Summary and Conclusions

Selection of earthquake GMs represents the most significant source of uncertainty in seismic assessment of structural systems. Therefore, appropriate selection of GMs is of vital importance. Existing methodologies for the selection of earthquake GMs focus on spectral matching in terms of either only central values or both central values and variability. In this study, a new versatile multi-objective framework for the optimum selection of earthquake GMs is presented that is fully compatible with major seismic design codes and sets, for first time, as selection objectives not only optimum matching with a target spectrum but also compliance with target values of parameters representing the seismology of the region, soil conditions, strong ground motion intensity and duration, the magnitude of scale factors and others. The additional objectives are reflecting not only peak demands but also cumulative seismic effects and the variability level of the GM selection.

The framework is based on the principles of multi-objective optimization and it is applied with the aid of the WSM that supports decision making both in the pre-processing and post-processing phase of the selection of GM sets. This is the first study employing the WSM to derive not a single GM set but a family of Pareto-optimal GM sets matching more than one selection objectives. It is also the first study deriving GM selection Pareto-fronts for more than two selection objectives. Following this approach, the proposed selection procedure provides a number of GM sets with the best trade-offs between spectral matching and compliance with the target parameters values in a computationally efficient manner. In a post-processing phase, the user is able to select the most suitable solution based on engineering judgement and/or decision-making theory.

The solution of the optimization problem is performed by an efficient GA that is able to track near global optimum solutions of constrained problems with both discrete and continuous selection variables. A parametric application of this algorithm is also conducted for first time herein to demonstrate the effects of its parameters on the selection of GMs. Useful conclusions are made regarding its efficient use in GM selection procedures.

The proposed framework is applied to the selection of GM sets for the seismic assessment of a structural system exposed to a specific earthquake scenario. It is shown that the developed framework is able to track sets of GMs that not only provide excellent spectral matching but they are also more representative of the earthquake scenario under investigation in terms of moment magnitude, epicentral distance, local soil conditions, intensity and duration of strong GMs. Similarly, it is found that, if set as selection objectives, the proposed framework is capable of limiting the variability and the magnitude of the scale factors of the selected GM sets while again maintaining excellent spectral matching in terms of central values. The previous are typically achieved by using higher weight values for spectral matching than the additional selection objectives. More particularly, by comparing the points of the developed Pareto-fronts of §3.3, it is usually found that weight values of up to 0.25 (points 4) and many times higher for the additional objectives improve significantly the compliance with their target values while they imperceptibly affect spectral matching with respect to the single objective solutions (points 5). However, further research is required to recommend weight values for ground motion selection in the general case and therefore it is generally recommended that the full Pareto-front of the optimal solutions is derived and the most suitable GM set to the seismic assessment investigated is selected. The previous observations drive to the conclusion that the developed framework represents a valuable tool for the efficient selection of earthquake GMs for the purposes of seismic assessment of structural systems.

References

[1] *fib*. Model Code 2010. Lausanne: Federation Internationale du Beton; 2012.

[2] Mergos PE. Optimum seismic design of reinforced concrete frames according to Eurocode 8 and *fib* Model Code 2010. Earthq Eng Struct Dyn 2017; 46(7): 1181-201.

[3] Padgett J, Desroches R. Sensitivity of seismic response and fragility to parameter uncertainty. J Struct Eng 2007; 133(12): 1710-8.

[4] Sextos AG, Katsanos EI, Manolis GD. EC8-based earthquake record selection procedure evaluation: validation study based on observed damage of an irregular R/C building. Soil Dyn Earthq Eng 2011; 31(4): 583-97.

[5] Bommer JJ, Acevedo AB. The use of real earthquake accelerograms as input to dynamic analysis. J Earthq Eng 2004; 8 (Special Issue 1): 43-91.

[6] Naeim F, Lew M. On the use of design spectrum compatible time histories. Earthq Spectra 1995; 11(1): 111-27.

[7] Sextos AG, Pitilakis KD, Kappos AJ. Inelastic dynamic analysis of RC bridges accounting for spatial variability of ground motin, site effects and soil-structure interaction phenomena. Part 1: Methodology and analytical tools. Earthq Eng Struct Dyn 2003; 32: 607–27.

