EARTHQUAKE ENGINEERING AND STRUCTURAL DYNAMICS
Earthquake Engng Struct. Dyn. 2011; **40**:691–708
Published online 23 September 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/eqe.1055

Strong ground motion record selection for the reliable prediction of the mean seismic collapse capacity of a structure group

Mohsen Ghafory-Ashtiany[1], Mehdi Mousavi[1] and Alireza Azarbakht[2,*,†]

[1]*International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran*
[2]*Department of Civil Engineering, Faculty of Engineering, Arak University, Arak, Iran*

SUMMARY

How to select a limited number of strong ground motion records (SGMRs) is an important challenge for the seismic collapse capacity assessment of structures. The collapse capacity is considered as the ground motion intensity measure corresponding to the drift-related dynamic instability in the structural system. The goal of this paper is to select, from a general set of SGMRs, a small number of subsets such that each can be used for the reliable prediction of the mean collapse capacity of a particular group of structures, i.e. of single degree-of-freedom systems with a typical behaviour range. In order to achieve this goal, multivariate statistical analysis is first applied, to determine what degree of similarity exists between each selected small subset and the general set of SGMRs. Principal Component analysis is applied to identify the best way to group structures, resulting in a minimum number of SGMRs in a proposed subset. The structures were classified into six groups, and for each group a subset of eight SGMRs has been proposed. The methodology has been validated by analysing a first-mode-dominated three-storey-reinforced concrete structure by means of the proposed subsets, as well as the general set of SGMRs. The results of this analysis show that the mean seismic collapse capacity can be predicted by the proposed subsets with less dispersion than by the recently developed improved approach, which is based on scaling the response spectra of the records to match the conditional mean spectrum.

Received 4 March 2010; Revised 11 July 2010; Accepted 12 July 2010

KEY WORDS: strong ground motion record (SGMR); structure group; seismic collapse capacity; record selection; multivariate statistical analysis; principal component analysis

1. INTRODUCTION

The seismic collapse of a structural system is one of the most important causes for the loss of life during and after a severe ground motion. Hence, quantification of the collapse potential of existing and newly designed structures is an important issue in performance-based earthquake engineering. Incremental dynamic analysis (IDA) is a widely used method for the estimation of the seismic collapse capacity of structures [1]. This method requires non-linear response-history analyses of the specific structure for an appropriate set of SGMRs, each scaled to different intensity levels, to cover a wide range of the structural response from elastic behaviour to global instability on the basis of an engineering demand parameter (EDP), e.g. the maximum inter-storey drift ratio.

*Correspondence to: Alireza Azarbakht, Department of Civil Engineering, Faculty of Engineering, Arak University, Arak, Iran.
†E-mail: a-azarbakht@araku.ac.ir

The SGMRs which are used for non-linear response-history analysis are usually chosen based on magnitude, distance and site conditions [2]. As there is a lack of knowledge about future earthquakes, it would be reasonable to use a relatively large set of SGMRs for the IDA analysis of a specific structure. However, this target is not very realistic, so some approximate methods have recently emerged. The approximate methods for IDA analysis usually involve the replacement of non-linear response-history analysis by a combination of the pushover analysis of a multi-degree-of-freedom (MDOF) model and a non-linear response-history analysis of a single degree-of-freedom (SDOF) model [3–5].

One of the challenges in the non-linear response-history analysis of sophisticated structural models is how to select a limited number of SGMRs. The selection of appropriate SGMRs needs to be performed with the goal of accurately estimating the response of a structure to a specified ground motion Intensity Measure (IM), as measured by the spectral acceleration corresponding to the first mode period of the structure, $S_a(T_1)$. The current code-based method of record selection (e.g. [6]) is based on a consideration of the magnitude and distance of the SGMRs, while matching the mean response spectrum to the uniform hazard spectrum (UHS) as a target spectrum. Based on this approach, some methods have been developed for the selection and scaling of a set of near-optimal records from a large database [7, 8]. As no single earthquake is likely to produce a spectrum as high as the UHS spectrum, the code-based procedure for record selection is usually conservatively biased [9–11]. Reduction of this bias and the variance of the resulting structural response can be achieved by considering the spectral shape in the record selection [11]. It has been demonstrated that spectral shape can be indicated by epsilon, which is defined as a measure of the difference between the spectral acceleration of a record and the mean obtained from a ground motion prediction equation at a given period [12]. It can, therefore, be concluded that one method to account for the spectral shape effect is through the selection of a set of SGMRs that is specific to the structure's fundamental period and the site hazard characteristics [13]. This selection presents a significant challenge when assessing the seismic collapse capacity of a large number of structures or when developing a systematic procedure, since it implies the need to assemble specific ground motion sets for each structure. An alternative method has been proposed in [13], whereby a general set of SGMRs is used to simulate collapse, and the resulting collapse capacity is adjusted to take into account the spectral shape effects that are not reflected in the selection of the general set. The major difficulty of this method is that it implies the need to apply a relatively large number of ground motion records for the collapse assessment of the structures involved.

The main object of this paper is to find a proper solution for a reduction in the number of SGMRs needed for a non-linear response-history analysis. The proposed approach is intended to suggest a few subsets from a general set of SGMRs, and to use each subset for the collapse simulation of a specific structure group with a typical range of behaviour. The main criterion for selection of each subset is the similarity of the subset to the general set for the prediction of the mean collapse capacity of the structures belonging to the relevant structure group. It is needed to emphasize that the collapse capacity is considered in this study as the ground motion IM corresponding to the drift-related dynamic instability in the structural system. Other modes of collapse, such as column shear failures, are not considered in this study. It is important to note that suggestion of subsets is done just once, and that the selected subsets can be used *a priori* for the prediction of the collapse capacity of any arbitrarily selected first-mode-dominated structure.

The first mode period and ductility are the two most important structural parameters which are considered here for the grouping of structures. In this study, the above-mentioned structural systems were modelled by means of simplified SDOF mathematical models. The application of simplified models (e.g. SDOF systems) for the selection of suitable SGMRs for the analysis of sophisticated structures was dealt with in previous investigations [14, 15]. By considering reasonable combinations of the structural parameters, a bank of non-linear SDOF models has been established, representing different realistic structures. The collapse capacity assessment of the considered SDOF bank of models results in a collapse capacity matrix, with the use of an appropriate general set of SGMRs, for further statistical analysis. Grouping of the SDOF models, together with the selection of the associated SGMR subsets, is performed by PC ('Principal Components') analysis of the

DOI: 10.1002/eqe

collapse capacity matrix, and the application of a Genetic Algorithm (GA) to find the near-optimal subsets. The details of the proposed methodology are described in the following section.

2. METHODOLOGY

As the non-linear response-history analysis of a particular structure is a time-consuming task, the selection of a subset of SGMRs of smaller size which can reproduce the general SGMR set response is a valuable objective. The general SGMR set can be chosen from a catalogue [6]. The second set of SGMRs, which is hereafter called a subset, can be selected from the general SGMR set, but has a smaller number of SGMRs. Each general set or subset of SGMRs can be used to predict the mean value of the structural seismic collapse capacity, based on a characteristic probability density function. Thus, the similarity of the structural seismic collapse capacity response, based on the two different SGMR sets, can be measured by comparing the corresponding probability density functions. For this purpose, Figure 1 summarizes the main features of the proposed methodology as follows:

1. Select a general set of SGMRs which is suitable for collapse capacity analysis of structures.
2. Analyse different combinations of SDOF models for the general SGMRs set in order to provide an appropriate collapse capacity database for statistical analysis.
3. Apply a statistical measurement to compare the similarity of two sets of SGMRs for collapse capacity analysis of a given structure group.
4. Apply the PC analysis in order to identify the best way to group the structures that results in a minimum number of SGMRs in a proposed subset.
5. Use GA, as an efficient optimization method, to find the optimal subsets corresponding to the structure groups.
6. Use the corresponding subsets *a priori* for the engineering purposes.

Now, after presenting the three important issues in the proposed methodology, the complete procedure is implemented for a general set in Section 3.

1. A multivariate statistical approach is presented whose aim is to quantify the similarity of the general SGMR set to a given subset for the mean collapse capacity assessment of a structure group.
2. An optimization approach is used to select near-optimal subsets of SGMRs for any given structure group.

