



A ground motion selection and modification method through stratified sampling

Aris-Artemis I. Theophilou¹

Received: 11 June 2018 / Accepted: 1 September 2018 / Published online: 22 September 2018
© The Author(s) 2018

Abstract

In this article a ground motion selection and modification (GMSM) method is presented, suitable for the probabilistic seismic assessment of the damage state and collapse potential of building structures. The objective is to predict the probability distribution of the engineering demand parameters in a future earthquake event, provided that the spectral acceleration at the fundamental period of the structure is given. The GMSM method uses a vector-valued intensity measure that incorporates the Normalized Spectral Area parameter. Through stratified sampling on the Normalized Spectral Area, optimised ground motion suites are formed. Its advantage over other GMSM methods is that it implicitly matches the multivariate distribution of the response spectrum in the region of the structure elongated period, and it adopts an unbiased estimator of the response central tendency. The GMSM method is applied in the probabilistic response assessment of a first-mode dominated multi-degree-of-freedom system that represents a ten-storey building. Substantial reduction in the computational work is achieved, compared to another method.

Keywords Ground motion selection method · Intensity measure · Normalized spectral area · Nonlinear response · Probabilistic seismic assessment

1 Introduction

A ground motion selection and modification (GMSM) method is presented, suitable for the probabilistic seismic assessment of the damage state and collapse potential of building structures. The objective is to predict the probability distribution of the engineering demand parameters (EDPs) when the structure is subjected to a ground motion having a spectral acceleration $S_a(T_1)$ at the fundamental period of the structure, T_1 . Its advantage over other GMSM methods is that it implicitly matches the multivariate distribution of the response spectrum in the region of the structure elongated period, and it adopts an unbiased estimator of the EDP central tendency.

✉ Aris-Artemis I. Theophilou
aris@theophilou.com

¹ Department of Civil and Environmental Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, UK

Ground motion intensity is expressed using the vector-valued intensity measure (IM) $\langle S_a(T_1), S_{dN}(T_1, T'_1) \rangle$ (Theophilou 2014; Theophilou et al. 2017), where T'_1 is an approximation of the elongated period of the structure due to inelasticity effects. Optimized suites of ground motions are obtained from a large ground motion dataset, through stratified sampling on $S_{dN}(T_1, T'_1)$. The optimized suites are used in the dynamic analysis of the structure, resulting in a response prediction that is optimized, compared to using a less efficient IM, or to random sampling. An optimized response prediction implies that a reduced number of ground motions is required to obtain the same level of prediction accuracy, or, conversely, an improved accuracy is achieved when the same number of ground motions is used. The concept is that stratified sampling on IM results in an optimized replication of the central tendency and the dispersion of the IM as well as the associated EDPs, provided that there is sufficiently high correlation between the IM and the EDP.

The proposed GSM method is applied to the dynamic analysis of a multi-degree-of-freedom (MDOF) structure with high participation of the fundamental mode, which is assumed to represent a real ten-storey building structure designed to Eurocode 2 (2004a) and Eurocode 8 (2004b) for ductility class ‘High’.

2 Motivation and background

Probabilistic seismic response assessment is of interest both in the design of new structures, and in the assessment of existing structures. In the design of new structures the goal is to ensure that the safety level required by the building codes is attained, while in the assessment of existing structures the goal is to quantify the inherent safety level.

Dynamic response-history analysis has various advantages over other methods of evaluating structural response, i.e. it is more accurate, it enables the explicit evaluation of the response at every time step, and it can be used in cases where other methods are unsuitable, such as with structures having complex configuration and complex nonlinear response. However, there are two significant impediments in adopting dynamic response-history analysis for meeting the above safety objectives. First, the dynamic response-history analysis of MDOF structures requires substantial computational work, especially when nonlinear behaviour is considered. Even though computers are becoming continuously more sophisticated in terms of processing power and memory resources, the sophistication and complexity of the mathematical models representing the structural system are also evolving. The specification of appropriately small suites of ground motions, consequently resulting in decreased computational work, will always be a topic of great interest and utility. Second, although the number of recorded ground motions is continuously increasing, there is still relative scarcity of high intensity records, which are of most interest. These two challenges call for the need to develop GSM methods, to predict the true response with sufficient accuracy and efficiency.

A convenient way of conveying the seismic intensity of the earthquake scenario to the structural engineer is through the spectral ordinate at the fundamental period of the structure, such as $S_a(T_1)$. One category of GSM methods aim at predicting the central tendency of the structural response, given $S_a(T_1)$. Such methods are the response spectrum matching (e.g. Eurocode 8 2004b; ICB 2015), which can be facilitated with software tools such as REXEL (Iervolino et al. 2010), the Conditional Mean Spectrum matching proposed by Baker (2011), the genetic algorithm selection and scaling proposed by Naeim et al. (2004), the selection based on a vector of record properties identified by proxy proposed by Watson-Lamprey and

Abrahamson (2006), the selection based on a precedence list proposed by Azarbakht and Dolsek (2007), the selection using a harmony search algorithm proposed by Kayhan et al. (2011), the scaling to target scenario with epsilon preservation and selection based on median dispersion minimization proposed by Ay and Akkar (2012), and the integrated software environment coupling ground motion selection with structural analysis proposed by Katsanos and Sextos (2013).

The proposed GSM method predicts, in addition to the central tendency, the dispersion of the response. A method consistent with this objective is the FEMA P-58 (2012) procedure. A weakness of this method is that by selecting all records to have essentially the same shape, the multivariate nature of the response spectrum distribution (Jayaram and Baker 2008) is suppressed. Jayaram et al. (2011) illustrate through application examples that GSM methods focusing on matching the response spectrum shape underestimate the response central tendency and dispersion. The semi-automated procedure proposed by Kottke and Rathje (2008), matches the central tendency and the dispersion of the entire suite. Since the procedure matches the median response spectrum to the target response spectrum, the response spectrum shape of the individual records within a suite could exhibit significant variation. However, this procedure does not match the multivariate distribution of the response spectrum.

Some GSM methods have been proposed that consider, directly or indirectly, the multivariate nature of the response spectrum distribution (Jayaram and Baker 2008). Shome et al. (1998) suggested forming bins of records based on magnitude and distance, and then forming record suites through random sampling and scaling them to $S_a(T_1)$; this method is used herein for comparison with the proposed GSM. Goulet et al. (2007) selected records to match the deaggregation results in terms of magnitude, distance, and epsilon (Baker and Cornell 2005), and subsequently scaled them to the target $S_a(T_1)$. Jayaram et al. (2011) proposed an algorithm for selecting suites of records to match simulated response spectra that have a specified mean, variance, and correlation between any two periods. Wang (2011) proposed a similar method, with the difference that the response spectra are conditioned on magnitude and distance, rather than $S_a(T_1)$. Bradley (2010) proposed the concept of the Generalized Conditional Intensity Measure, which is a vector-valued IM that contains a multitude of different IMs; records are selected using the algorithm proposed by Bradley (2012) such that the empirical distribution function of the suite matches the target IM distribution.

The proposed GSM method is a contribution to the methods that match the multivariate nature of the response spectrum distribution.