[8] Hancock J, Watson-Lamprey J, Abrahamson NA, Bommer JJ, Markatis A, McCoy E, Mendis E. An improved method of matching response spectra of recorded earthquake GM using wavelets. J Earthq Eng 2006; 10 (Special Issue 1): 67-89.

[9] Giaralis A, Spanos PD. Wavelet-based response spectrum compatible synthesis of accelerograms-Eurocode application (EC8). Soil Dyn Earthq Eng 2009; 29: 219-35.

[10] Katsanos EI, Sextos AG. Structure-specific selection of earthquake GMs for the reliable design and assessment of structures. B Earthq Eng 2018; 16(2): 583-611.
[11] ASCE. Minimum design loads and associated criteria for buildings and other structures. Reston, Virginia: ASCE/SEI Standard 7-16; 2016.
[12] CEN. Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings. Brussels: European Standard EN 1998-1; 2004.
[13] Stewart JP, Chiou SJ, Bray JD, Graves RW, Somerville PG, Abrahamson NA. Ground motion evaluation procedures for performance-based design. Soil Dyn Earthq Eng 2002; 22(9): 765–72.
[14] Shome N, Cornell CA, Bazzurro P, Carballo JE. Earthquakes, records, and nonlinear responses. Earthq Spectra 1998; 14(3): 469–500.
[15] Haselton CB, Baker JW, Bozorgnia Y, Goulet CA, Kalkan E, Luco N, Shantz T, Shome N, Stewart JP, Tothong P, Watson-Lamprey J, Zareian F. Evaluation of ground motion Selection and Modification Methods: Predicting Median Interstory Drift Response of Buildings. Berkeley, California: PEER Technical Report 2009/01; 2009.
[16] Baker JW. Ground motion selection for performance-based engineering, and the Conditional Mean Spectrum as a selection tool. Proc. 10th Pacific Conf. Earthq. Eng. Build. an Earthquake-Resilient Pacific; 2015.
[17] Luco N, Bazzurro P. Does amplitude scaling of GM records result in biased nonlinear structural drift responses? Earthq Eng Struct Dyn 2007; 36: 1813-35.
[18] Iervolino I, Galasso C, Cosenza E. REXEL: Computer aided record selection for code-based seismic structural analysis. B Earthq Eng 2010; 8(2): 339-62.
[19] Beyer K, Bommer JJ. Selection and scaling of real accelerograms for bi-directional loading: a review of current practice and code provisions. J Earthq Eng 2007; 11(1): 13-45.
[20] Smerzini C, Galasso C, Iervolino I, Paoluccia R. Ground motion record selection based on broadband spectral compatibility. Earthq Spectra 2014; 30(4): 1427-48.
[21] Katsanos EI, Sextos AG. An integrated software environment for structure-specific earthquake ground motion selection. Adv Eng Softw 58: 70-85.
[22] Cimellaro GP, Marasco S. A computer-based environment for processing and selection of seismic GM records: OPENSIGNAL. Frontiers in Built Environment 2015; 1: 1:13.
[23] Baker JW. Conditional mean spectrum: tool for ground-motion selection. J Struct Eng 2011; 137: 322-31.
[24] Kottke A, Rathje E. A semi-automated procedure for selecting and scaling recorded earthquake motions for dynamic analysis. Earthq Spectra 2008; 24(4): 911-32.
[25] Bradley BA. A generalized conditional intensity measure approach and holistic ground-motion selection. Earthq Eng Struct Dyn 2010; 39(12): 1321–42.
[26] Jayaram N, Lin T, Baker JW. A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. Earthq Spectra 2011; 27(3): 797–815.
[27] Lin T, Harmsen SC, Baker JW, Luco N. Conditional spectrum computation incorporating multiple causal earth- quakes and ground-motion prediction models. B Seismol Soc Am 2013; 103(2A): 1103–16.
[28] Kwong NS, Chopra AK, McGuire RK. A ground motion selection procedure for enforcing hazard consistency and estimating seismic demand hazard curves. Earthq Eng Struct Dyn 2015; 44: 2467–87.
[29] Ha SH, Han SW. A method for selecting ground motions that considers target response mean and variance as well as correlation structure. J Earthq Eng 2016; 20(8): 1263-77.
[30] Ha SH, Han SW. An efficient method for selecting and scaling ground motions matching target response spectrum mean and variance. Earthq Eng Struct Dyn 2016; 45: 1381-87.
[31] Baker JW, Lee C. An improved algorithm for selecting ground motions to match a conditional spectrum. J Earthq Eng 2018; 22(4): 708-23.
[32] Katsanos EI, Sextos AG, Manolis GD. Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering perspective. Soil Dyn Earthq Eng 2010; 30(4): 157-69.
[33] Sextos AG. Selection of GMs for Response History Analysis. Encyclopedia of Earthquake Engineering. https://doi.org/10.1007/978-3-642-36197-5; 2014.
[34] Kwong NS, Chopra AK. Selection and Scaling of GMs for Nonlinear Response History Analysis of Buildings in Performance-Based Earthquake Engineering. Berkeley, California: PEER Technical Report 2015/11; 2015.
[35] Kwong NS, Chopra AK. Evaluation of the exact conditional spectrum and generalized conditional intensity measure methods for GM selection. Earthq Eng Struct Dyn 2016; 45: 757–77.
[36] Moschen L, Medina R, Adam C. A ground motion record selection approach based on multi-objective optimization. J Earthq Eng 2017; published online: https://doi.org/10.1080/13632469.2017.1342302.
[37] Mergos PE, Beyer K. Loading protocols for European regions of low to moderate seismicity. B Earthq Eng 2014; 12(6): 2507–30.