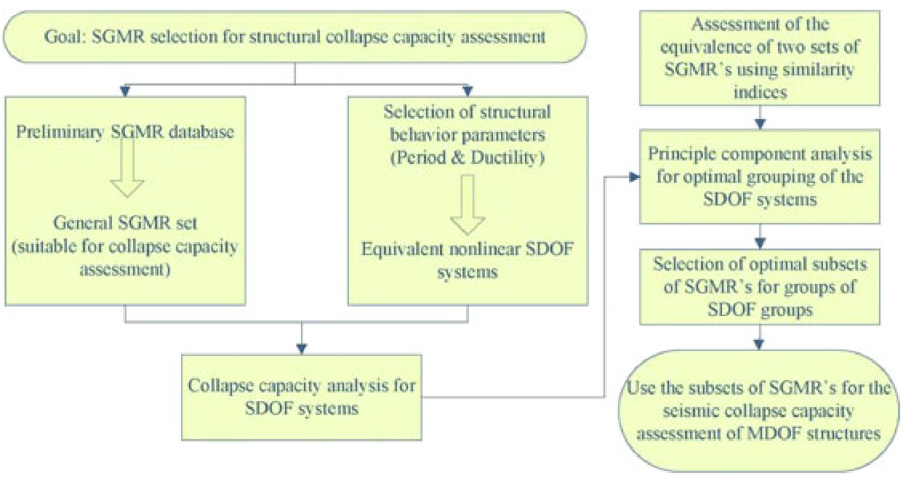

Figure 1. The main features of the proposed methodology.

DOI: 10.1002/eqe

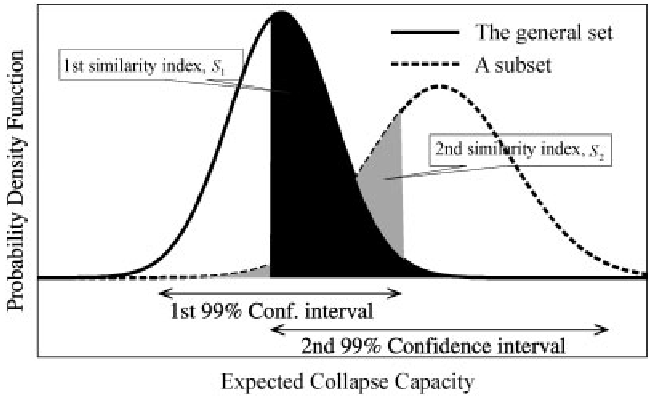

Figure 2. Definition of the two similarity indices for the univariate case.

3. As an increase in the size of a structure group results in a significant increase in the size of the associated subset of SGMRs, PC analysis is applied to limit this size.

2.1. Quantification of the similarity of two sets of SGMRs

Assume that a statistical sample of structural seismic collapse capacities in terms of an IM is available, based on a set of non-linear response-history analyses, for a particular SDOF system. Estimation of the distribution of the population mean and its confidence bounds can be performed by using the t-Student distribution function [16], by assuming a normal distribution for the logarithm of the seismic collapse capacity values. Assuming that $X_1, X_2, \ldots, X_n$ are samples from a normal population with the mean μ_x, the confidence interval for μ_x can be defined as [16]

$$\overline{X} - t_{n-1}(\alpha/2t)\frac{s_x}{\sqrt{n}} \leqslant \mu_x \leqslant \overline{X} + t_{n-1}(\alpha/2)\frac{s_x}{\sqrt{n}} \tag{1}$$

where $\bar{X}$, s_x are, respectively, the mean and the standard deviation of the samples, n is the size of the sample space and $t_{n-1}(\alpha/2)$ is the upper $100(\alpha/2)$th percentile of the t-distribution function with $n-1$ degrees of freedom. Now assume that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ are samples from a population which are called, respectively, the general set and the subset. Based on Equation (1), two plausible ranges can be defined for the population mean by considering a 99% confidence interval ($\alpha = 0.01$). The degree of similarity of the two mean values can be quantified by defining two similarity indices. The first index, S_1, is the probability that the population mean determined from the general set falls into the plausible range of the mean which is obtained from the subset. The second similarity index, S_2, is defined as the probability that the population mean determined from the subset falls into the plausible range of the mean obtained from the general set. The similarity indices are computed, as shown in Figure 2, by integrating the related probability density functions over the associated ranges.

The above-described approach for the assessment of the similarity of two sets of SGMRs can be extended to apply to an analysis of a set of SDOF systems categorized as a structure group. For this purpose, the univariate case, as shown in Figure 2, can be developed into the multivariate case by using the multivariate statistical analysis technique. Assume that a group of p SDOF systems is analysed for n SGMRs. The seismic collapse capacities of the structures are calculated, consequently, as $\langle X_1^1, X_1^2, \ldots, X_1^p \rangle$, $\langle X_2^1, X_2^2, \ldots, X_2^p \rangle$, $\ldots$, $\langle X_n^1, X_n^2, \ldots, X_n^p \rangle$ which define a p-dimensional statistical sample from a population with a $p \times 1$ mean vector μ_x and a $p \times p$ covariance matrix. A $100(1-\alpha)\%$ joint confidence interval for the mean of the population can be defined by the following Equation [17]:

$$n(\overline{X} - \mu_x)^{\mathrm{T}}(S_x)^{-1}(\overline{X} - \mu_x) < \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha) \tag{2}$$

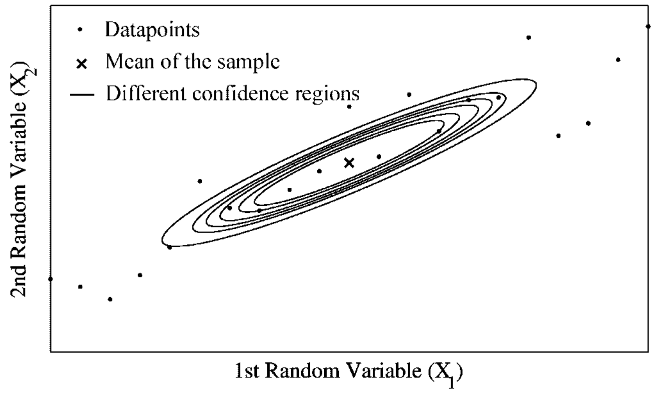

Figure 3. The different confidence regions for a two-dimensional statistical sample.

where $\overline{X}$, S_x are, respectively, the mean vector and the covariance matrix of the sample. $F_{p,n-p}$ denotes a random variable with an F-distribution with p and $n-p$ degrees of freedom. The designation $()^{\mathrm{T}}$ indicates a transposed matrix or vector. Some aspects of the multivariate analysis can be reflected by plotting the associated confidence regions. As an example, the different confidence regions are shown schematically, for a given $p=2$ dimensional sample, in Figure 3.

The orientation of the confidence ellipses is proof of the correlation between the variables, i.e. between the structural seismic collapse capacities. So it is a reasonable inference that the p-correlated variables must be analysed jointly, which is a key aspect of multivariate statistical analysis [17]. The similarity of the two sets of SGMRs for a structure group of size p can be quantified by extending the previously defined similarity indices to the multivariate case. For this purpose, it is necessary to evaluate the joint probability density function of the mean of the structure group capacity for each of the SGMR sets, and then calculate the p-dimensional volume below each function with a 99% confidence region of the other set.

As a well-known approach, the similarity of the general set mean with the subset mean can be evaluated by applying the standard statistical hypothesis test [16]. The null hypothesis is defined as the equality of the two means, and the alternative test defines the difference between them. This test is generally performed to a characteristic significant level, i.e. 0.01. This significant level is generally interpreted as the first type of error which is defined as the probability of the false rejection of the null hypothesis. This hypothesis test has enough strength for the rejection of the null hypothesis, but cannot systematically quantify the similarity of two means. This imperfection can be resolved by use of the second type of error. For this purpose, a reasonable value for the difference of the means can be defined as an alternative hypothesis, and consequently the probability of the false acceptance of the null hypothesis can be calculated. The optimal subsets can be found by minimizing the second type of error. The similarity indices that are defined in this paper are different from the above-mentioned types of error, but still have a common terminology and may be comparable. The advantage of the similarity indices which are used in this paper is that there is no need to define a reasonable value for the difference of the means which makes them convenient for the similarity assessment of the means.

2.2. Selection of a near-optimal SGMR subset

The near-optimal subset of SGMRs can be selected from the general SGMR set by maximizing the similarity indices. Here, GA, as an efficient optimization search algorithm, is applied [18, 19]. The advantage of GA is that there is a high probability that the objective function converges to near the global optimum of the problem, and not to a local optimum, which is not always the case for gradient-based searching techniques [18].

Generally, by increasing the size of the subset, the similarity indices will increase. On the other hand, by limiting the minimum value of the similarity indices to 90%, the minimum size of the

DOI: 10.1002/eqe

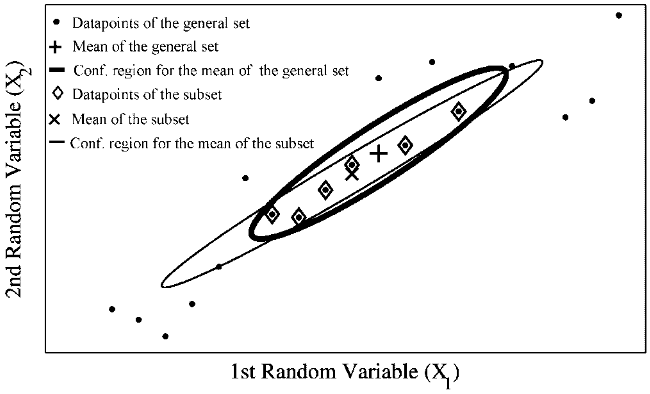

Figure 4. The mean values and corresponding confidence intervals for the general set and a subset of SGMRs for the two-dimensional problem.

subset can be calculated. For example, for the dataset used in Figure 4, the minimum number of SGMRs in the subset that guarantees similarity indices of at least 90% is 6. The selected subset and the corresponding confidence region are shown in Figure 4.

