

Marquette University  
**e-Publications@Marquette**

---

Civil and Environmental Engineering Faculty  
Research and Publications

Civil and Environmental Engineering, Department  
of

---

8-1-2011

# A Computationally Efficient Ground-Motion Selection Algorithm for Matching a Target Response Spectrum Mean and Variance

Nirmal Jayaram  
*Stanford University*

Ting Lin  
*Marquette University, [ting.lin@marquette.edu](mailto:ting.lin@marquette.edu)*

Jack W. Baker  
*Stanford University*

---

Published version. *Earthquake Spectra*, Vol. 27, No. 3 (August 2011): 797-815. DOI. © 2011

Earthquake Engineering Research Institute. Used with permission.

Ting Lin was affiliated with Stanford University at time of publication

This article may be downloaded for personal use only. Any other use requires prior permission of the Earthquake Engineering Research Institute.

# A Computationally Efficient Ground-Motion Selection Algorithm for Matching a Target Response Spectrum Mean and Variance

Nirmal Jayaram,<sup>a)</sup> Ting Lin,<sup>a)</sup> M.EERI, and Jack W. Baker<sup>a)</sup> M.EERI

Dynamic structural analysis often requires the selection of input ground motions with a target mean response spectrum. The variance of the target response spectrum is usually ignored or accounted for in an ad hoc manner, which can bias the structural response estimates. This manuscript proposes a computationally efficient and theoretically consistent algorithm to select ground motions that match the target response spectrum mean and variance. The selection algorithm probabilistically generates multiple response spectra from a target distribution, and then selects recorded ground motions whose response spectra individually match the simulated response spectra. A greedy optimization technique further improves the match between the target and the sample means and variances. The proposed algorithm is used to select ground motions for the analysis of sample structures in order to assess the impact of considering ground-motion variance on the structural response estimates. The implications for code-based design and performance-based earthquake engineering are discussed.

## INTRODUCTION

Dynamic structural analysis is commonly used in performance-based earthquake engineering to predict the response of a structure subjected to earthquake ground motions. In the past, a wide variety of techniques have been developed for selecting input ground motions for the structural analysis (e.g., Haselton et al., 2009; Katsanos, 2010). One commonly used approach is to select recorded or simulated ground motions whose response spectra match a target mean response spectrum (e.g., Beyer and Bommer, 2007; Shantz, 2006; Watson-Lamprey and Abrahamson, 2006). Commonly used target spectra include the uniform hazard

---

<sup>a)</sup> Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, USA

spectrum and more recently, the conditional mean spectrum (Baker, 2010). Irrespective of the procedure used to obtain a target response spectrum, there are several methods for selecting input ground motions that match a desired spectrum. One approach is to select ground motions (sometimes after scaling) that individually deviate the least from the target response spectrum. The deviation can be measured using the sum of squared differences between the response spectrum of the record and the target response spectrum (e.g., Youngs et al., 2007). An alternate approach is to select a ground-motion set, rather than one record at a time, by minimizing the mean spectrum of the selected records from the target response spectrum. This is a more complicated subset selection optimization problem, and this optimization has been carried out in the past using genetic algorithms (Naeim et al., 2004).

In some situations, matching only a target mean response spectrum is not sufficient since the approach ignores the inherent variance that may exist in the response spectrum. For instance, the conditional mean spectrum is derived by conditioning on spectral acceleration at only a single period,  $S_a(T^*)$ , so the response spectra at other periods have variance. This is discussed in more detail in a subsequent section. Another approach might be to choose ground motions associated with a scenario earthquake (having a specified magnitude, distance, etc.), but not conditioned on any spectral values (e.g., Jayaram and Baker, 2010). In that case, the response spectrum will have variance at all periods. The commonly used uniform hazard spectrum (UHS), on the other hand, is an envelope of the spectral accelerations at all periods that are exceeded with a specified rate, as computed using probabilistic seismic hazard analysis (PSHA). PSHA already accounts for variability in spectral accelerations at each period being considered, and construction of the UHS is a conservative method of combining these spectral values. No further variance need be applied to the UHS, as varying the spectral values is equivalent to varying the associated exceedance rate of the spectral accelerations from period to period (an operation which is unlikely to produce meaningful results). In other words, if the UHS is used as the target mean spectrum, the ground motions should be chosen so that their spectra match the UHS and not have additional variance about the UHS.

In cases where a response spectrum has a non-zero variance, estimates of structural response obtained using the ground motions selected only based on the target mean spectrum will show smaller than ‘actual’ dispersion (logarithmic standard deviation) and may also have a biased mean value. Selecting ground motions based only on a target response spectrum

mean is computationally inexpensive since it can be done by choosing ground motions whose response spectra individually deviate the least from the target, and so the search time is proportional to the number of ground motions considered. When matching a target response spectrum mean and variance, it does not suffice to treat ground motions individually, but rather requires comparisons of the mean and variance of sets of ground motions to the target values. That is, the suitability of a particular ground motion can only be determined in the context of the complete ground-motion set in which it might be included. There is generally an intractably large number of possible ground-motion sets, and so identifying the best set is a computationally expensive combinatorial optimization problem. No automated procedures are available in the literature, however, to select ground motions that match the response spectrum mean and variance. One notable work in this regard is that of Kottke and Rathje (2008), who proposed a semi-automated procedure that first selects ground motions based on matching the mean spectrum, and subsequently applies individual scale factors on the ground motions to achieve the target variance. This technique, however, does not easily scale to work with large ground-motion datasets, and also cannot be used for the selection of unscaled ground motions.