3 Intensity measure

The proposed GSM method uses the vector-valued IM shown below

$$\langle S_a(T_1), S_{dN}(T_1, T'_1) \rangle \quad (1)$$

The vector-valued IM is comprised of the spectral acceleration, $S_a(T_1)$, which is an absolute measure, and the Normalized Spectral Area parameter, $S_{dN}(T_1, T'_1)$ (Theophilou 2014; Theophilou et al. 2017), which is a relative measure, given by

$$S_{dN}(T_1, T'_1) = \frac{1}{S_d(T_1)T_N T_1} \int_{T_1}^{T'_1} S_d(T) dT, \quad T_1 < T'_1 \quad (2)$$

where T_1 is the initial fundamental period of the structure, T'_1 is an approximation of the elongated period due to inelasticity effects, $T_N = 1.0$ s is a normalizing constant, and $S_a(T_1)$ is the response spectrum displacement at period T_1 .

Due to the normalization to $S_a(T_1)$, the $S_{dN}(T_1, T'_1)$ value does not change with scaling. In this way, $S_{dN}(T_1, T'_1)$ captures the effect of the excitation spectral characteristics (i.e. frequency content) on the response. Thus, it is a measure of the intensity that affects the inelastic response associated with period elongation. In turn, the degree of period elongation depends on the frequency content, which is unique for each ground motion. Hence, the purpose of integrating the response spectrum is to capture the elongated period within appropriately estimated bounds. In this context, $S_{dN}(T_1, T'_1)$ can be seen as a descriptor of the local response spectrum shape between periods T_1 and T'_1 .

Theophilou et al. (2017) review a number of similar vector-valued IMs proposed by other researchers.

4 Specification of large-sample datasets

A dataset comprised of a large number, N_{gm} , of ground motions is initially formed, which is sufficiently large, e.g. $N_{gm} \geq 30$, so that the central tendency and the dispersion of the ground motion parameters can be accurately captured (Walpole et al. 2007). It is therefore termed ‘ground motion large-sample dataset’, denoted as

$$G = \{GM_i | i = 1, 2, 3, \dots, N_{gm}\} \quad (3)$$

where GM_i is ground motion i . The GSM method performs sampling on dataset G , to form optimized suites of ground motions.

The vectors representing the IM and the EDPs at a particular $S_a(T_1)$ are denoted as \mathbf{IM} and \mathbf{EDP} , respectively. More specifically, \mathbf{IM}_i and \mathbf{EDP}_i represent the \mathbf{IM} and \mathbf{EDP} , respectively, that correspond to ground motion i . The dataset of the \mathbf{IM}_i and \mathbf{EDP}_i pairs is therefore termed ‘dynamic analysis large-sample dataset’, denoted as

$$S = \{(\mathbf{IM}_i, \mathbf{EDP}_i) | i = 1, 2, 3, \dots, N_{gm}\} \quad (4)$$

\mathbf{IM} is the vector-valued IM presented in (1). The \mathbf{EDP} vector is comprised of scalar EDPs that are considered as relevant for the particular problem studied. The EDPs considered should be appropriately selected so as to collectively provide a representative description of the structural damage state and collapse potential, consistent with the objective of structural assessment. It is, thus, desirable that EDPs representative of different damage/failure mechanisms are included.

5 Methodology

The GSM method aims to replicate the true inelastic response distribution of the structure, for a given $S_a(T_1)$. In the present section the GSM method is presented, and in the following section the sample statistics are described. Use is made of the IM and EDP distributions, appropriately transformed so that they are normally distributed, and denoted as follows: \mathbf{IMT} is defined as the $S_{dN}(T_1, T'_1)$ element of \mathbf{IM} transformed so that its distribution is normal; \mathbf{EDPT} is defined as a scalar element of \mathbf{EDP} transformed likewise.

The proposed GSM method consists of the following steps:

Step 1: Formation of ground motion large-sample dataset

Initially, dataset G is formed, having a sufficiently large size (e.g. $N_{gm} \geq 30$). Ground motions selected have seismological parameters (such as magnitude, and source-to-site distance) consistent with the earthquake scenario. It is strongly recommended to check the distribution of the dataset with respect to a ground motion prediction model representative of the expected earthquake.

Step 2: Ground motion normalization to $S_a(T_1)$

All ground motions in dataset G are normalized to $S_a(T_1)$. An upper limit can be imposed to the scale factor, e.g. 3 as proposed by Shome et al. (1998) or 4 by Iervolino and Cornell (2005), beyond which scaling is considered inappropriate (i.e. the frequency characteristics of the scaled motion are not consistent with its amplitude); ground motions exceeding this limit should be discarded.

Step 3: Determination of T_1 and T'_1

The integration interval periods T_1 and T'_1 are determined, which are then used in the evaluation of $S_{dN}(T_1, T'_1)$. T'_1 represents the ‘ultimate’ elongated period and can be calculated using the simplified procedure proposed by Theophilou et al. (2017), or the procedure proposed in FEMA 440 (2005). Alternatively, the optimum T'_1 can be evaluated by first carrying out dynamic analyses on a single-degree-of-freedom system using all ground motions in dataset G . Then, through regression analysis for a range of candidate T'_1 values, the final T'_1 value is selected as the one with the highest correlation between IMT and $EDPT$.

Step 4: Calculation of $S_{dN}(T_1, T'_1)$

$S_{dN}(T_1, T'_1)$ is calculated for each ground motion using Eq. (2).

Step 5: Evaluation of IMT mean and variance

The IMT mean and variance of dataset G are evaluated.

Step 6: Partition of IMT distribution into N_s strata

The IMT domain is partitioned into N_s strata, such that all strata have equal probability of occurrence. It is strongly recommended that at least $N_s = 5$ strata are used.

Step 7: Formation of optimized suites

In the last step, optimized suites are formed through ‘stratified sampling’, by selecting N_i ground motions from each stratum. The total number of selected ground motions is $N_{g,s} = N_s \cdot N_i$.

6 Estimation of response distribution through stratified sampling

6.1 Statistical dependence between intensity and response

Simple linear regression analysis is performed between the independent variable IMT , and the dependent variable $EDPT$, using the data in dataset S ,

$$EDPT = \alpha + \beta \cdot imt + \varepsilon \quad (5)$$

where α is the axis intercept, β is the slope, ε is the random error that is assumed to be normally distributed with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$, and imt is the value of IMT . The regression equation parameters α and β are determined, together with the variance, σ^2 , and the correlation, ρ . By visual observation of the regression analyses in Theophilou (2014), σ^2 can be inferred to be constant with respect to IMT . These are used next, in the evaluation of the stratified sampling statistics.

6.2 Stratified sampling on *IMT*

The theoretical distribution of *IMT* is normal with mean μ_{IMT} and standard deviation σ_{IMT} . The area under the curve is divided into N_s strata, which have equal probability of occurrence. Random variable IMT_i represents variable *IMT* within stratum i . Within any stratum i the theoretical mean is denoted as μ_{IMTi} , and the theoretical standard deviation as σ_{IMTi} . The sampling fraction of stratum i , w_{IMTi} , is equal to $1/N_s$ for all strata.

The estimator of the sample mean IMT_i , \overline{IMT}_i , is an unbiased estimator of μ_{IMTi} , on the assumption of simple random sampling within stratum i (e.g. Ang and Tang 1975). Consequently, the estimator of the sample mean *IMT*, \overline{IMT} , is an unbiased estimator of μ_{IMT} (e.g. Cochran 1977).