It follows from Equation (2) that, for a structure group of size$p$, at least $(p+1)$ records are needed for the subset. Hence, the size of the subset will increase significantly when the size of the structure group is increased. In the following section, a statistical solution is proposed in order to obtain a solution to this problem.

2.3. Reduction in the size of the SGMR subset by PC analysis

One of the most important inherent difficulties in the above-mentioned procedure for the selection of the SGMR subset for a structure group is the growth in the size of the SGMR subset in the case of an increase in the size of the structure group. Dimension reduction, which is a key feature of multivariate statistical analysis, can be performed in order to try to find a solution to this problem. Whether or not dimension reduction can be effectively applied depends strongly on the variance–covariance arrangement of the considered data. For example, if the response of two structures is highly correlated, it is possible to reduce the dimensions of the problem from two to one. Such a dimension reduction leads to a reduction in the number of SGMRs needed in the subset. On the other hand, the possibility of reducing the number of SGMRs will decrease if the considered correlation is low.

The PC analysis is a systematic approach which can be used to explain the variance–covariance arrangement of p random variables, and to reduce the problem dimension [17]. The PC analysis method transforms the current set of variables into a new set, called PCs. The PCs are weighted combinations of the original variables, and represent the same data in an aggregated form. Some of the aspects of PC analysis are summarized below.

1. Each PC is a linear combination of the original variables, and all the PCs are orthogonal to one another.
2. Mathematically, PC analysis is equivalent to eigen-value analysis of the covariance of the data matrix.
3. The first few PCs usually represent the majority of the total variation in the original variables.
4. The importance of each PC is measured based on its respective variation. The problem dimension can be reduced by neglecting the less important PCs, with the least amount missing from the inherent information.

The role of PC analysis in the reduction of the subset size is shown in Figure 5. A high correlation between the two variables, as seen in Figure 5(a), makes it possible to reduce the dimensions of the data-set. The transformed data is shown in Figure 5(b), which confirms the independence of the PCs. The first PC, as seen in Figure 5(c), contains 95% of the total variation in the data, so

DOI: 10.1002/eqe

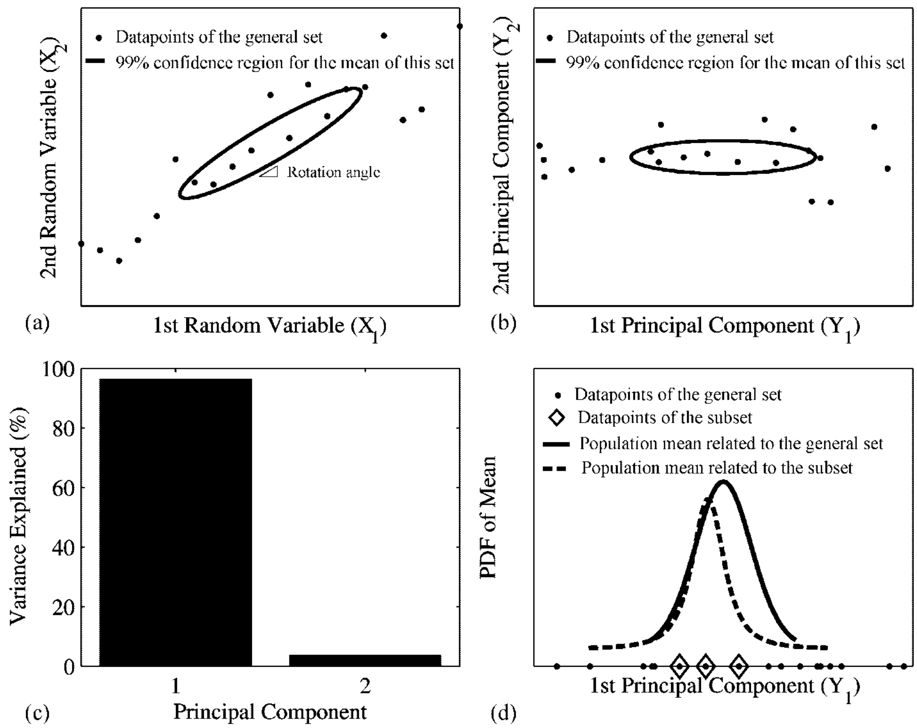

Figure 5. The role of PC analysis in reduction of subset size: (a) a high correlation between the two variables can be observed; (b) the transformed data confirm the independency of the PCs; (c) the first PC contains 95% of the required information and (d) there is good agreement between the probability density functions corresponding to the general and the subset.

that neglecting the second PC is a rational decision. In this case, only three data points are needed to supply the minimum 90% of the similarity indices (compared to the six SGMRs, as presented in Section 2.2). The similarity indices, as well as the probability density functions, are presented, for the general set and for the subset of SGMRs, in Figure 5(d). The good agreement between the probability density functions, associated with the general set and the near-optimal subset, confirms that the use of the first PC results in a reduction in the minimum size of the subset from 6 to 3.

The reduction in the problem dimensions will clearly result in the minimum size of the subset of SGMRs. On the other hand, the missing inherent information of the initial data, which occurs if the higher PCs are neglected, should be limited. The minimum number of PCs which are needed to provide at least 90% of the total variation from the general set has been considered in this study. It is clear that due to this condition, a smaller number of considered PCs for a structure group results in a corresponding subset of smaller size. For this reason PC analysis can provide valuable information about the minimum size of a subset corresponding to a given structure group. As a result, PC analysis is an efficient tool which can be used to identify the best pattern for grouping structures, which results in the minimum size of the proposed subsets of SGMRs.

3. SELECTION OF SGMR SUBSETS FROM THE GENERAL SET USED IN THE ATC63 PROJECT

The proposed methodology was applied to a specific far-field set consisting of 44 SGMRs. This general set was used in the Applied Technology Council Project (FEMA 2008) as a procedure to validate the provisions for seismic structural design [20]. Detailed information about the SGMRs, as well as the complete list of selection criteria, are provided in [21]. Figure 6 shows the

DOI: 10.1002/eqe

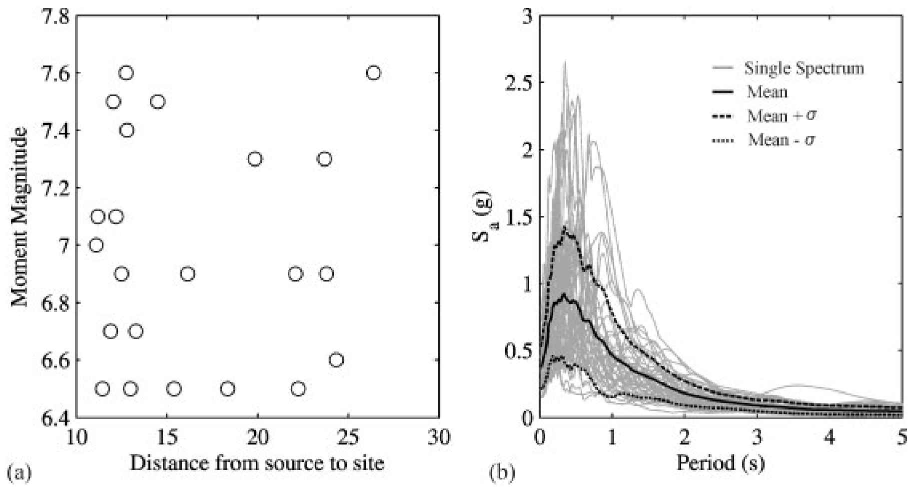

Figure 6. (a) The magnitude–distance distribution and (b) the 5% damped acceleration response spectra of the general set of SGMRs.

magnitude–distance distribution, as well as the 5% damped acceleration response spectra of the general set.

As already mentioned, the main aim of this paper is to suggest a few small subsets from this general set for the analysis of different structure groups. Also, in order to validate the independence of the grouping of structures from the applied set of SGMRs, another set of SGMRs was used to study this issue. The second set of SGMRs includes 192 records, and is documented in [22].