The current work proposes a new, computationally fast and theoretically consistent ground-motion selection algorithm to match the target response spectrum mean and variance. The selection algorithm first uses Monte Carlo simulation to probabilistically generate multiple response spectra from a distribution parameterized by the target means and variances. For each simulated response spectrum, a ground motion with a similar response spectrum is then selected. Since the Monte Carlo simulated response spectra have the desired mean and variance, the response spectra of the selected recorded ground motions will also have the desired mean and variance. A greedy optimization technique then further improves the match between the target and the sample means and variances. This step replaces one previously selected ground motion at a time with a record from the ground-motion database that causes the best improvement in the match between the target and the sample means and variances.

The proposed algorithm is then used to select ground motions for estimating the seismic response of sample single-degree-of-freedom (SDOF) and multiple-degree-of-freedom (MDOF) structures, in order to assess the impact of considering ground-motion variance on

the structural response estimates. The implications for code-based design and performance-based earthquake engineering are discussed.

### **GROUND-MOTION SELECTION ALGORITHM**

The objective of the proposed algorithm is to select a suite of ground motions whose response spectra have a specified mean and variance. This algorithm is based on the empirically verified observation that the set of logarithmic spectral accelerations ( $\ln S_a$ ) at various periods is a random vector that follows a multivariate normal distribution (Jayaram and Baker, 2008). The first step in this algorithm is to parameterize the multivariate normal distribution of  $\ln S_a$ 's at multiple periods. The parameters of the multivariate normal distribution are the means and the variances of the  $\ln S_a$ 's at all periods and the correlations between the  $\ln S_a$ 's at all pairs of periods. Equivalently, the distribution can be parameterized using the means of the  $\ln S_a$ 's and the covariances between the  $\ln S_a$ 's at all pairs of periods. In order to achieve the desired properties in the selected ground motions, these parameters should be set to their target values (i.e., target means and variances for the ground motions to be selected). A subsequent section illustrates this parameterization.

Once the distribution means and covariances are set equal to the desired target values, Monte Carlo simulation is used to probabilistically generate response spectra from the above mentioned multivariate normal distribution. This can be performed using a standard function in many programming languages. The number of response spectra to be simulated equals the desired number of ground motions. For each simulated response spectrum, a ground motion with a similar response spectrum is then selected. The similarity between a ground-motion response spectrum and a Monte Carlo simulated response spectrum is evaluated using the sum of squared errors (*SSE*) described below:

$$SSE = \sum_{j=1}^p \left( \ln S_a(T_j) - \ln S_a^{(s)}(T_j) \right)^2 \quad (1)$$

where  $\ln S_a(T_j)$  is the logarithmic spectral acceleration of the (optionally scaled) ground motion in consideration at period  $T_j$ ,  $\ln S_a^{(s)}(T_j)$  is the target  $\ln S_a$  at period  $T_j$  from the simulated response spectrum,  $p$  is the number of periods considered and *SSE* is the sum of squared errors, which is a measure of dissimilarity. The measure of similarity defined by Equation 1 is not unique, and discussion of other measures of similarity can be found in

Beyer and Bommer (2007) and Buratti et al. (2010). The selection is done by computing  $SSE$  for each ground motion in the database, and then choosing the ground motion having the smallest  $SSE$ . Other ground motion properties can be accounted for at this stage by, for example, considering only ground motions falling within a specified range of magnitudes and distances. Note that this is identical to comparison procedures in use today, except that here we are comparing to simulated spectra rather than a target mean spectrum.

The mean and the variance of the simulated response spectra will approximately match the corresponding target values because they were sampled from the desired distribution. This match will be nearly exact if a large number of spectra are simulated and will be approximate otherwise. Since the simulated response spectra have approximately the desired mean and variance, the response spectra selected using this approach will also have approximately the desired mean and variance. Additionally, this ground-motion selection approach also ensures that the selected set has the target correlation structure (i.e., correlation between  $\ln S_a$ 's at pairs of periods) specified while parameterizing the distribution of the response spectrum. This implies that, in the particular case where the logarithmic response spectrum follows a multivariate normal distribution, the proposed algorithm actually matches the entire response spectrum distribution. Another advantage of this approach is that this algorithm allows the selection of unscaled ground motions (Jayaram and Baker, 2010).

As mentioned above, when ground motions are selected using the approach described above, the sample means and variances may deviate slightly from the target values, particularly when the number of ground motions selected is small. Therefore, a 'greedy' optimization technique is used to further improve the match between the sample and the target means and variances. In this approach, each ground motion selected previously is replaced one at a time with a ground motion from the database that causes the best improvement in the match between the target and the sample means and variances. If none of the potential replacements causes an improvement, the original ground motion is retained. The mismatch is estimated as the sum of squared differences between the target and the sample means and variances over the period range of interest. The deviation of the set mean and variance from the target mean and variance (denoted  $SSE_s$ ) is estimated as follows:

$$SSE_s = \sum_{j=1}^p \left[ \left( \hat{m}_{\ln S_a(T_j)} - \mu_{\ln S_a(T_j)}^{(t)} \right)^2 + w \left( \hat{\sigma}_{\ln S_a(T_j)} - \sigma_{\ln S_a(T_j)}^{(t)} \right)^2 \right] \quad (2)$$

where  $SSE_s$  is the sum of squared errors of the set, which is the parameter to be minimized,  $\hat{m}_{\ln S_a(T_j)}$  is the set mean  $\ln S_a$  at period  $T_j$ ,  $\mu_{\ln S_a(T_j)}^{(t)}$  is the target mean  $\ln S_a$  at period  $T_j$ ,  $\hat{s}_{\ln S_a(T_j)}$  is the set standard deviation of the  $\ln S_a$  at period  $T_j$ ,  $\sigma_{\ln S_a(T_j)}^{(t)}$  is the target standard deviation of the  $\ln S_a$  at period  $T_j$ ,  $w$  is a weighting factor indicating the relative importance of the errors in the standard deviation and the mean (A possible value for  $w$  is 1, but it can be chosen depending on the desired accuracy in the match between the sample and the target means and standard deviations), and  $p$  is the number of periods ( $T_j$ ) at which the error is computed.

The set mean and standard deviation can be calculated as follows:

$$\hat{m}_{\ln S_a(T_j)} = \frac{1}{n} \sum_{i=1}^n \ln S_{a_i}(T_j) \quad (3)$$

$$\hat{s}_{\ln S_a(T_j)} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \left( \ln S_{a_i}(T_j) - \hat{m}_{\ln S_a(T_j)} \right)^2} \quad (4)$$

where  $\ln S_{a_i}(T_j)$  denotes the  $\ln S_a$  of the  $i^{th}$  record in the set at period  $T_j$ , and  $n$  denotes the number of records in the set.

Note that the greedy optimization technique does not explicitly account for the correlation structure of selected sets. This correlation structure is captured in the initial selection step, and is approximately retained after the greedy optimization as well.

The steps involved in the greedy optimization technique are summarized below.

- Step 1: Set  $j = 1$ .
- Step 2: Set  $i = 1$ . Denote the  $SSE_s$  of the set as  $SSE_{s,old}$
- Step 3: If the  $i^{th}$  database ground motion ( $G_i$ ) is not already present in the set, replace the  $j^{th}$  ground motion in the set with  $G_i$ . Compute  $SSE_{s,i}$  (i.e, the  $SSE_s$  of the set after the replacement is carried out).
- Step 4: Reverse the replacement carried out in Step 3. Increment  $i$  by 1.
- Step 5: If  $i$  is less than or equal to the size of the ground-motion database, go to Step 3. Otherwise, identify the ground motion  $\tilde{i}$  that results in the minimum value of  $SSE_{s,\tilde{i}}$ .

$SSE_{s,\tilde{i}} < SSE_{s,old}$ , replace the  $j^{th}$  ground motion in the set with the  $\tilde{i}^{th}$  ground motion in the database.

- Step 6: Increment  $j$  by 1. If  $j$  is less than the size of the set, go to Step 2. Otherwise, terminate the algorithm.

This is called a ‘greedy’ optimization technique because it tries to improve the match between the target and the sample properties in every iteration locally without necessarily achieving a global optimum solution. In this application, the initial simulation and selection steps result in a ground motion set that is already approximately optimal (for reasonably large sets), so it has been observed that only this greedy technique is needed to find solutions that are essentially globally optimal. Observational experience suggests that this algorithm never produces sets of ground motions with poor matches between the sample and the target means and variances (even for sets with as few as 10 ground motions, as illustrated in a subsequent section).

Appendix “An Alternate Ground-Motion Selection Algorithm” describes an alternate selection algorithm that does not require knowledge of the response spectrum distribution or the correlation structure.

## **ILLUSTRATIVE GROUND-MOTION SELECTION**

This section describes the application of the proposed algorithm for selecting structure-specific ground motions that have a specified spectral acceleration at the structure’s fundamental period. In this example, the target response spectrum mean and covariance matrices are obtained using the conditional mean spectrum (CMS) method (Baker, 2010), which provides the mean and variance (and correlations) of the response spectrum conditioned on the specified spectral acceleration. It is to be noted that while this example uses the targets from the CMS method, the proposed algorithm can be used with any arbitrary target mean and covariance (e.g., Jayaram and Baker, 2010).

### **PARAMETERIZATION OF THE TARGET RESPONSE SPECTRUM DISTRIBUTION**

As described in the previous section, the first step in the algorithm is to parameterize the multivariate normal distribution of the  $\ln S_a$ ’s using the means and the variances of the spectral accelerations (chosen to equal the target mean and the target variance respectively)



and the correlations between the spectral accelerations at two different periods. The steps involved in parameterizing the distribution using the CMS method are listed below.