6.3 *EDPT* statistics

The method proposed, based on regression analysis principles, finds the relationship that best fits to the data in dataset S . The simple linear regression model is adopted to describe the relationship between the sample mean *EDPT* of stratum i , \overline{EDPT}_i , and the value of \overline{IMT}_i , \overline{imt}_i , as shown below

$$\overline{EDPT}_i = \alpha + \beta \cdot \overline{imt}_i + \varepsilon \tag{6}$$

The sample mean *EDPT*, \overline{EDPT} , can be calculated from \overline{EDPT}_i , as shown below

$$\overline{EDPT} = \sum_{i=1}^{N_s} w_{EDPT_i} \cdot \overline{EDPT}_i \tag{7}$$

where w_{EDPT_i} is equal to w_{IMTi} , because of the linear relationship assumed.

It is thus possible to calculate \overline{EDPT} from \overline{imt}_i and from \overline{imt} as

$$\overline{EDPT} = \frac{\sum_{i=1}^{N_s} (a + b \cdot \overline{imt}_i)}{N_s} = a + b \frac{\sum_{i=1}^{N_s} \overline{imt}_i}{N_s} = a + b \cdot \overline{imt} \tag{8}$$

where a , and b are the estimates of α and β . Since \overline{IMT} is an unbiased estimator of the mean *IMT*, and \overline{EDPT} is a linear function of \overline{IMT} , it is proved that \overline{EDPT} is an unbiased estimator of the theoretical mean *EDPT*.

The standard error of the mean *EDPT*, $\sigma_{\overline{EDPT},s}$, is (Cochran 1977)

$$\sigma_{\overline{EDPT},s} = \sqrt{\sum_{i=1}^{N_s} w_{EDPT_i}^2 \cdot \sigma_{EDPT_i}^2} = \sqrt{\sum_{i=1}^{N_s} \frac{\sigma^2}{N_s^2}} = \frac{\sigma}{\sqrt{N_s}} \tag{9}$$

where $\sigma_{EDPT_i}^2$ is the variance of *EDPT* in stratum i , assumed to be equal to σ^2 .

6.4 Comparison to random sampling

Stratified sampling on *IMT* nearly always results in a standard error of the mean that is lower than that obtained through random sampling (Cochran 1977). In this section it is shown that, as a consequence, the standard error of the mean *EDPT* is also lower, by a degree that depends on ρ .

The ratio $\sigma_{EDPT,s}/\sigma_{EDPT,r}$ can be evaluated assuming an equal number of ground motions between stratified and random sampling, i.e. $N_{g,s} = N_{g,r}$, where $N_{g,r}$ is the number of randomly sampled ground motions, and $\sigma_{EDPT,r}$ is the standard error of the mean *EDPT* through random sampling. The following equation gives the relationship of $\sigma_{EDPT,s}/\sigma_{EDPT,r}$ to ρ for the case of $N_i = 1$

$$\frac{\sigma_{EDPT,s}}{\sigma_{EDPT,r}} = \frac{\sigma/\sqrt{N_s}}{\sigma/(\sqrt{N_{g,r}}\sqrt{1-\rho^2})} = \sqrt{1-\rho^2} \tag{10}$$

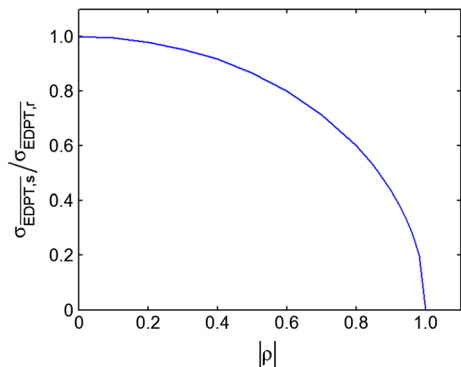
Plotted in Fig. 1 is $\sigma_{EDPT,s}/\sigma_{EDPT,r}$, for $|\rho|$ values from 0 to 1.0.

It can be inferred from Fig. 1 that at one extreme, at $\rho = 0$, $\sigma_{EDPT,s}$ is equal to $\sigma_{EDPT,r}$, which means that stratified sampling is not more accurate than random sampling. At the other extreme, at $|\rho| = 1.0$, $\sigma_{EDPT,s}$ is equal to zero, which means that the *IMT* obtained through stratified sampling on *IMT* can be used to evaluate the exact *EDPT*. In the intermediate region, $\sigma_{EDPT,s}/\sigma_{EDPT,r}$ is decreasing as $|\rho|$ is increasing, which means that, stratified sampling on *IMT* is potentially more efficient than random sampling, because $\sigma_{EDPT,s}$ is lower than $\sigma_{EDPT,r}$. The decrease becomes more pronounced at high $|\rho|$ values, because $\sigma_{EDPT,s}$ is decreasing at a higher rate than $\sigma_{EDPT,r}$.

6.5 Size of suites

The size of the suites plays an important role in the accuracy of *IMT* and *EDPT* estimation. Theophilou (2014) applied the method to a single-degree-of-freedom system assumed to represent an idealized structure. Using suites of $N_{g,s} = 8$ ground motions ($N_s = 8, N_i = 1$) resulted in a standard error in the mean *IMT* of $< 2\%$, and in a standard error in the mean *EDPT* of approximately 10%. The variance of *IMT* was overestimated by about 10%, and the variance of *EDPT* by about 10%. This estimation accuracy is deemed acceptable.

Fig. 1 Relationship between $\sigma_{EDPT,s}/\sigma_{EDPT,r}$ and $|\rho|$



Using suites of $N_{g,s} = 5$ ground motions ($N_s = 5$, $N_i = 1$) resulted in less accurate, but still acceptable estimation.

Further research could lead to further optimization of the method with respect to the N_s and N_i combinations used, e.g. choosing between ($N_s = 4$, $N_i = 2$) and ($N_s = 8$, $N_i = 1$).

6.6 Multivariate distribution of response spectrum

The objective of the proposed GSM method is to match the multivariate distribution of the response spectra between T_1 and T'_1 . This is achieved in an implicit manner, by matching the dispersion of $S_{dN}(T_1, T'_1)$, given that the response spectra are conditioned on $S_a(T_1)$. The match is implicit, in the sense that instead of matching the multivariate distribution of the response spectra over a range of discrete periods, the dispersion of the scalar $S_{dN}(T_1, T'_1)$ is matched.

7 Application to a building structure

The GSM method was applied in the probabilistic response assessment of a first-mode dominated MDOF system, representative of a real building structure.

7.1 Structure description

The structure examined is a ten-storey reinforced concrete building, used for residential occupancy. Each storey has a rectangular floor plan of dimensions 24 m × 16 m, and a height of 3.2 m. The columns are located in a 5 × 5 array, spaced 6 m apart in the longitudinal direction, and 4 m apart in the transverse direction. The structure was designed to Eurocode 2 (2004a) and Eurocode 8 (2004b) for ductility class ‘High’. Further details of the structure are provided in Theophilou (2014).

7.2 Finite element model

A two-dimensional finite element model was developed, representing one interior frame in the longitudinal direction of the structure. The model was analysed using the computer program OpenSees version 2.4.0 (PEER 2013). Beams and columns were represented as elastic beam-column elements. At the locations of the beam and column joints, and at the location of the column and foundation joints, the beam and column frame elements are connected to the respective nodes by zero-length ‘*nonlinear spring*’ elements that allow rotation only. The moment-rotation relationship assigned to the nonlinear springs is obtained using the modified Ibarra-Krawinkler hysteresis model (Lignos and Krawinkler 2012a). The nonlinear spring parameters were determined using the empirical relationships and values given by Haselton and Deierlein (2007), and by Lignos and Krawinkler (2012b). The hysteresis rules are consistent with the modified Clough-Johnston model (Clough and Johnston 1966; Mahin and Lin 1983), with the exception that they are modified to conform with the multilinear backbone curve. The finite element model considered second-order (P-Delta) effects.