3.1. Seismic collapse capacity database for SDOF systems

A seismic collapse capacity database was established for 84 SDOF systems with periods ranging between 0.1 and 2.0 s ($T = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.25, 1.5, 1.75 and 2.0), six ductility values ($\mu_f = 2$, 4, 6, 8, 10 and 12) and a mass proportional critical damping ratio equal to 5%. A tri-linear backbone curve model with a zero hardening slope ($\alpha_s = 0$) and fixed capping ductility ($\mu_f = 0.9\mu_f$), as shown in Figure 7(a), was selected. $P-\Delta$ effects and cyclic deterioration were neglected for the purpose of simplicity, although these non-linear effects could have an influence on the seismic collapse capacity [23].

Based on an analysis of the seismic collapse capacity of 14×6 SDOF systems, and taking into account 44 records of the general set, an 84×44 matrix of collapse capacity data was created for the statistical analysis. The seismic collapse capacity of a particular SDOF system was determined using the Hunt and Fill tracing algorithm [1] for each of the 44 SGMRs. $S_a(T, 5\%)$ was used as the IM for this seismic collapse capacity assessment. The collapse capacity tracing procedure is presented in Figure 7(a) for an arbitrary SDOF system and for the 44 SGMRs. The application of the Kolmogorov–Smirnov [16] test to the collapse capacity values corresponding to each SDOF system, as shown in Figure 7(b), confirmed the goodness of fit of a normal distribution to the logarithm of the collapse capacity IM values. The minimum, average and maximum p-values for the considered SDOF systems in the Kolmogorov–Smirnov test were equal to 0.17, 0.67 and 0.99, respectively. The p-value indicates the smallest level of significance that would lead to rejection of the null hypothesis, i.e. fitness of the normal distribution to the collapse capacity values of a considered SDOF. It would take p-value lower than 0.05 to cast doubt on the null hypothesis.

3.2. Analysis of different patterns for the grouping of structures

First, the SDOF systems were grouped, based on their periods, with varying ductility (μ_f). The SDOF seismic collapse capacity response can be summarized in a 44×6 matrix for each SDOF system with a particular period value, which reflects the seismic collapse capacity IMs

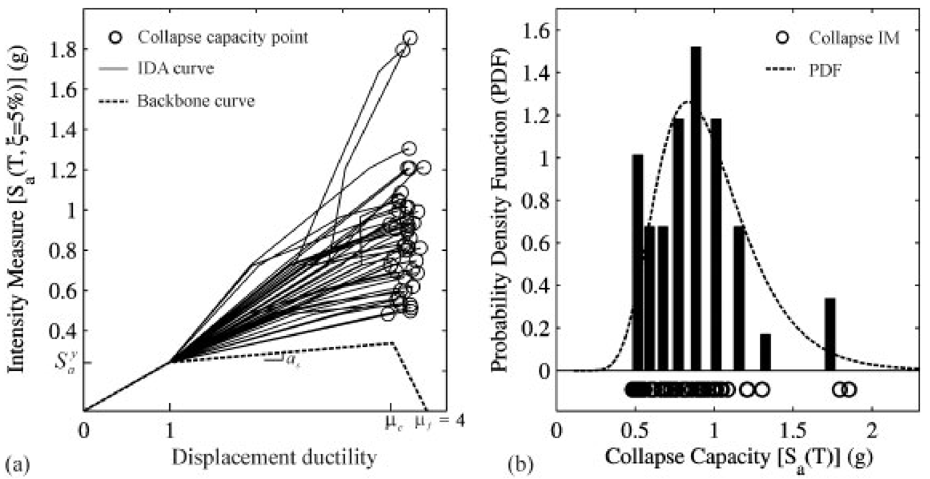

Figure 7. (a) The seismic collapse capacity points in the IDA curves for an SDOF system and 44 SGMRs and (b) the log-normal curve fitted to the probability density function of the collapse capacity points.

corresponding to the 44 SGMRs and six ductility values. The results of the PC analysis confirmed that the first PC is dominant, which means that the first PC includes a considerable proportion of the inherent information located in the related data matrix. Figure 8 shows the results for the group of structures with a period of 0.9 s. The IMs corresponding to the SDOF seismic collapse capacity for two different ductility values (4 and 10) are shown in Figure 8(a). A high correlation can be observed in the response, which confirms the high influence of the first PC, shown in Figure 8(b). The first and the second PCs contain, respectively, 80 and 15% of the total variation. By considering the first two PCs for the near-optimal record selection, as shown in Figure 8(c), only six SGMRs are required to satisfy the similarity indices criterion (at least 90%). However, by using only the first PC, three SGMRs can be found which provide similarity indices of up to 98% (Figure 8(d)). On the other hand, if the first three PCs are taken into account, at least 16 records should be used to satisfy the similarity indices criterion. This trend shows that taking the upper PCs into account may result in a substantial increase in the minimum size of the corresponding SGMR subset.

The percentage of the total variance for the first three PCs is summarized in Table I for 14 different groups of structures, each group containing structures with the same period value. The results, which are shown in the second column of Table I, indicate the minimum number of SGMRs that need to be used for the prediction of the IM corresponding to the seismic collapse capacity with a high level of similarity indices and independent of the ductility values. It can be concluded that, for an acceptable value of the similarity indices (e.g. 90%), only three to eight SGMRs are needed for the reliable assessment of the mean IM of the seismic collapse capacity of all structures having the same period, independent of their ductility values.

In order to estimate the possibility of grouping structures together, based on their ductility, the 84 SDOF systems were grouped into six categories. A 44×14 matrix contains the seismic collapse capacity IMs corresponding to the 44 SGMRs and 14 period values for each of the six groups of structures. The results of the PC analysis for the different groups of structures, with ductilities equal to 4 and 10, as shown in Figure 9, indicate that the variance fraction of the first PC is less than 40% of the total variance. Also, at least eight PCs need to be taken into account in order to cover 90% of the total variation. This grouping pattern is not convenient for reducing the required number of SGMRs. In other words, the estimation of the mean IM of the seismic collapse capacity by a limited number of SGMRs depends strongly on the period of the structure. For clarification, the three to eight SGMRs that were proposed for each of the period-based structure groups need to be recalled. A second similarity index, corresponding to each of the 14 structure groups, was calculated, as shown in Table II, based on all of the subsets in Table I, in order to determine

DOI: 10.1002/eqe

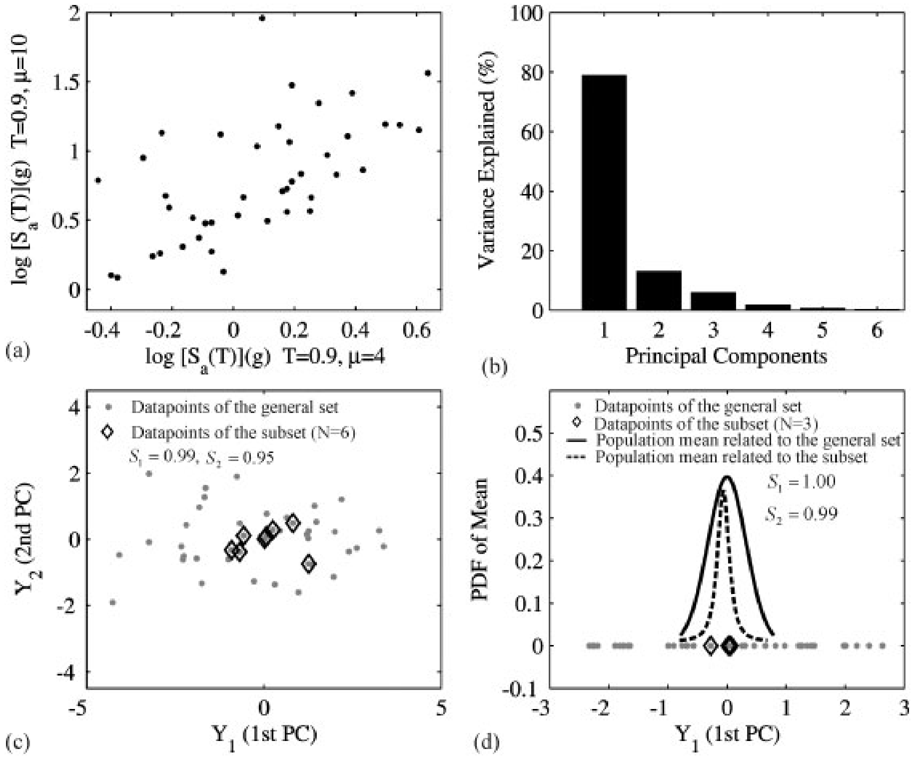

Figure 8. The subset of SGMRs for structure groups with similar periods and different ductility values: (a) high correlation can be observed between the structural seismic collapse capacities; (b) the contribution of different PCs to the period-based grouping; (c) the subset of SGMRs based on the two first PCs; and (d) the subset of SGMRs based on the first PC.

Table I. The relationship between the considered PCs and the minimum number of required SGMRs.