- Step 1: Determine the target spectral acceleration ( $S_a$ ) at a given period  $T^*$  (e.g., the fundamental period of the structure), and the associated magnitude ( $M$ ), distance to source ( $R$ ) and  $\varepsilon(T^*)$ , where  $\varepsilon(T^*)$  is the number of standard deviations by which a given  $\ln S_a$  differs from the mean predicted (by a ground-motion model)  $\ln S_a$  at the period of interest  $T^*$ . In general,

$$\varepsilon(T) = \frac{\ln S_a(T) - \mu_{\ln S_a(T)}}{\sigma_{\ln S_a(T)}} \quad (5)$$

where  $\ln S_a(T)$  is the ground motion's logarithmic spectral acceleration at period  $T$ , and  $\mu_{\ln S_a(T)}$  and  $\sigma_{\ln S_a(T)}$  are the predicted mean and standard deviation, respectively, of  $\ln S_a(T)$  given  $M$ ,  $R$ , etc. (e.g., Campbell and Bozorgnia, 2008). The values of  $M$ ,  $R$  and  $\varepsilon(T^*)$ , can be obtained from deaggregation (e.g., USGS, 2008).

- Step 2: For all  $T_j$  of interest, compute the unconditional mean and the unconditional standard deviation of the response spectrum, given  $M$  and  $R$ . In other words, compute  $\mu_{\ln S_a(T)}$  and  $\sigma_{\ln S_a(T)}$ .
- Step 3: Compute the mean of  $(\ln S_a(T_1), \ln S_a(T_2), \dots, \ln S_a(T_n))$  conditioned on  $\varepsilon(T^*)$ . This mean matrix (denoted  $\boldsymbol{\mu}$ ) is computed as follows:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_{\ln S_a(T_1)} + \rho(T_1, T^*)\varepsilon(T^*)\sigma_{\ln S_a(T_1)} \\ \mu_{\ln S_a(T_2)} + \rho(T_2, T^*)\varepsilon(T^*)\sigma_{\ln S_a(T_2)} \\ \vdots \\ \mu_{\ln S_a(T_n)} + \rho(T_n, T^*)\varepsilon(T^*)\sigma_{\ln S_a(T_n)} \end{bmatrix} \quad (6)$$

where  $\rho(T_j, T^*)$  is the correlation between  $\varepsilon(T_j)$  and  $\varepsilon(T^*)$  provided by, for instance, Baker and Jayaram (2008).

- Step 4: Compute the covariance of  $(\ln S_a(T_1), \ln S_a(T_2), \dots, \ln S_a(T_n))$  conditioned on  $\varepsilon(T^*)$ . This covariance matrix (denoted  $\boldsymbol{\Sigma}$ ) is estimated as follows:

Let  $\Sigma_0$  denote the (unconditional) covariance matrix of the vector  $(\ln S_a(T_1), \ln S_a(T_2), \dots, \ln S_a(T_n))$ .

$$\Sigma_0 = \begin{bmatrix} \sigma_{\ln S_a(T_1)}^2 & \rho(T_1, T_2) \sigma_{\ln S_a(T_1)} \sigma_{\ln S_a(T_2)} & \cdots & \rho(T_1, T_n) \sigma_{\ln S_a(T_1)} \sigma_{\ln S_a(T_n)} \\ \rho(T_2, T_1) \sigma_{\ln S_a(T_2)} \sigma_{\ln S_a(T_1)} & \sigma_{\ln S_a(T_2)}^2 & \cdots & \rho(T_2, T_n) \sigma_{\ln S_a(T_2)} \sigma_{\ln S_a(T_n)} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \rho(T_n, T_1) \sigma_{\ln S_a(T_n)} \sigma_{\ln S_a(T_1)} & \rho(T_n, T_2) \sigma_{\ln S_a(T_n)} \sigma_{\ln S_a(T_2)} & \cdots & \sigma_{\ln S_a(T_n)}^2 \end{bmatrix} \quad (7)$$

Let  $\Sigma_1$  denote the covariance between  $(\ln S_a(T_1), \ln S_a(T_2), \dots, \ln S_a(T_n))$  and  $\ln S_a(T^*)$ , defined as follows:

$$\Sigma_1 = \begin{bmatrix} \rho(T_1, T^*) \sigma_{\ln S_a(T_1)} \sigma_{\ln S_a(T^*)} \\ \rho(T_2, T^*) \sigma_{\ln S_a(T_2)} \sigma_{\ln S_a(T^*)} \\ \cdot \\ \cdot \\ \rho(T_n, T^*) \sigma_{\ln S_a(T_n)} \sigma_{\ln S_a(T^*)} \end{bmatrix} \quad (8)$$

The covariance matrix of  $(\ln S_a(T_1), \ln S_a(T_2), \dots, \ln S_a(T_n))$  conditioned on  $\ln S_a(T^*)$  can be computed as follows (e.g., Johnson and Wichern, 2007):

$$\Sigma = \Sigma_0 - \frac{1}{\sigma_{\ln S_a(T^*)}^2} \Sigma_1 \Sigma_1' \quad (9)$$

where  $\Sigma_1'$  denotes the transpose of  $\Sigma_1$ . The conditional standard deviation of the  $\ln S_a$ 's is the square root of the diagonals of  $\Sigma$ , also given by Equation 10.

$$\sigma_{\ln S_a(T) | \ln S_a(T^*)} = \sigma_{\ln S_a(T)} \sqrt{1 - \rho(T, T^*)^2} \quad (10)$$

Figure 1 shows the target conditional response spectrum mean and standard deviation obtained corresponding to magnitude = 7, distance to the rupture = 10km,  $T^* = 2.63$ s and  $\varepsilon(T^*) = 2.0$ . These values have been chosen to be compatible with ground-motion studies carried out by Haselton et al. (2009). The unconditional  $\ln S_a$  means and standard deviations corresponding to this scenario,  $\mu_{\ln S_a(T_j)}$  and  $\sigma_{\ln S_a(T_j)}$ , are obtained from the Campbell and Bozorgnia (2008) ground-motion model. (Since  $\ln S_a$ 's at multiple periods follow a

multivariate normal distribution, the exponential of the mean  $\ln S_a$  equals the median spectral acceleration. This is why the axis of Figure 1 is labeled as ‘Median  $S_a$ ’.)