Table 1 Natural modes of structure

Mode	Period (s)	Mass participation ratio (%)
1	1.452	81.0
2	0.479	9.9
3	0.276	3.7

7.3 Natural modes

The first three natural periods of the structure and the corresponding mass participation ratios are summarized in Table 1. As anticipated, the structure is first-mode dominated, which is consistent with the concept of selecting ground motions based on the inelastic response of the first mode, adopted in the proposed GSM method.

7.4 Ground motion large-sample dataset

The ground motion large-sample dataset (i.e. dataset G) was formed by selecting 40 ground motions, listed in Table 2, with seismological characteristics consistent with the earthquake scenario. The earthquake scenario can be described as a strong earthquake, at a close distance to the fault, but sufficiently long to avoid near-fault effects. The main criteria used in the selection of the records were a moment magnitude higher than 6, and closest distance to fault less than about 30 km. The source of the records was the NGA Database 2005 (PEER 2005) and the European Strong-Motion Database (Ambraseys et al. 2002). The graph in Fig. 2 presents the distribution of the moment magnitude and closest distance of the records. It can be observed that these two variables are widely distributed throughout.

To ensure that the dataset of ground motions is a representative sample of the population, its statistics were compared to the Boore and Atkinson (2008) ground motion prediction model. Epsilon was evaluated in the period range 0.5–2.0 s and was found to match very well the standard normal distribution, which is the theoretical distribution of epsilon.

7.5 Ground motion suites

One dataset of 2000 optimized suites was formed through stratified sampling, using $N_s = 8$ strata, sampling $N_i = 1$ ground motion from each stratum, resulting in $N_{g,s} = 8$ ground motions per suite. For comparison, another dataset of 2,000 suites was formed through random sampling, using $N_{g,r} = 8$ ground motions per suite.

Figure 3 shows the displacement response spectra of the ground motions, normalized to $S_d(1.45\text{ s})$ so as to facilitate the observation of the spectral shape in the region $T_1 = 1.45\text{ s}$ to $T'_1 = 2.90\text{ s}$. Figure 3a shows the spectral distribution of dataset G , and Fig. 3b shows suite No. 1 obtained through stratified sampling. It can be observed that the distribution of the suite approximates well the distribution of dataset G .

The “Random” sampling method referred to herein, corresponds to the method described in PEER Report 2009/01 (Haselton et al. 2009) as “ $S_a(T_1)$ Scaling with Bin Selection” proposed by Shome et al. (1998). With this GSM method a bin of ground

Table 2 Ground motion records

Earthquake	Station	<i>M</i>	<i>R</i> (km)	Azimuth/direction
Loma Prieta	CDMG 47379	6.93	28.64	000
18/10/1989—00:05	Gilroy Array #1			090
Victoria, Mexico	Cerro Prieto	6.33	33.73	045
09/06/1980—03:28				315
Coalinga	CDMG 46175	6.36	33.52	045
02/05/1983—23:42	Slack Canyon			315
San Fernando	CDMG 126	6.61	24.19	111
09/02/1971—14:00	Lake Hughes #4			201
Duzce, Turkey	Lamont 531	7.14	27.74	000
12/11/1999				090
Kozani, Greece	ITSAK 99999	6.40	18.27	L
13/05/1995—08:47	Kozani			T
Irpinia, Italy	ENEL 99999	6.20	22.29	000
23/11/1980—19:35	Bagnoli Irpinio			270
Whittier Narrows	CDMG 24399	5.99	19.56	000
01/10/1987—14:42	Mt Wilson			090
Basso Tirreno	Milazzo	6.00	34	NS
15/04/1978—23:33				EW
Montenegro	Hercegnovi Novi	6.90	65	NS
15/04/1979—06:19	Pavicic School			EW
Tabas, Iran	9102 Dayhook	7.35	20.63	LN
16/09/1978				TR
Umbria Marche	Assisi-Stallone	6.00	21	NS
26/09/1997				EW
North Palm Springs	CDMG 12206 Silent Valley	6.06	20.70	000
08/07/1986				090
Loma Prieta	USGS 1032 Hollister	6.93	49.52	270
18/10/1989—00:05				360
Chi-Chi, Taiwan	CWB 99999 TCU045	7.62	77.50	N
20/09/1999				E
Northridge	USC 90059 Burbank Howard	6.69	23.18	060
17/01/1994—12:31				330
San Fernando	USGS 266 Pasadena	6.61	39.17	180
09/02/1971—14:00				270
Whittier Narrows	USC 90017	5.99	28.48	075
01/10/1987—14:42	LA Wonderland			165
Northridge	USGS 5080	6.69	19.19	270
17/01/1994—12:31	Monte Nido			360
Irpinia, Italy	Auletta	6.90	33.10	000
23/11/1980—19:34				270

Fig. 2 Magnitude-distance distribution of records (each point corresponds to one station and hence two horizontal records)

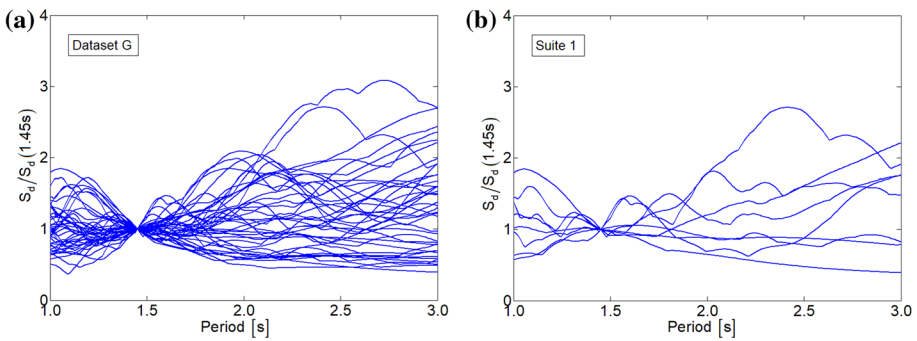
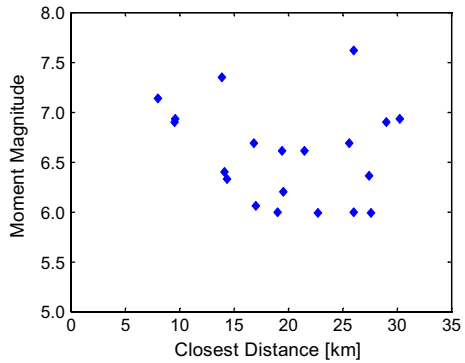


Fig. 3 Ground motion displacement response spectra normalized to $S_d(1.45\text{ s})$ **a** dataset *G*, **b** Suite No. 1 through stratified sampling

motions is first selected with moment magnitude and distance criteria. Then a suite of ground motions is formed through random sampling without replacement, and scaling to $S_a(T_1)$.