	Variation explained by the PCs (%)			Minimum size of the subsets		
Period	First PC	Second PC	Third PC	Use of 1 PC	Use of 2 PCs	Use of 3 PCs
0.1	94.7	4.3	0.8	3	6	14
0.2	93.9	4.7	1.1	3	8	16
0.3	91.1	6.2	1.9	3	6	16
0.4	86.7	9.4	2.9	3	6	16
0.5	90.1	6.8	1.9	3	6	14
0.6	84.6	10.8	3.5	3	6	18
0.7	91.2	4.8	2.6	3	6	16
0.8	83.1	9.2	5.7	3	6	18
0.9	78.8	12.9	5.8	3	6	16
1.0	74.4	18.1	4.7	3	6	14
1.25	86.2	8.8	4.1	3	6	14
1.50	88.3	8.0	2.8	3	6	16
1.75	85.4	11.2	2.3	3	6	14
2.0	84.1	10.7	3.9	3	8	18

whether or not a subset of records is convenient for the other groups of structures. The results, as seen in Table II, confirm that the selection of SGMRs strongly depends on the period of the structure.

For engineering applications, the possibility of reducing the dependency of the record selection process on the period of structures has been studied. For this objective, a third grouping pattern

DOI: 10.1002/eqe

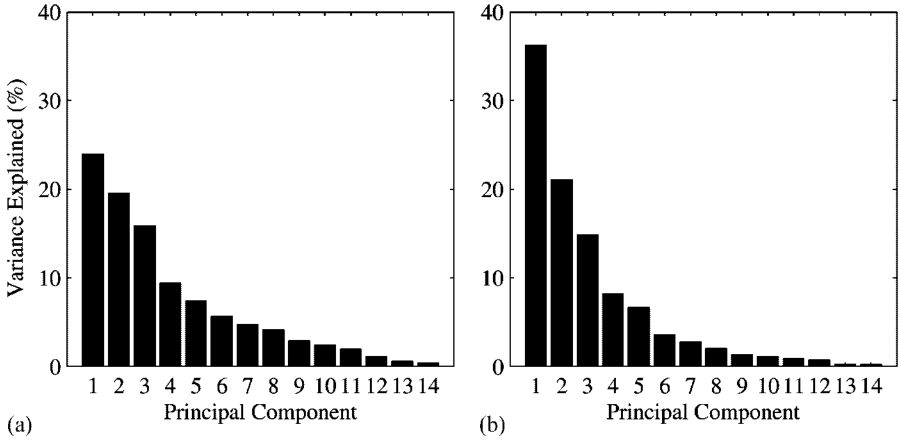

Figure 9. The contribution of individual PCs in the ductility-based grouping for (a) the structure group with $\mu_f = 4$ and (b) the structure group with $\mu_f = 10$.

Table II. The minimum similarity indices achieved by using the period-based subset of SGMRs for the different structure groups.

	Period-based subsets													
	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14
T1	0.93	0.43	0.48	0.21	0.53	0.37	0.28	0.64	0.62	0.73	0.31	0.20	0.16	0.36
T2	0.30	0.95	0.51	0.30	0.24	0.22	0.08	0.17	0.11	0.12	0.49	0.53	0.38	0.14
T3	0.61	0.25	0.92	0.08	0.45	0.19	0.05	0.35	0.45	0.61	0.41	0.58	0.31	0.44
T4	0.11	0.02	0.53	0.92	0.54	0.37	0.10	0.25	0.15	0.11	0.07	0.21	0.36	0.24
T5	0.33	0.17	0.08	0.42	0.92	0.21	0.27	0.43	0.24	0.27	0.47	0.40	0.18	0.44
T6	0.06	0.10	0.13	0.05	0.55	0.91	0.58	0.09	0.14	0.03	0.11	0.31	0.11	0.47
T7	0.30	0.15	0.30	0.32	0.22	0.38	0.93	0.49	0.18	0.26	0.28	0.37	0.31	0.22
T8	0.14	0.13	0.40	0.18	0.23	0.44	0.28	0.91	0.41	0.48	0.12	0.02	0.02	0.00
T9	0.27	0.60	0.43	0.29	0.38	0.11	0.28	0.54	0.92	0.73	0.38	0.36	0.10	0.16
T10	0.45	0.35	0.20	0.00	0.20	0.23	0.15	0.02	0.19	0.90	0.31	0.19	0.20	0.23
T11	0.10	0.31	0.31	0.39	0.51	0.25	0.13	0.44	0.47	0.15	0.93	0.55	0.32	0.22
T12	0.36	0.03	0.05	0.24	0.06	0.51	0.19	0.24	0.11	0.26	0.17	0.90	0.26	0.28
T13	0.31	0.14	0.24	0.34	0.28	0.36	0.29	0.20	0.14	0.21	0.36	0.56	0.92	0.54
T14	0.26	0.23	0.50	0.10	0.40	0.59	0.59	0.17	0.25	0.46	0.45	0.65	0.55	0.91

was considered, based on merging of the neighbouring period-based groups of structures. The criterion for this re-grouping was that the first and second PCs of each group cover at least 90% of the total variation, in order to limit the number of SGMRs needed for each group. As illustrated in Table III, based on the results of these analyses, all of the 84 SDOF systems were grouped into six discrete groups, and eight near-optimal records were found for each group. Table IV shows the characteristics of the SGMRs which are proposed for each of the groups described in Table III. In the next section, the proposed SGMR subset will be applied to estimate the mean IM of the seismic collapse capacity of a test structure, in order to investigate the efficiency of the proposed methodology.

The dependency of the SDOF systems grouping procedure on the general set selection has been investigated using a second general set, which was introduced in the previous section [22]. The analyses were repeated for the second general set of SGMRs, and the results are presented in Table V. The results confirm that the grouping trend is not affected by changing the general SGMR set.

DOI: 10.1002/eqe

Table III. The near-optimal SGMRs for different period ranges.

Ground motion subset	Period	Variance explained by the first PC	Variance explained by the second PC	SGM's ID	First similarity index	Second similarity index
I	0.1–0.3	59.6%	26.1%	3–8–14–20–21–24–27–28	0.99	0.92
II	0.3–0.5	65.4%	23.7%	2–4–10–12–20–21–23–30	0.99	0.93
III	0.5–0.7	70.9%	18.8%	1–4–6–10–12–15–17–23	0.99	0.95
IV	0.7–0.9	76.8%	12.3%	1–4–12–22–23–24–25–26	0.99	0.95
V	0.9–1.25	64.3%	24.3%	8–9–12–15–16–22–23–29	0.99	0.95
VI	1.25–2.0	69.8%	19.8%	5–7–13–15–19–23–28–31	0.99	0.92

Table IV. The SGMRs selected from the general set.

Event, Mw, Year	ID	Station, Dir	Vs_30 (m/s)	Campbell distance (km)	Joyner-Boore dist. (km)	PGA (g)
Northridge, 6.7, 94	1	W Lost Cany, 000	309	12.4	11.4	0.41
	2	W Lost Cany, 270	309	12.4	11.4	0.48
Hector Mine, 7.1, 99	3	Hector, 000	685	12	10.4	0.27
Imperial Valley, 6.5, 79	4	Delta, 262	275	22.5	22	0.24
	5	Delta, 352	275	22.5	22	0.35
	6	El Centro Array #11, 140	196	13.5	12.5	0.36
	7	El Centro Array #11, 230	196	13.5	12.5	0.38
Kobe, Japan, 6.9, 95	8	Nishi-Akashi, 090	609	25.2	7.1	0.50
	9	Shin-Osaka, 000	256	28.5	19.1	0.24
Kocaeli, Turkey, 7.5, 99	10	Duzce, 180	276	15.4	13.6	0.31
	11	Duzce, 270	276	15.4	13.6	0.36
	12	Arcelik, 000	523	13.5	10.6	0.22
Landers, 7.3, 1992	13	Yermo Fire Station, 270	354	23.8	23.6	0.24
	14	Yermo Fire Station, 360	354	23.8	23.6	0.15
	15	Coolwater, LN	271	20	19.7	0.28
	16	Coolwater, TR	271	20	19.7	0.42
Loma Prieta, 6.9, 89	17	Capitola, 000	289	35.5	8.7	0.53
	18	Capitola, 090	289	35.5	8.7	0.44
	19	Gilroy Array #3, 000	350	12.8	12.2	0.56
Manjil, Iran, 7.4, 90	20	Abbar, T	724	13	12.6	0.50
Superstition Hills, 6.5, 87	21	El Centro Imp. Co. Cent, 000	192	18.5	18.2	0.36
	22	Poe Road (temp), 270	208	11.7	11.2	0.45
	23	Poe Road (temp), 360	208	11.7	11.2	0.30
Cape Mendocino, 7.0, 92	24	Rio Dell Overpass—FF, 270	312	14.3	7.9	0.39
	25	Rio Dell Overpass—FF, 360	312	14.3	7.9	0.55
Chi-Chi, Taiwan, 7.6, 99	26	CHY101, E	259	15.5	10	0.35
	27	TCU045, E	705	26.8	26	0.47
	28	TCU045, N	705	26.8	26	0.51
San Fernando, 6.6, 71	29	LA—Hollywood Stor FF, 090	316	25.9	22.8	0.21
	30	LA—Hollywood Stor FF, 180	316	25.9	22.8	0.17
Friuli, Italy, 6.5, 76	31	Tolmezzo, 000	425	15.8	15	0.35