[38] Mergos PE, Kappos AJ. A combined local damage index for seismic assessment of existing RC structures. Earthq Eng Struct Dyn 2013; 42(6): 833-52.
[39] Yang XS. Nature-inspired optimization algorithms. London: Elsevier Insights; 2014.
[40] Naeim F, Alimoradi A, Pezeshk S. Selection and scaling of GM time histories for structural design using genetic algorithms. Earthq Spectra 2004; 20(2): 413-26.
[41] Macedo L, Castro JM. SelEQ: An advanced GM record selection and scaling framework. Adv Eng Softw 2017; 114: 32-47.
[42] Stefanidou SP, Sextos AG, Kotsoglou AN, Lesgidis N, Kappos AJ. Soil-structure interaction effects in analysis of seismic fragility of bridges using an intensity-based ground motion selection procedure. Eng Struct 2017; 151: 366-80.
[43] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE T Evolut Comput 2002; 6: 182-97.
[44] European Strong-motion database (ESD). http://www.isesd.cv.ic.ac.uk/ESD/ frameset.htm.
[45] PEER GM Database. https://ngawest2.berkeley.edu.
[46] Reiter L. In: Earthquake hazard analysis issues and insights. Columbia University Press: New York; 1990.
[47] Abrahamson N, Silva W. Empirical response spectral attenuation relations for shallow crustal earthquakes. Seismol Res Lett 1997; 68(1): 94-127.
[48] Boore DM, Atkinson GM. Ground-Motion prediction equations for the average horizontal component of PGA, PGV, and 5%-Damped PSA at Spectral Periods between 0.01 s and 10.0 s. Earthq Spectra 2008; 24(1): 99-138.
[49] Messac A. Optimization in practice with MATLAB for engineering students and professionals. Cambridge: Cambridge University Press; 2015.
[50] MathWorks. MATLAB R2017a – Global Optimization Toolbox. Natick, MA, USA: The MathWorks Inc.; 2017.
[51] Holland J. Adaptation in natural and artificial systems. Ann Arbor, MI, USA: University of Michigan Press; 1975.
[52] Deep K, Singh KP, Kansal ML, Mohan C. A real coded genetic algorithm for solving integer and mixed integer optimization problems. Appl Math Comput 2009; 212(2): 505–18.
[53] Deb K. An efficient constraint handling method for genetic algorithms. Comput Method Appl M 2000; 186(2–4): 311–38.
[54] FEMA P-58. Next-generation Seismic Performance Assessment for Buildings, Volume 1 – Methodology. Washington, D.C.: Federal Emergency Management Agency; 2012.
[55] Bommer J, Martinez-Pereira A. The effective duration of earthquake strong ground motions. J Earthq Eng 1999; 3(2): 127-72.
[56] Koutrakis SI, Karakaisis GP, Hatzidimitriou PM, Koliopoulos PK, Margaris VN. Seismic hazard in Greece based on different strong ground motion parameters. J Earthq Eng 2002; 6(1): 75-109.