7.6 Incremental dynamic analysis

Incremental dynamic analysis (Vamvatsikos and Cornell 2002) was carried out by incrementing $S_a(T_1)$ up to 1.2 g. $S_{dN}(T_1, T'_1)$ is evaluated by integration of the elastic displacement spectrum from $T_1 = 1.45\text{ s}$ to $T'_1 = 2.90\text{ s}$, i.e. $T'_1 = 2.0T_1$. The upper intensity limit corresponds to the highest $S_a(T_1 = 1.0\text{ s})$ found on seismic hazard maps (e.g. Southern California) with probability of exceedance of 2% in 50 years. At this probability of exceedance, safety against collapse is evaluated. The ground motions used at each intensity increment were scaled to the target $S_a(T_1)$. The response parameters investigated were the Park-Ang overall structural damage index (Park and Ang 1985), *OSDI*, and the maximum interstorey drift ratio, *MIDR*.

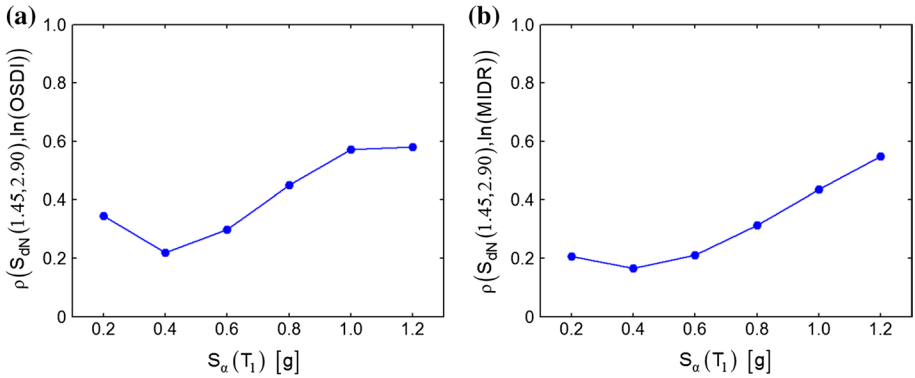


Fig. 4 Correlation coefficient between $S_{dN}(1.45, 2.90)$ and **a** $\ln(OSDI)$, **b** $\ln(MIDR)$

7.7 Regression analysis

The probability distribution of $S_{dN}(1.45, 2.90)$ was found to conform well to the normal distribution, and to the lognormal distribution, using Lilliefors (1967) test at a significance level of 5%. The normal distribution was assumed in the regression analysis, as it resulted in the highest correlation with the response parameters. Similarly, by Lilliefors (1967) test it was found that the probability distributions of the response parameters $OSDI$, and $MIDR$, conform well to the lognormal distribution, at a significance level of 5%.

Thus, regression analyses were carried out between $S_{dN}(1.45, 2.90)$ and the natural logarithms of the response parameters, at various $S_a(T_1)$ intensities, using the simple linear model. The correlation coefficient, ρ , between $S_{dN}(1.45, 2.90)$ and (a) $\ln(OSDI)$ is shown in Fig. 4a, and (b) $\ln(MIDR)$ in Fig. 4b.

It can be observed that, in general, ρ is low (in the range of 0.2–0.3) at the low non-linearity levels (at intensity $S_a(T_1)$ between 0.2 and 0.6 g), and moderate (in the range of 0.5–0.6) at the high nonlinearity levels (at intensity $S_a(T_1)$ between 1.0 and 1.2 g).

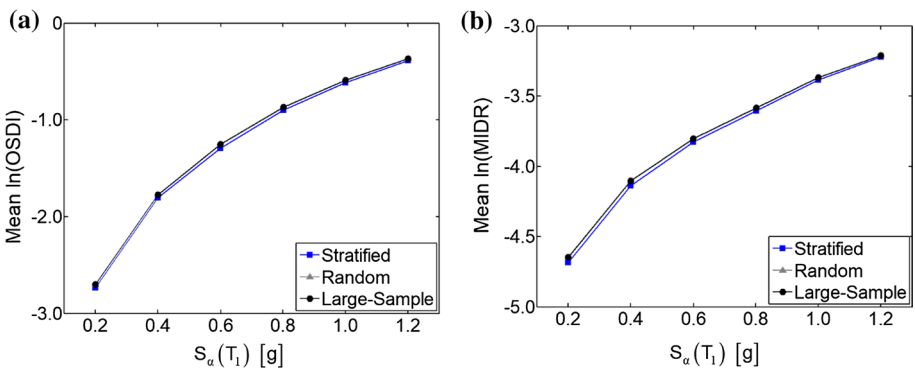


Fig. 5 Mean **a** $\ln(OSDI)$, **b** $\ln(MIDR)$

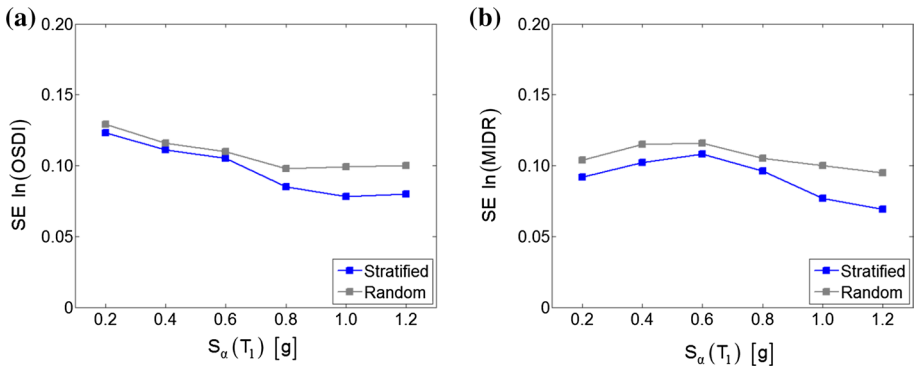


Fig. 6 Standard error (SE) of the mean **a** $\ln(OSDI)$, **b** $\ln(MIDR)$

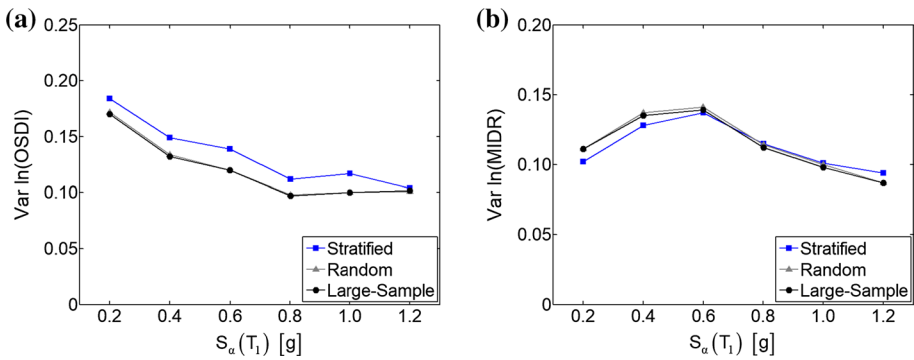


Fig. 7 Variance of **a** $\ln(OSDI)$, **b** $\ln(MIDR)$

7.8 Response sample statistics

Figure 5 shows the mean $\ln(OSDI)$, and $\ln(MIDR)$, plotted against $S_a(T_1)$. It appears that, in general there is excellent conformity between the random sampling, the stratified sampling, and the large-sample dataset results. This is an empirical proof that estimator *EDPT* through stratified sampling is unbiased.