4. APPLICATION OF THE PROPOSED SGMR SUBSETS FOR THE ANALYSIS OF A MDOF TEST STRUCTURE

The test structure, as shown in Figure 10, is a three-storey asymmetric-reinforced concrete frame, for which a pseudo-dynamic experiment was performed at full scale at the ELSA Laboratory [24]. The post-test mathematical model [25] generated within the OpenSees program [26] was used for the analyses in this study. For simplicity, the non-linear dynamic analyses were performed by subjecting the structure to seismic loads in the weak X direction only (Figure 10). The period of the dominant mode shape was 0.85 s. The EDP is defined as the maximum inter-storey drift ratio of

DOI: 10.1002/eqe

Table V. The results of the PC analysis for SDOF groups based on the second general set of SGMRs.

Ground motion subset	Period	Variance explained by the first PC	Variance explained by the second PC	First similarity index	Second similarity index
I	0.1–0.3	69.3%	20.6%	0.99	0.95
II	0.3–0.5	68.4%	20.9%	0.98	0.92
III	0.5–0.7	75.4%	14.7%	0.99	0.95
IV	0.7–0.9	79.1%	11.5%	0.99	0.95
V	0.9–1.25	76.2%	14.2%	0.99	0.88
VI	1.25–2.0	83.1%	10.2%	0.99	0.94

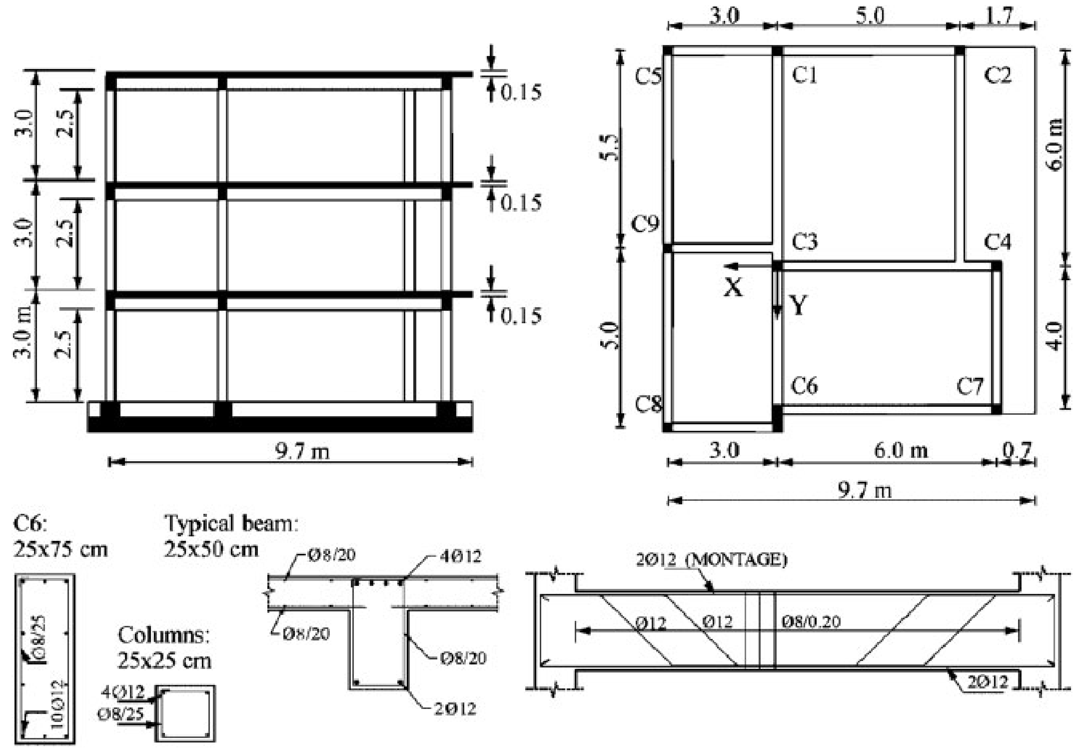

Figure 10. The elevation (upper left) and plan (upper right) view of the test structure, showing typical reinforcement details (bottom).

the structure. The collapse is considered here as the dynamic sidesway instability in one or several storeys of the structural system. Other modes of collapse, such as loss of vertical load-carrying capacity in individual structural components or progressive collapse, are not considered because of the inadequacy of the structural models to capture such modes.

Figure 11(a) shows the EDP versus IM and the corresponding seismic collapse capacity points of the structure obtained from the IDA analysis, using the 44 SGMRs. The implementation of the Kolmogorov–Smirnov test to the collapse capacity values of this structure, as shown in Figure 11(b), confirmed the goodness of fit of the normal distribution to the logarithm of the collapse capacity IM values. The p-value for this test was 0.91.

4.1. *Efficiency of the proposed SGMR subset*

The first mode period of the test structure implies that subset IV of SGMRs, as described in Table III, would be appropriate for the assessment of the mean IM of the seismic collapse capacity. In order to

DOI: 10.1002/eqe

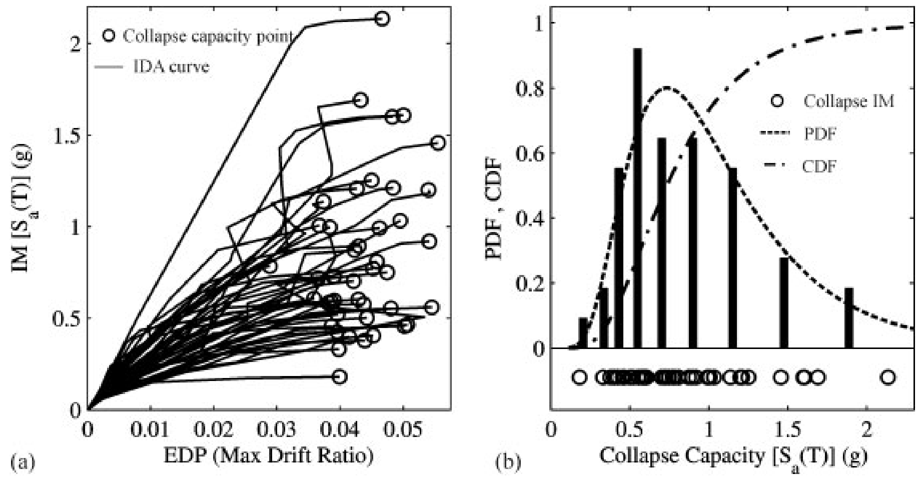

Figure 11. (a) The collapse capacity tracing algorithm for the test structure and (b) the log-normal curves fitted to the probability density and probability distribution functions of the collapse capacity IM values.

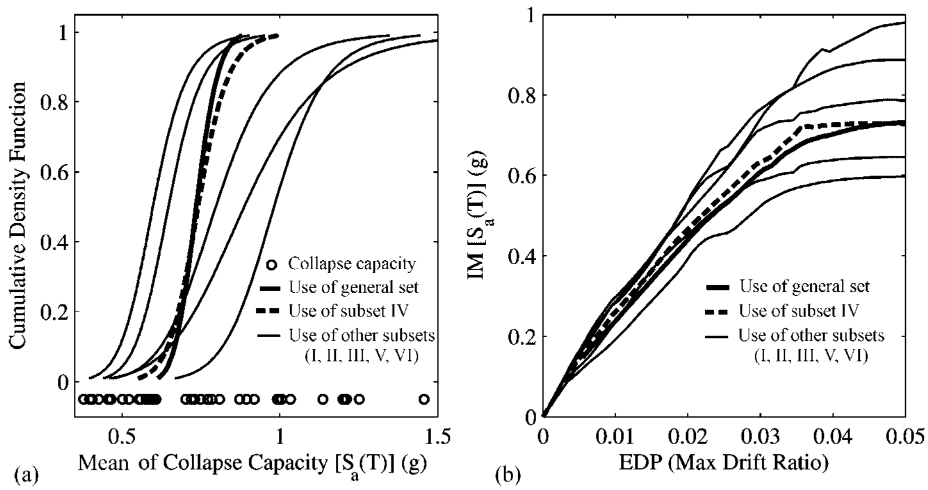

Figure 12. Efficiency assessment of the proposed subsets: (a) the CDF of the mean value of the seismic collapse capacity and (b) the median of the full IDA analysis using different proposed subsets and the general set.

evaluate the efficiency of the proposed subset records, the CDF of the mean IM value of the seismic collapse capacity based on the general SGMR set, as well as the proposed SGMR subsets (see Table III), are shown in Figure 12(a). The CDF for the SGMR subset IV, as shown in Figure 12(a), shows good agreement with the general records ($S_1 = 0.99$, $S_2 = 0.95$), whereas the other subsets could not produce a good estimate for the mean IM of the seismic collapse capacity. Table VI shows the mean values as well as the 99% confidence interval of the IM of the seismic collapse capacity according to each subset, and the associated similarity indices. The median IDA curve based on the general set and on each of the investigated SGMR subsets are shown in Figure 12(b). Good agreement can be seen between the median IDA curve based on the general SGMRs and the median IDA curve obtained from subset IV. This result is due to this fact that a large range of ductility values have been used for the selection of the appropriate SGMR set for each structure group.