**Figure 1.** (a) Response spectrum mean (b) Response spectrum standard deviation.

## RESPONSE SPECTRUM SIMULATION

Forty response spectra are simulated (using Monte Carlo simulation) by sampling from a multivariate normal distribution with the mean and covariance matrices defined by Equations 6 and 9 for the target scenario described above. The response spectra are simulated at 20 periods logarithmically spaced between 0.05s and 10.0s, and are shown in Figure 2a. A large period range is used to ensure a good match in the entire response spectrum that covers regions of higher modes and nonlinearity. Because individual spectra may vary while still achieving a target mean and variance of the overall set, there is often little penalty in considering a broad period range in this step.

**Figure 2.** (a) Simulated response spectra (b) Response spectra of ground motions selected before greedy optimization (c) Response spectra of ground motions selected after greedy optimization.

Figure 1a compares the mean of the Monte Carlo simulated response spectra to the target mean. It can be seen that the mean values agree reasonably well. Figure 1b shows a reasonable agreement between the standard deviation of the simulated  $\ln S_a$  values and the target standard deviation. The small deviation seen in these figures is because the sample mean and standard deviation for moderately small sample sizes do not necessarily match the target mean and standard deviation.

## SELECTION OF GROUND MOTIONS TO MATCH SIMULATED SPECTRA

Forty ground motions are selected from the Next Generation Attenuation (NGA) database (Chiou et al., 2008) that individually match the forty response spectra simulated in the previous step. For two-dimensional structural models, a single ground motion component is required as an input for every time history analysis. (For three-dimensional structural models, two ground motion components can be selected by considering their geometric mean response spectrum, as described in Jayaram and Baker 2010.) Here, each horizontal component of a recording from the same station in the NGA database is treated separately as an individual ground motion. No constraints on, for example, the magnitudes and distances of the selected recordings are used, but such constraints are easily accommodated by simply

restricting the set of ground motions considered for selection. Prior to selection, each of the available 7102 ground motions in the NGA database is scaled so that its  $S_a(T^*)$  matches the target  $S_a(T^*)$  from the target mean spectrum (seen in Figure 1a) when  $T^*$  is equal to 2.63s. Figure 2b shows the response spectra of the selected ground motions. The sample and the target means and standard deviations are shown in Figure 1. The figure shows that the sample and the target response spectrum mean and variance match reasonably well. Additionally, the selected ground motion spectra also match the specified target correlation structure (specified by the non-diagonal terms of the covariance matrix in Equation 9) reasonably well, as indicated by a mean absolute error between the sample and the target correlations of 0.12.

The computational time required for selecting the set of 40 ground motions is 10 seconds using a MATLAB implementation on an 8GB RAM 2.33GHz quad core processor. This computational efficiency allows for the algorithm to be optionally applied multiple times if one wants several candidate sets to choose from. While selecting the ground motions shown in Figure 2, we applied the algorithm multiple times (twenty times, in particular) to obtain multiple candidate ground-motion sets and chose the set with the minimum value of  $SSE$ . This approach is also beneficial to work around situations where recorded ground motion spectra that adequately match one or more of the simulated spectra are not available.

## **GREEDY OPTIMIZATION TECHNIQUE**

The greedy optimization technique is used to modify the ground-motion suite selected in the previous step. The spectra of the selected ground motions are shown in Figure 2c. The means and the standard deviations of the set are shown in Figure 1, and have a near perfect match with the target means and standard deviations. The mean absolute error between the sample and the target correlations is 0.15.

In total, the computational time required to select the set of 40 ground motions from the 7102 available ground motions is about 180 seconds using a MATLAB implementation on an 8GB RAM 2.33GHz quad core processor. A MATLAB implementation of the proposed ground-motion selection algorithm can be downloaded from [http://www.stanford.edu/~bakerjw/gm\\_selection.html](http://www.stanford.edu/~bakerjw/gm_selection.html).

## SELECTION OF A SMALLER NUMBER OF GROUND MOTIONS

To test the effectiveness of the algorithm in sampling smaller ground motion sets, it is repeated to select a set of 10 ground motions for the scenario described earlier (magnitude = 7, distance to rupture = 10km,  $T^* = 2.63\text{s}$  and  $\varepsilon(T^*) = 2$ ). The response spectra of the selected records are shown in Figure 3a. The set means and standard deviations are compared to the target means and standard deviations in Figure 3b-c. It can be seen that the matches are good, illustrating the effectiveness of the algorithm in selecting small sets of ground motions. The mean absolute error between the sample and the target correlations is 0.17. The computational time required to select the set of 10 ground motions is about 25 seconds using a MATLAB implementation on an 8GB RAM 2.33GHz quad core processor. The computational time required for selecting the set of 10 ground motions without using the greedy optimization technique is 4 seconds.