Figure 6 shows the standard error of the mean $\ln(OSDI)$, and $\ln(MIDR)$, plotted against $S_a(T_1)$. It appears that, in general, stratified sampling results in a lower standard error than random sampling, and that there is a marked reduction at $S_a(T_1) \geq 0.8$ g. In particular, the maximum reduction in the standard error is 22% for $\ln(OSDI)$, and 27% for $\ln(MIDR)$. The standard error obtained using both methods, between 7 and 13%, is deemed acceptable.

Figure 7 shows the variance of $\ln(OSDI)$, and $\ln(MIDR)$, plotted against $S_a(T_1)$. It appears that, there is excellent conformity between the random sampling and the large-sample dataset results. The stratified sampling results exhibit some discrepancy from the large-sample dataset results, reaching a maximum of 15% overestimation in the case of $\ln(OSDI)$ (corresponding to 7% overestimation in the standard deviation).

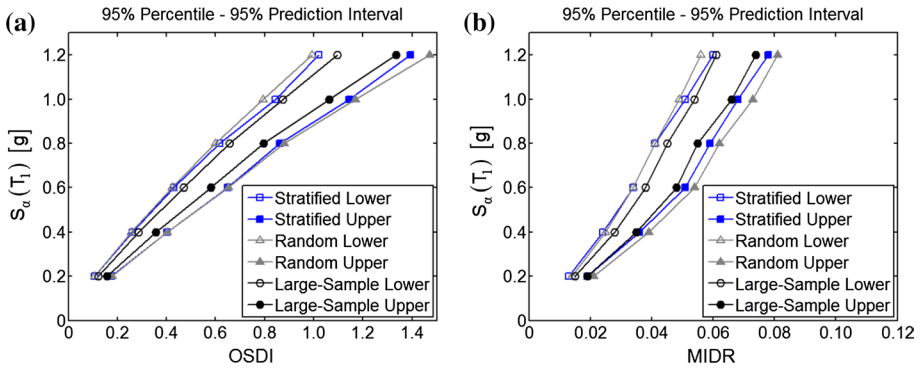


Fig. 8 Prediction intervals of **a** $OSDI$, **b** $MIDR$

7.9 Response prediction

In Fig. 8 response parameters $OSDI$, and $MIDR$ are plotted at 95% probability of occurrence as 95% prediction intervals. It can be observed that the stratified sampling prediction interval is generally narrower than the random sampling prediction interval. It is also observed that the stratified sampling prediction interval converges towards the large-sample prediction interval. It appears that despite the small overestimation in the variance, the proposed GSM method is still more efficient than random sampling.

The efficiency of the GSM method in reducing the prediction interval can be quantified through the ‘Efficiency Index’, EI , defined as

$$EI = \frac{(R_U - R_L) - (S_U - S_L)}{(R_U - R_L) - (L_U - L_L)} \tag{11}$$

where S , R , and L denote the interval bounds of stratified sampling, random sampling, and the large-sample, respectively, and subscripts U and L denote the upper and lower interval bounds, respectively. The higher the EI , the more efficient the GSM method; at zero EI the GSM method is not more efficient than random sampling, whereas at unity EI

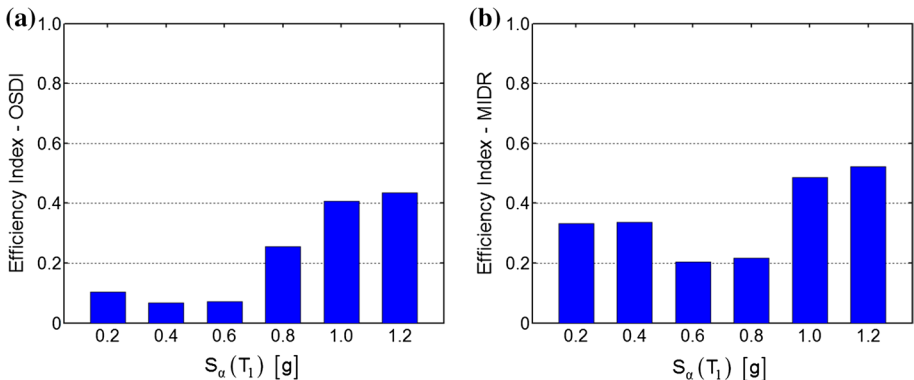


Fig. 9 Efficiency index for **a** $OSDI$, **b** $MIDR$

the GSM method achieves the same accuracy as the large-sample dataset. EI is plotted in Fig. 9 against $S_a(T_1)$. It appears that EI for $OSDI$ is in the range of 0.07–0.43, and for $MIDR$ in the range of 0.20–0.52. The highest EI was observed at $S_a(T_1)$ equal to 1.0 g and 1.2 g; this is attributed to the higher correlations obtained at the high nonlinearity levels.

The computational work needed using the proposed GSM method, expressed as number of ground motions required to obtain the same standard error as through random sampling, was reduced to about 72% at the moderate nonlinearity levels, and to about 50% at the high nonlinearity levels.

7.10 Elongated period estimation

After the application of the GSM method in the response prediction of the MDOF system, in which the elongated period was assumed to be equal to $T'_1 = 2.0T_1$ for practical purposes (Theophilou et al. 2017), the elongated period observed was investigated. The elongated period is the predominant period of oscillation when the structure enters the inelastic range and was estimated from the response power spectrum.

Initially, the power spectrum of the top node response displacement was calculated. Figure 10 shows the response spectrum for the Loma Prieta 090 ground motion at intensities $S_a(T_1) = 0.2$ g and $S_a(T_1) = 1.0$ g. In the $S_a(T_1) = 0.2$ g spectrum it is obvious that the dominant frequency is close to 0.69 Hz, which corresponds to the natural period of the fundamental mode $T_1 = 1.45$ s. The spectral power at frequencies lower than 0.69 Hz is close to zero, inferring a predominantly elastic response. In the $S_a(T_1) = 1.0$ g spectrum it is observed that the spectral power in the frequencies lower than 0.69 Hz is substantial, which is attributed to period elongation due to inelasticity effects. The abrupt increase in the spectral power at frequencies close to zero observed in both spectra has various possible explanations, however this range is not of practical interest. This range is excluded by seeking the elongated period at frequencies above 0.2 Hz.

Within the frequency range of interest, i.e. 0.2 Hz and 0.69 Hz, the elongated period was calculated. The dominant frequencies were first identified as the localised peaks on the power spectrum. The frequency corresponding to the elongated period was calculated as the average of the frequencies, weighted with respect to the spectral power. The curve