In order to compare the efficiency of the proposed methodology with the typical spectral-based record selection procedure, the conditional mean spectrum (CMS) [27] was used to select an

Table VI. The accuracy achieved when using the proposed records for the estimation of the seismic collapse capacity of the test structure compared to the other subsets.

Ground motion class	Expected value of seismic collapse capacity	99% confidence interval for seismic collapse capacity	First similarity index	Second similarity index
All records	0.73	0.61–0.88	—	—
I	0.60	0.40–0.90	0.99	0.48
II	0.65	0.44–0.95	0.99	0.69
III	0.80	0.47–1.35	0.99	0.67
IV	0.74	0.55–1.00	0.99	0.92
V	0.98	0.67–1.45	0.98	0.26
VI	0.89	0.46–1.71	0.99	0.47

alternative subset for analysis of the test structure. The target epsilon for the determination of the CMS was chosen to be equal to the mean epsilon value of the general set. The individual and mean epsilon values of the records of the general set, obtained by using [28], are shown, versus period, in Figure 13(a). The mean value of epsilon at $T=0.85\,\text{s}$ is equal to 0.24, as can be seen in Figure 13(b). By considering an idealized site, where the ground motion hazard is dominated by a single characteristic event of magnitude 7.0, a distance to fault of 10 km, and site soil condition corresponding to $V_{s30}=360\,\text{m/s}$, the mean spectrum was determined by using [28], and the CMS associated with this spectrum and conditioned to the above-mentioned epsilon mean value was calculated [27]. The mean spectrum, the mean spectrum plus 0.24 standard deviation and the CMS have been plotted in Figure 13(c). The eight SGMRs, which are compatible with the CMS spectrum, were calculated using the GA optimization technique [7] and are shown in Figure 13(d). The criteria for these eight SGMRs were (1) consistency with the CMS spectrum and (2) minimum dispersion in the period range between $0.2T$ and $2.0T$. The second similarity index (S_2) for the resulting subset is 0.75 (compared with the subset suggested in this paper, which provides an S_2 value of 0.92). The CMS-based selection procedure was repeated for different numbers of SGMRs, and the S_2 values are shown in Figure 14(a). It is clear that, as the size of subset increases, the value of S_2 also increases. A selection of 15 records, based on Figure 14(a), resulted in a value of S_2 equal to 0.91. It can thus be claimed that the use of the proposed subset, containing eight records, is equal to the use of 15 CMS-based selected records. To further demonstrate the benefits of the proposed methodology, the random selection approach was taken into account. As shown in Figure 14(b), a large number of random subsets of size 8 were selected, and the S_2 values associated with them were determined. The median value of the resulting S_2 is 0.58, with a dispersion of 0.2, which is a lot less than the value obtained by using the CMS method or the proposed methodology. This simulation was repeated for different numbers of SGMRs, and the median and dispersion of the S_2 values were determined, as shown in Figure 14(c). The selection of 28 random records resulted in a median S_2 value equal to 0.94, and a dispersion of 0.05. It can thus be tentatively claimed that use of the eight proposed subsets is equivalent to the use of 15 CMS-based selected records, or the use of 28 randomly selected records.

4.2. Discussion of the results

Reliable assessment of the mean IM of the seismic collapse capacity of the three-storey concrete test-structure subjected to 44 SGMRs requires about 37 h of computational effort, with a 2.8 GHz Pentium 4 processor, and 1 GB DDR RAM, which is not convenient for engineering design purposes. The significantly smaller subset of records shown in Table III can be used, instead of the general set which is based on the first mode period of the structure. The use of subset IV for the test-structure provides an average of 92% for the second similarity index. This similarity index indicates that the application of subset IV results in a low bias and low dispersion in the mean IM of the seismic collapse capacity assessment. On the other hand, if the CMS method is used for the selection of different compatible subsets, the resulting second similarity index is 0.75. This lower

DOI: 10.1002/eqe

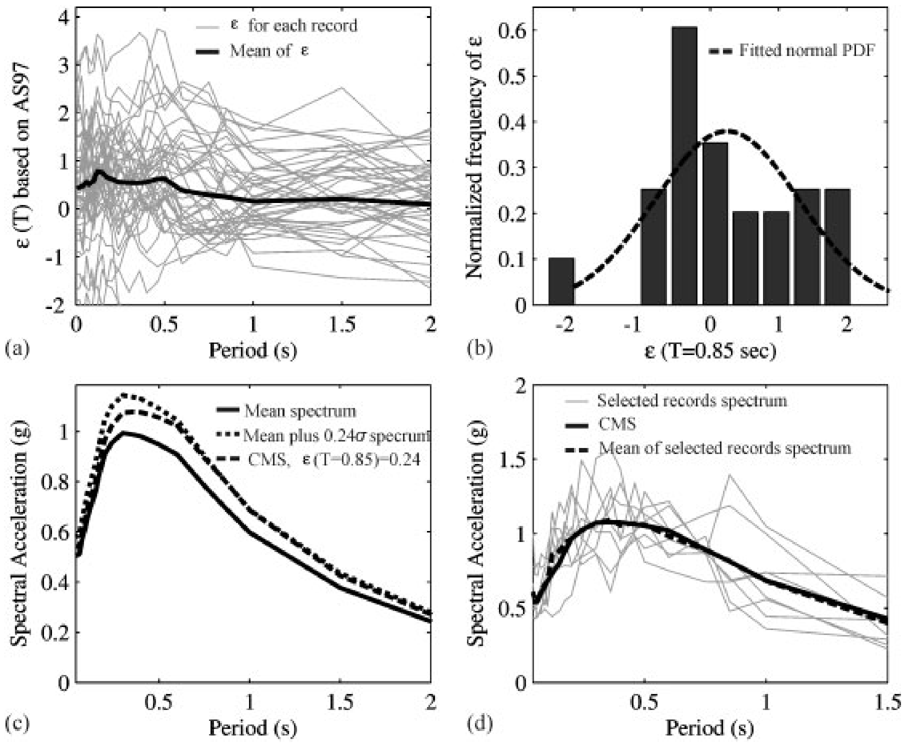

Figure 13. The code-based selection of SGMRs: (a) the individual and the mean of epsilon versus period based on the Abrahamson-Silva 97 [28] attenuation relationship; (b) the probability density function of epsilon at $T=0.85$ s; (c) the mean spectrum based on the general set of SGMRs as well as on the CMS spectra and (d) the eight SGMRs which are compatible with the CMS spectrum.

value of the median of S_2 demonstrates the greater reliability of the proposed subset compared to the CMS method. It has been shown that, if results of similar reliability are to be achieved, it is necessary to select 18 records based on the CMS method. The results also show that this efficiency can be accomplished by random selection when the number of selected records is 28.

It should be noted that the proposed subsets were determined without any consideration of the hazard level. As a solution for this problem, the mean IM of the seismic collapse capacity resulting from the application of a suitable subset should be modified by use of the procedure proposed in [13] to meet the target hazard level. This procedure is based on an adjustment of the difference between the target epsilon value (as an indicator of the hazard level) and the mean epsilon value of the general set using an empirical relationship.