**Figure 3.** (a) Response spectra of 10 selected ground motions (b) Response spectrum mean (c) Response spectrum standard deviation.

## IMPACT OF MATCHING SPECTRUM VARIANCE ON STRUCTURAL RESPONSE

Code-based structural design and performance-based earthquake engineering applications require statistics such as the mean (e.g., ASCE, 2005) or the median and the dispersion (e.g., ATC-58, 2009) of the structural response. It is of interest in this section to evaluate the impact of ground-motion selection considering a target response spectrum mean and variance (as compared to considering only a target mean) on these statistics.

### GROUND-MOTION SELECTION

The ground motions used for evaluating structural response are selected using the method described in the previous section for a target scenario with magnitude = 7, distance to rupture = 10km,  $V_{s30} = 400\text{m/s}$ , and a strike-slip mechanism. The Campbell and Bozorgnia (2008) ground-motion model is used to estimate the mean and variance of the response spectrum. The values of  $\varepsilon$  and period  $T^*$  are varied to obtain multiple test scenarios. Three typical  $\varepsilon$  values of 0, 1 and 2 are considered. The structures considered in this work have periods ( $T^*$ ) ranging between 0.5s and 2.63s.

In order to investigate the impact of matching response spectrum variance (Equation 9) on the structural response statistics, sets of forty ground motions are selected using two methods: ‘Method 1’ in which only the target mean is matched (a common approach in current practice, e.g., Baker and Cornell, 2006 and Method 300 in Haselton et al., 2009) and ‘Method 2’ in which both the target mean and the target variance are matched using the approach proposed here. The target response spectrum mean and covariance matrices are evaluated using Equations 6 and 9 for each combination of  $\varepsilon$  and  $T^*$ . Figure 4 shows example response spectra of ground motions selected using these two methods (for  $\varepsilon = 2$  and  $T^* = 2.63\text{s}$ ).

**Figure 4.** Response spectra of 40 selected ground motions for  $\varepsilon = 2$  and  $T^* = 2.63\text{s}$  (a) Using Method 1: Match target response spectrum mean, and (b) Using Method 2: Match target response spectrum mean and variance.

## STRUCTURAL RESPONSE

This section describes the response of sample nonlinear single-degree-of-freedom (SDOF) structures and multiple-degree-of-freedom (MDOF) buildings designed according to modern building codes. In this work, we consider only maximum displacement for the SDOF structures and maximum interstory drift ratio (*MIDR*) for the MDOF structures.

### Description of structural systems

The SDOF structures considered in this work follow a non-deteriorating, bilinear force-displacement relationship (Chopra, 2007). They have  $T^* = 0.5\text{s}$ , 5% damping and post-yielding stiffness equal to 10% of elastic stiffness. SDOF structures with ‘R factors’ (the ratio of the target spectral acceleration at the period of the structure,  $S_a(T^*)$ , to the yield spectral acceleration =  $\omega^2 * \text{yield displacement}$ , where  $\omega$  is the structure’s fundamental circular frequency) of 1, 4 and 8 are considered to study varying levels of non-linear behavior. The R factor is controlled by varying the yield displacements of the SDOF structures relative to the  $S_a(T^*)$  value obtained from the target spectrum. The SDOF structures are non-deteriorating systems, so structural collapse is not considered.

The MDOF structures used in this study were designed per modern building codes and modeled utilizing the Open System for Earthquake Engineering Simulation (OpenSEES) (2007) by Haselton and Deierlein (2007). The structural models consider strength and stiffness deterioration (Ibarra et al., 2005) unlike in the SDOF case. The designs for these

buildings have been checked by practicing engineers as part of the Applied Technology Council Project ATC-63 (FEMA, 2009). They have also been used for previous extensive ground-motion studies (Haselton et al., 2009). The two buildings used in the current study are a 4-story reinforced concrete moment frame structure with  $T^* = 0.94\text{s}$ , and a 20-story reinforced concrete moment frame structure with  $T^* = 2.63\text{s}$ . The buildings show deterioration, and collapse is said to occur if dynamic instability (large increases in the drift for small increases in the ground-motion intensity) is reached in the model (Haselton and Deierlein, 2007).

### **Response of SDOF systems**

Table 1 shows the mean, median and dispersion (dispersion refers to logarithmic standard deviation) of ductility ratios (spectral displacement divided by the yield displacement) of the SDOF structures under the different ground-motion scenarios described earlier. The ductility statistics are estimated using the two sets of 40 ground motions selected using Method 1 (ground motions selected by matching only the target response spectrum mean) and Method 2 (ground motions selected by matching the target response spectrum mean and variance). It can be seen from Table 1 that the median ductilities are similar across the two ground-motion selection methods, while the mean and the dispersion of the response are higher in Method 2, when the ground-motion variance is considered. The higher dispersion of the response seen while using Method 2 is a result of considering the uncertainty in the response spectra, which is ignored in Method 1. As expected, the increase in dispersion is particularly significant at large R values when the structure behaves in a non-linear manner. Note that there are no differences between the methods when  $R = 1$ , because the response is dependent only on  $S_a(T^*)$ , which is identical in both cases.