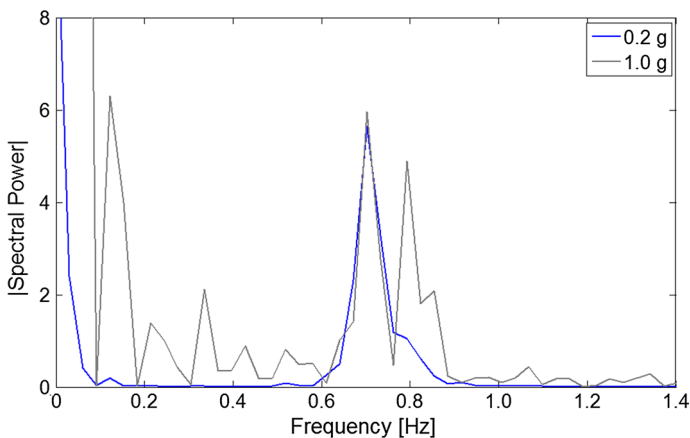


Fig. 10 Power spectrum of Loma Prieta 090 record

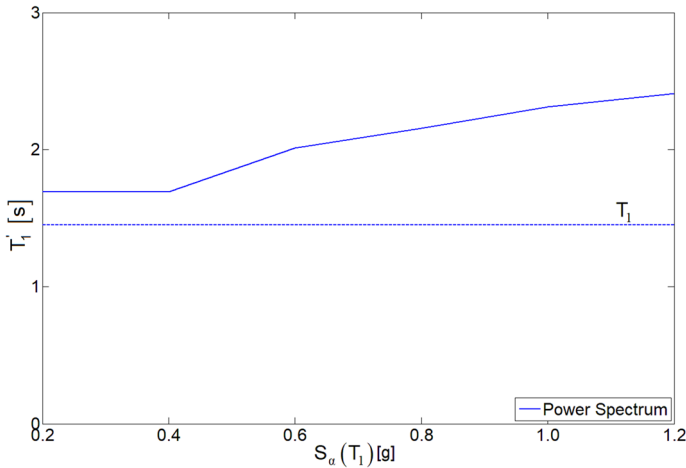


Fig. 11 Mean estimated elongated period

resulting from this approach is presented in Fig. 11. The mean estimated elongated period, T'_1 , at $S_a(T_1)$ equal to 1.0 g, and 1.2 g, was found to be 2.31 s, and 2.41 s, respectively, using the power spectrum method. This is a reasonable agreement with the assumed elongated period of 2.90 s.

8 Differentiation to other methods

Goulet et al. (2007) suggested selecting records to match the deaggregation results in terms of magnitude, distance, and epsilon (Baker and Cornell 2005), and subsequently scaling them to the target $S_a(T_1)$. In their method the spectral shape (i.e. epsilon) is assumed to be dependent on the magnitude and distance, thus a sufficiently large number of records (e.g. in their example 34) is required to cover all magnitude-distance combinations. The difference of the present GSM method is that the selection criteria of magnitude and distance are relaxed. Thus, the method is applicable when the normalized spectral shape between T_1 and T'_1 (i.e. Normalized Spectral Area) does not change significantly with magnitude and distance, as supported by the findings of Shome et al. (1998). In this way the number of required records for the same magnitude and distance ranges is reduced.

Jayaram et al. (2011) proposed a method that initially generates simulated acceleration response spectra, which collectively match the central tendency, dispersion, and multivariate distribution of the target response spectrum. Subsequently, ground motions are selected that match the simulated response spectra. As Jayaram et al. (2011) states, the suite of selected ground motions resulting from the main procedure may deviate slightly from the target central tendency and dispersion. For this reason, a supplementary “greedy” procedure is specified to replace the ground motions one-by-one and thus minimize the residuals.

Bradley (2012) proposed an algorithm that generates random simulations of response spectra from the conditional multivariate distribution of intensity measures, obtained from the Generalized Conditional Intensity Measure (Bradley 2010). The method considers a multitude of intensity measures (including spectral acceleration), in contrast to the Jayaram

et al. (2011) method that considers only spectral acceleration. It can also utilize the full seismic hazard disaggregation probability. Bradley (2012) also discusses the issue of variability between the matching ground motion suites, particularly for small suites, and proposed selecting the suite with the lowest residual between the empirical and target conditional distribution.

Both the Jayaram et al. (2011) method and the Bradley (2012) method place stringent requirements over the matching spectral shape. As Bradley (2012) states, finding a ground motion with identical spectral shape to the simulation is generally not possible. For this reason, ground motion selection is based on minimizing the spectral shape deviation from the simulations. The proposed GSM method adopts a different approach, in which selection is based on $S_{dN}(T_1, T'_1)$, which is a scalar parameter descriptive of the spectral shape, rather than the actual spectral shape. By limiting the matching period range between T_1 and T'_1 , the number of matching ground motions is higher. In addition, it adopts an unbiased estimator of the EDP central tendency, and it has been found to match the EDP dispersion well, thus avoiding the problem of residuals associated with the Jayaram et al. (2011) and Bradley (2012) methods.

9 Concluding remarks

A GSM method has been proposed with which optimized suites of ground motions can be sampled, suitable for a probabilistic seismic response assessment of building structures. The key aim of the method is to estimate with reasonable accuracy the central tendency and the dispersion of the EDPs. It can be used in cases wherein seismic intensity can be defined in terms of $S_a(T_1)$. Its advantage over other GSM methods is that it implicitly matches the multivariate distribution of the response spectrum in the region of the structure elongated period. It also adopts an unbiased estimator of the EDP central tendency, and has been found to match the EDP dispersion well.

Ground motion intensity is expressed using the vector-valued IM $\langle S_a(T_1), S_{dN}(T_1, T'_1) \rangle$. Optimized suites of ground motions are formed by partitioning the distribution of $S_{dN}(T_1, T'_1)$ into N_s strata, and then selecting N_i ground motions from each stratum. The proposed method replicates the mean and variance of *IMT*, whereas the standard error is reduced, compared to random sampling. At the same time the mean and the variance of *EDPT* are also replicated. The advantage over random sampling is that when there is high enough correlation between *IMT* and *EDPT*, stratified sampling results in a reduced standard error in the mean *EDPT*. Consequently, an optimized response prediction is achieved.

The method was applied in the analysis of a first-mode dominated MDOF system, representing a ten-storey building frame designed to Eurocode 2 (2004a) and Eurocode 8 (2004b) for ductility class High. The central tendency of the response was found to have excellent conformity to the large-sample dataset, while some discrepancy in the dispersion estimation was observed. Expressing the 95% percentile response as 95% prediction interval it was found that the highest 'Efficiency Index' achieved was 0.43 for *OSDI*, and 0.52 for *MIDR*. The computational work needed to obtain the same accuracy as through random sampling was reduced to about 72% at the moderate nonlinearity levels, and to about 50% at the high nonlinearity levels. The elongated period, estimated using the power spectrum method, was found to be in reasonable agreement with the assumed elongated period at the higher nonlinearity levels.

In conclusion, the proposed GSM method is most efficient in predicting response at moderate to high nonlinearity levels, due to the significant reduction in the standard error of the *EDPT*, which is a result of the sufficiently high correlations observed between the proposed *IMT* and the *EDPT*. It is therefore suitable for first-mode dominated structures when the probabilistic assessment focuses on limit states associated with moderate to severe damage and collapse.