5. CONCLUSIONS

A new approach has been proposed in this paper for the categorization of SDOF systems into groups, and then a subset of SGMRs has been proposed for each of them. All of the presented subsets were chosen from a general set of SGMRs, and each of them is equivalent to the general set for the assessment of the mean IM of the seismic collapse capacity of the associated SDOF group. The grouping features are period and ductility. For this purpose, a statistical approach is first used to quantify the similarity of two sets of SGMRs for the assessment of the mean IM of the seismic collapse capacity of an assumed SDOF group. Based on this approach, two similarity indices are defined, and GA is applied in order to determine the optimal subset of SGMR for the considered SDOF group. PC analysis is then used to explore the best pattern for grouping the SDOF systems. The best pattern for grouping the SDOF systems is the one that results in the minimum size of the corresponding SGMR subsets. The results of PC analysis have shown that the best selection can

DOI: 10.1002/eqe

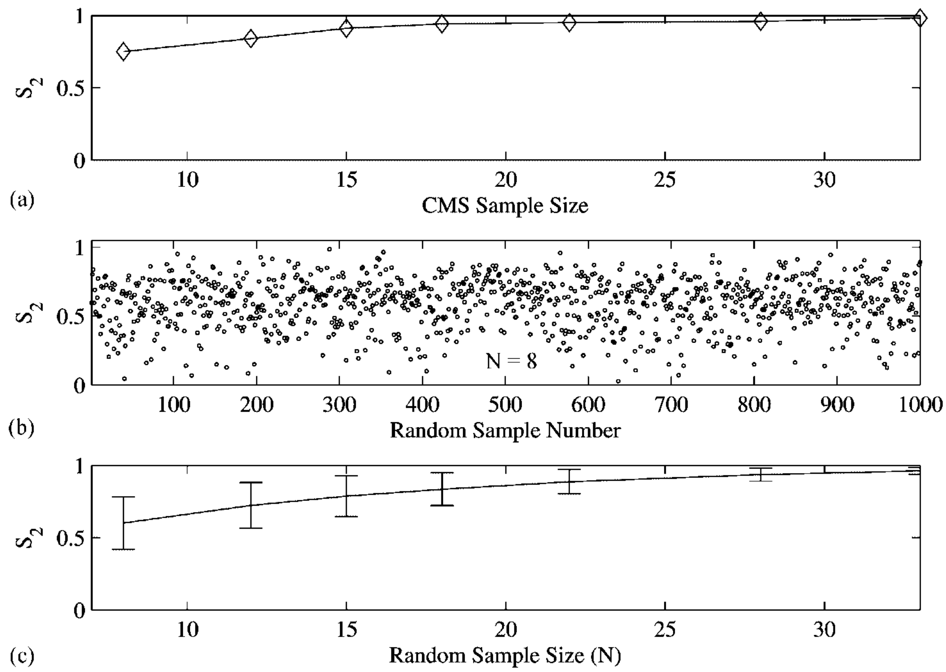

Figure 14. Comparison between randomly selected records and the CMS-based selection: (a) the second similarity index for different sample sizes of the CMS-based selected records; (b) the second similarity index for the use of eight randomly selected records versus different iterations; and (c) the median and dispersion of the second similarity index for different sample sizes of randomly selected records.

be achieved by grouping the SDOF systems, based on the period feature. Taking this into account, the range of the studied structural period values (0.1–2.0 s) has been divided into six groups, and eight records are proposed for each group (see Table III). The results confirm the efficiency of the presented subsets of SGMRs for the mean IM of the seismic collapse capacity assessment. Also, the proposed subset shows good agreement with the general set for the prediction of the IM in the full range of EDP, i.e. maximum drift.

The use of the presented subsets is more efficient than the application of the CMS-based record selection procedure. The results obtained show that the application of the presented subset, with eight records, is equivalent to the use of a CMS-based selected subset with 18 records.

As the proposed methodology is constrained by the SDOF models, the results are limited to first-mode-dominated structures. Investigation of higher mode effects and other factors is open to future research. The current study is also limited to the prediction of the mean seismic collapse capacity. An extension of this study to evaluate the variance of structural capacity could be valuable for the development of the damage assessment of different structures.

ACKNOWLEDGEMENTS

The research conducted by the authors has been funded by the International Institute of Earthquake Engineering and Seismology (IIEES) under Grant Number 7134. This support is gratefully acknowledged. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the funding body. The authors are also very grateful to both anonymous reviewers for their valuable comments.

REFERENCES

1. Vamvatsikos D, Cornell CA. Incremental dynamic analysis. *Earthquake Engineering and Structural Dynamics* 2002; **31**(3):491–514.
2. Iervolino I, Cornell CA. Records selection for nonlinear seismic analysis of structures. *Earthquake Spectra* 2005; **21**(3):658–713.

3. Han SW, Chopra AK. Approximate incremental dynamic analysis using the modal pushover analysis procedure. *Earthquake Engineering and Structural Dynamics* 2006; **35**(15):1853–1873.
4. Dolšek M, Fajfar P. Simplified non-linear seismic analysis of infilled reinforced concrete frames. *Earthquake Engineering and Structural Dynamics* 2005; **34**(1):49–66.
5. Vamvatsikos D, Cornell CA. Direct estimation of the seismic demand and capacity of multi-degree-of-freedom systems through incremental dynamic analysis of single degree of freedom approximation. *Journal of Structural Engineering* 2005; **131**(4):589–599.
6. ASCE7-05, *Minimum Design Loads for Structures and other Structures*. American Society of Civil Engineers: Reston, VA, 2005.
7. Naeim F, Alimoradi A, Pezeshk S. Selection scaling of ground motion earthquakes for structural design using genetic algorithms. *Earthquake Spectra* 2004; **20**(2):413–426.
8. Alimoradi A, Pezeshk S, Naeim F, Frigui H. Fuzzy pattern classification of strong ground motion records. *Journal of Earthquake Engineering* 2005; **9**(3):307–332.
9. Bommer JJ, Scott SG, Sarma SK. Hazard-consistent earthquake scenarios. *Soil Dynamics and Earthquake Engineering* 2000; **19**:219–231.
10. Naeim F, Lew M. On the use of design spectrum compatible time histories. *Earthquake Spectra* 1995; **11**:111–127.
11. Baker JW, Cornell CA. Spectral shape, epsilon and record selection. *Earthquake Engineering and Structural Dynamics* 2006; **35**(9):1077–1095.
12. Baker JW, Cornell CA. Vector-valued intensity measures incorporating spectral shape for prediction of structural response. *Journal of Earthquake Engineering* 2008; **12**(4):534–554.
13. Haselton CB, Baker JW, Liel AB, Deierlein GG. Accounting for ground motion spectral shape characteristics in structural collapse assessment through an adjustment for epsilon. *ASCE Journal of Structural Engineering* **136**(10). DOI: 10.1061/(ASCE)ST.1943-541X.0000103.
14. Azarbakht A, Dol²ek M. Prediction of the median IDA curve by employing a limited number of ground motion records. *Earthquake Engineering and Structural Dynamics* 2007; **36**(15):2401–2421.
15. Watson JL, Abrahamson N. Selection of ground motion time series and limits on scaling. *Soil Dynamics and Earthquake Engineering* 2006; **26**(5):477–482.
16. Mont gomery CD, Runger CG. *Applied Statistics and Probability for Engineers*. Wiley: New York, 2003.
17. Johnson AR, Wichern WD. *Applied Multivariate Statistical Analysis* (4th edn). Prentice-Hall: Englewood Cliffs, NJ, 1998.
18. Kroittmaier J. *Optimizing Engineering Designs*. McGraw-Hill: London, U.K., 1993.
19. Goldberg DE. *Genetic Algorithms in Search*, *Optimization and Machine Learning*. Addison-Wesley: Reading, MA, 1989.
20. ATC 63. Recommended methodology for quantification of building system performance and response parameters. 90% *Draft Report*, Applied Technology Council, Redwood City, CA, 2008.
21. Haselton CB. Deierlein GG. Assessing seismic collapse safety of modern reinforced concrete frame. *PEER Report 2007/08*, Pacific Engineering Research Center, University of California, Berkeley, CA, 2007.
22. Haselton CB. Evaluation of ground motion selection and modification methods: predicting median interstorey drift response of buildings. *Report No. 2009/01*, Pacific Engineering Research Center, University of California, Berkeley, CA, 2009.
23. Ibarra LF, Krawinkler H. Global collapse of frame structures under seismic excitations. *Report No. 152*, John A. Blume Earthquake Engineering Center, 2005.
24. Negro P, Mola E, Molina FJ, Magonette GE. Full-scale testing of a torsionally unbalanced three-storey nonseismic RC frame. *Proceedings of the 13th World Conference on Earthquake Engineering*, Vancouver, Canada, 2004; 968.
25. Fajfar P, Dolšek M, Marušic D, Stratan A, Pre- and post-test mathematical modeling of a plan-asymmetric reinforced concrete frame structure. *Earthquake Engineering and Structural Dynamics* 2006; **35**(11):1359–1379.
26. McKenna F, Fenves GL, Jeremic B, Scott MH. *Open System for Earthquake Simulations*, 2000. Available from: http://opensees.berekely.edu.
27. Baker JW. The conditional mean spectrum: a tool for ground motion selection. *ASCE Journal of Structural Engineering* **136**(10). DOI: 10.1061/(ASCE)ST.1943-541X.0000215.
28. Abrahamson NA, Silva WJ. Empirical response spectral attenuation relations for shallow crustal earthquakes. *Seismological Research Letters* 1997; **68**:94–126.

DOI: 10.1002/eqe