**Table 1.** Ductility ratio of SDOF structure.

$\varepsilon$	R	Median Ductility		Dispersion of Ductility		Mean Ductility	
		Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
0	1	1.00	1.00	0	0	1.00	1.00
	4	3.93	3.76	0.24	0.31	4.21	4.18
	8	10.76	9.97	0.28	0.42	10.82	10.74
1	1	1.00	1.00	0	0	1.00	1.00
	4	3.55	3.35	0.22	0.33	3.79	3.93
	8	8.04	8.16	0.28	0.46	8.57	9.46
2	1	1.00	1.00	0	0	1.00	1.00
	4	3.27	3.04	0.19	0.28	3.39	3.34
	8	6.90	7.44	0.24	0.41	7.34	7.98

Figure 5 shows the fraction of response analyses that result in a ductility less than a specified value for the SDOF structure with  $R = 8$  in the  $\varepsilon = 1$  scenario, estimated using Methods 1 and 2. This type of plot is referred to as an empirical cumulative distribution function, or CDF. The CDFs intersect at a value of approximately 0.5 due to the similarity in the median response in both cases. The CDF obtained using Method 2 is flatter with heavier tails as a result of the larger dispersion observed in this case. As seen from Figure 5a, the upper tails of the CDFs are heavier than the lower tails. Since the mean response is the area above the CDF (the mean of a random variable is the area under the complementary CDF, which equals  $1 - \text{CDF}$ ), it can be visually observed that the difference in the heaviness of the upper tails results in a larger mean value of the response in case of Method 2 as compared to Method 1. This is a graphical evidence of the larger mean values reported earlier in Table 1. Analytically, if the responses were to follow a lognormal distribution (a common assumption in performance-based earthquake engineering), the properties of the lognormal distribution will imply that a larger dispersion results in a larger mean for a fixed median, which also explains the larger means observed in Method 2.

**Figure 5.** Distribution of the structural response of the SDOF structure corresponding to  $R = 8$  and  $\varepsilon(T^*) = 1$ : (a) Linear scale (b) Logarithmic scale.

### Response of MDOF systems

Table 2 summarizes the maximum interstory drift ratio (*MIDR*) estimates for the MDOF structures considered in this study under various ground-motion scenarios, estimated using



Methods 1 and 2. The distributions of responses are summarized using the probability of collapse (i.e., counted fraction of responses indicating collapse) and the median and the dispersion of the non-collapse responses.

**Table 2.** Maximum interstory drift ratio (*MIDR*) of 20-story and 4-story moment frames.

Building	$\varepsilon$	Median <i>MIDR</i>		Dispersion of <i>MIDR</i>		Collapse Probability	
		Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
20-story moment frame	0	0.0044	0.0043	0.18	0.32	0	0
	1	0.0096	0.0086	0.24	0.29	0	0
	2	0.0186	0.0196	0.25	0.43	0	0.05
4-story moment frame	0	0.0072	0.0072	0.09	0.09	0	0
	1	0.0137	0.0139	0.26	0.29	0	0
	2	0.0279	0.0237	0.28	0.46	0.10	0.20

From Table 2, it can be seen that, as observed in the SDOF case, the medians are similar across Methods 1 and 2 in all the considered scenarios. The dispersions are larger, however, when the ground-motion variance is considered in Method 2. The increase in the dispersion also results in an increased probability of observing large values of structural response. This can result in an increased probability of structural collapse while using Method 2, as evidenced, for example, when  $\varepsilon = 2$  in Table 2.

Figure 6 shows the empirical CDF of the *MIDR* of the 20-story frame corresponding to the  $\varepsilon = 2$  ground-motion scenario. As seen in the SDOF case, the CDF obtained using Method 2 is flatter and has heavier tails on account of larger dispersion. The maximum plotted values of the CDFs differ from one, and the difference equals the probability of collapse.

**Figure 6.** Distribution of the structural response of the 20 story moment frame building corresponding to  $\varepsilon(T^*) = 2$ : (a) Linear scale (b) Logarithmic scale.

In summary, the response estimates for the SDOF and the MDOF structures across several ground-motion scenarios show that the consideration of the response spectrum variance while selecting ground motions does not significantly impact the median structural response, but tends to increase the mean response and the dispersion in the response. The increased dispersion can result in more extreme responses, which can lead to a larger probability of structural collapse.

These example analysis cases serve to illustrate the potential importance of matching response spectrum variance. More detailed investigations regarding the impact are important, and will be carried out in the future.

## **IMPLICATIONS**

Code-based design is often concerned with the average response of the structure (e.g., ASCE, 2005). The average response is typically interpreted as the mean response, although sometimes it is interpreted as the median. If median structural response is of interest, the consideration of the response spectrum variance while selecting ground motions does not have a significant impact in the limited investigation performed here. On the other hand, if mean structural response is of interest, the consideration of the response spectrum variance appears to increase the mean structural response and may thus impact code-based design calculations.

Performance-based earthquake engineering (PBEE), in contrast, often requires knowledge about the full distribution of structural response (ATC-58, 2009). Matching target response spectrum variance increases the dispersion of structural response, thereby affecting the distribution of structural response and consequently the damage state and loss estimation computations in PBEE. The increase in the dispersion leads to higher and lower extremes of structural response and the associated damage states and losses. The increased dispersion can also lead to a larger probability of structural collapse. PBEE calculations will thus almost certainly be affected by this issue.