Acknowledgements The author would like to express his sincere gratitude to Prof Marios K. Chryssanthopoulos and to Prof Andreas J. Kappos, who provided valuable feedback to this article.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Ambraseys N, Smit P, Sigbjornsson R, Suhadolc P, Margaris B (2002) Internet-Site for European Strong-Motion Data, European Commission, Research-Directorate General, Environment and Climate Programme. <http://smbase.itsak.gr/>. Accessed 19 Sept 2012
- Ang AH, Tang WH (1975) Probability concepts in engineering planning and design: volume I—basic principles. Wiley, New York
- Ay BO, Akkar S (2012) A procedure on ground motion selection and scaling for nonlinear response of simple structural systems. *Earthq Eng Struct Dyn* 41(9):1693–1707. <https://doi.org/10.1002/eqe.1198>
- Azarbakht A, Dolsek M (2007) Prediction of the median IDA curve by employing a limited number of ground motions. *Earthq Eng Struct Dyn* 36(15):2401–2421. <https://doi.org/10.1002/eqe.740>
- Baker JW (2011) Conditional mean spectrum: tool for ground motion selection. *J Struct Eng Am Soc Civ Eng* 137(3):322–331. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0000215](https://doi.org/10.1061/(ASCE)ST.1943-541X.0000215)
- Baker JW, Cornell CA (2005) A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. *Earthq Eng Struct Dyn* 34(10):1193–1217. <https://doi.org/10.1002/eqe.474>
- Boore DM, Atkinson GM (2008) 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* 24(1):99–138. <https://doi.org/10.1193/1.2830434>
- Bradley BA (2010) A generalized conditional intensity measure approach and holistic ground motion selection. *Earthq Eng Struct Dyn* 39(12):1321–1342. <https://doi.org/10.1002/eqe.995>
- Bradley BA (2012) A ground motion selection algorithm based on the generalized conditional intensity measure approach. *Soil Dyn Earthq Eng* 40:48–61. <https://doi.org/10.1016/j.soildyn.2012.04.007>
- Clough RW, Johnston SB (1966) Effect of stiffness degradation on earthquake ductility requirements. In: Proceedings of the Japan earthquake engineering symposium
- Cochran WG (1977) Sampling techniques, 3rd edn. Wiley, New York
- European Committee for Standardization (2004a) BS EN 1992-1-1:2004. Eurocode 2—design of concrete structures. Part 1-1: general rules and rules for buildings. British Standards Institution
- European Committee for Standardization (2004b) BS EN 1998-1:2004. Eurocode 8—design of structures for earthquake resistance. Part 1: general rules, seismic actions and rules for buildings. British Standards Institution
- Federal Emergency Management Agency (2005) Improvement of nonlinear static seismic analysis procedures, Report FEMA 440, Washington DC
- Federal Emergency Management Agency (2012) Seismic performance assessment of buildings, volume 1—methodology, Report FEMA P-58-1, Washington DC
- Goulet CA, Haselton CB, Mitrani-Reiser J, Beck JL, Deierlein GG, Porter KA, Stewart JP (2007) Evaluation of the seismic performance of a code-conforming reinforced concrete frame building—from seismic hazard to collapse safety and economic losses. *Earthq Eng Struct Dyn* 36(13):1973–1997. <https://doi.org/10.1002/eqe.694>

- Haselton CB, Deierlein GG (2007) Assessing seismic collapse safety of modern reinforced concrete moment frame buildings. *Report No. 156*. John A. Blume Earthquake Engineering Research Center, Stanford University
- 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 (2009) Evaluation of ground motion selection and modification methods: predicting median interstory drift of buildings. PEER Report 2009/01, Pacific Earthquake Engineering Centre
- Iervolino I, Cornell CA (2005) Record selection for nonlinear seismic analysis of structures. *Earthq Spectra* 21(3):685–713. <https://doi.org/10.1193/1.1990199>
- Iervolino I, Galasso C, Cosenza C (2010) REXEL: computer aided record selection for code-based seismic structural analysis. *Bull Earthq Eng* 8(2):339–362. <https://doi.org/10.1007/s10518-009-9146-1>
- International Code Council (2015) International building code. International Code Council, Inc., Country Club Hills, IL
- Jayaram N, Baker JW (2008) Statistical tests of the joint distribution of spectral acceleration values. *Bull Seismol Soc Am* 98(5):2231–2243. <https://doi.org/10.1785/0120070208>
- Jayaram N, Lin T, Baker JW (2011) A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. *Earthq Spectra* 27(3):797–815. <https://doi.org/10.1193/1.3608002>
- Katsanos EI, Sextos GS (2013) ISSARS: an integrated software environment for structure-specific earthquake ground motion selection. *Adv Eng Softw* 58:70–85. <https://doi.org/10.1016/j.advengsoft.2013.01.003>
- Kayhan AH, Korkmaz KA, Irfanoglu A (2011) Selecting and scaling real ground motions using a harmony search algorithm. *Soil Dyn Earthq Eng* 31(7):941–953. <https://doi.org/10.1016/j.soildyn.2011.02.009>
- Kottke A, Rathje EM (2008) A semi-automated procedure for selecting and scaling recorded earthquake motions for dynamic analysis. *Earthq Spectra* 24(4):911–932. <https://doi.org/10.1193/1.2985772>
- Lignos DG, Krawinkler H (2012a) Sidesway collapse of deteriorating structural systems under seismic excitations. Rep. No. 177, JA Blume Center, Department of Civil and Environmental Engineering, Stanford University
- Lignos DG, Krawinkler H (2012b) Development and utilization of structural component databases for performance-based earthquake engineering. *J Struct Eng Am Soc Civ Eng* 139(8):1382–1394. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0000646](https://doi.org/10.1061/(ASCE)ST.1943-541X.0000646)
- Lilliefors HW (1967) On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J Am Stat Assoc* 62:399–402
- Mahin SA, Lin J (1983) Construction of inelastic response spectra for single degree of freedom systems. UCB/EER-83/17, Earthquake Engineering Research Center, University of California at Berkeley
- Naem F, Alimoradi A, Pezeshk S (2004) Selection and scaling of ground motion time histories for structural design using genetic algorithms. *Earthq Spectra* 20(2):413–426. <https://doi.org/10.1193/1.1719028>
- Pacific Earthquake Engineering Research Center (2005) NGA Database. University of California at Berkeley
- Pacific Earthquake Engineering Research Center (2013) OpenSees software version 2.4.0. The Regents of the University of California. <http://opensees.berkeley.edu/index.php>. Accessed 4 July 2016
- Park YJ, Ang AH-S (1985) Mechanistic seismic damage model for reinforced concrete. *J Struct Eng Am Soc Civ Eng* 111(4):722–739. [https://doi.org/10.1061/\(ASCE\)0733-9445\(1985\)111:4\(722\)](https://doi.org/10.1061/(ASCE)0733-9445(1985)111:4(722))
- Shome N, Cornell CA, Bazzurro P, Carballo JE (1998) Earthquakes, records and nonlinear responses. *Earthq Spectra* 14(3):469–500. <https://doi.org/10.1193/1.1586011>
- Theophilou AI (2014) A ground motion selection and modification method suitable for probabilistic seismic assessment of building structures. Ph.D. Thesis, University of Surrey, UK
- Theophilou AI, Chryssanthopoulos MK, Kappos AJ (2017) A vector-valued ground motion intensity measure incorporating normalized spectral area. *Bull Earthq Eng* 15(1):249–270. <https://doi.org/10.1007/s10518-016-9959-7>
- Vamvatsikos D, Cornell CA (2002) Incremental dynamic analysis. *Earthq Eng Struct Dyn* 31(3):491–514. <https://doi.org/10.1002/eqe.141>
- Walpole PE, Myers RH, Myers SL, Ye K (2007) Probability and statistics for engineers and scientists. Pearson Prentice Hall, Upper Saddle River
- Wang G (2011) A ground motion selection and modification method capturing response spectrum characteristics and variability of scenario earthquakes. *Soil Dyn Earthq Eng* 31(4):611–625
- Watson-Lamprey JA, Abrahamson NA (2006) Selection of ground motion time series and limits on scaling. *Soil Dyn Earthq Eng* 26(5):477–482. <https://doi.org/10.1016/j.soildyn.2005.07.001>