In summary, the example analyses presented above and engineering intuition suggest that the target response spectrum variance used when selecting ground motions has an impact on the distribution of structural responses obtained from resulting dynamic analysis. It appears that this is true for both code-based design checks and performance-based earthquake engineering analysis. Further study is needed to quantify the magnitude of these impacts, and this new algorithm will facilitate such studies.

## **CONCLUSIONS**

A computationally efficient, theoretically consistent ground-motion selection algorithm was proposed to enable selection of a suite of ground motions whose response spectra have a target mean and a target variance. The algorithm first uses Monte Carlo simulation to

probabilistically generate multiple realizations of response spectra from a target distribution, and then selects recorded ground motions whose response spectra individually match the simulated response spectra. A greedy optimization technique then further improves the match between the target and the sample means and variances by replacing one previously selected ground motion at a time with a record from the ground-motion database that causes the best improvement in the match. It was shown empirically that this selection algorithm selects ground motions whose response spectra have the target mean and variance.

The proposed algorithm was then used to select ground motions for estimating the seismic response of sample single-degree-of-freedom (SDOF) and multiple-degree-of-freedom (MDOF) structures, in order to assess the impact of considering response spectrum variance on the structural response estimates. SDOF structures with different levels of non-linearity (as indicated by their R factors) were analyzed using the selected ground motions. It was seen that considering the response spectrum variance does not significantly affect the resulting median response, but slightly increases the mean response and considerably increases the dispersion (logarithmic standard deviation) of the response. The increase in the mean and the dispersion is larger for more non-linear SDOF structures. Two code-compliant MDOF structures with heights of 4 and 20 stories were also analyzed using the selected ground motions. As with the SDOF structures, it was seen that considering the response spectrum variance does not significantly affect the median response but increases the dispersion of the response and the probability of observing collapse. These observations have implications for applications where the dispersion of the response is an important consideration, such as in many performance-based engineering evaluations. A MATLAB implementation of the proposed ground-motion selection algorithm can be downloaded from [http://www.stanford.edu/~bakerjw/gm\\_selection.html](http://www.stanford.edu/~bakerjw/gm_selection.html)

### **ACKNOWLEDGMENTS**

The authors thank Curt Haselton, Tom Shantz, Nilesh Shome, Peter Stafford and an anonymous reviewer for their helpful reviews of the manuscript. Also, thanks to Curt Haselton for providing the structural models used for the example analyses. This work was supported by the State of California through the Transportation Systems Research Program of the Pacific Earthquake Engineering Research Center (PEER), and by Cooperative Agreement Number 08HQAG0115 from the United States Geological Survey. Any opinions, findings,

conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the funding agencies.

## APPENDIX A      A GREEDY GROUND-MOTION SELECTION TECHNIQUE

The ground-motion selection algorithm described in the body of this manuscript selects an initial set of ground motions whose response spectra match a set of simulated response spectra. These simulations are obtained from a multivariate normal distribution parameterized by the target mean and covariance matrices. A greedy optimization technique then further improves the match between the target and the sample means and variances and obtains the final set of ground motions.

Sometimes, it may not be possible to completely parameterize the distribution of the response spectra using the mean and covariance information. This includes situations where ground motions are selected to match the UHS (where only the mean spectrum needs to be considered) or where the mean and the variance information, but not the correlation information, are available. There may also be situations where the response spectrum does not follow a multivariate normal distribution. For such situations, the authors propose the following technique for selecting the initial ground-motion set that can be subsequently improved by the greedy optimization technique. The steps involved in the technique are summarized below.

- Step 1: Initialize the algorithm with an empty ground-motion set.
- Step 2: Set  $i = 1$ .
- Step 3: If the  $i^{\text{th}}$  database ground motion ( $G_i$ ) is not already present in the ground-motion set, include it in the set and compute  $SSE_{s,i}$  (i.e, the  $SSE_s$  of the set after  $G_i$  is included, where  $SSE_s$  is defined in Equation 2).
- Step 4: Delete  $G_i$  from the set, if included in Step 3. Increment  $i$  by 1.
- Step 5: If  $i$  is less than or equal to the size of the ground-motion database, go to Step 3. Otherwise, identify the ground motion  $\tilde{i}$  that results in the minimum value of  $SSE_{s,\tilde{i}}$ . Add the  $\tilde{i}^{\text{th}}$  ground motion in the database to the ground-motion set.
- Step 5: If the size of the set equals the desired number of ground motions, terminate the algorithm. Otherwise, go to Step 2.

This selection technique will provide a reasonable starting set of ground motions that can be subsequently improved using the greedy optimization technique described earlier in the manuscript. This selection technique does not take advantage of the knowledge of the response spectrum distribution or the correlation structure, but is therefore more general in its application. It is also empirically seen to produce sets of ground motions with response spectrum mean and variance closely matching the corresponding target values.

To test the effectiveness of the technique, it is used to select a set of 40 ground motions for the scenario described earlier (magnitude = 7, distance to rupture = 10km,  $T^* = 2.63$ s and  $\varepsilon(T^*) = 2$ ). The response spectra of the selected records are shown in Figure 7a. The ground-motion set means and standard deviations are compared to the target means and standard deviations in Figure 7b-c. It can be seen that the matches are good, illustrating the effectiveness of the technique. Incidentally, despite the fact that the technique does not use the correlation information, it is seen that the mean absolute error between the sample and the target correlations (Equation 7) is only 0.15.

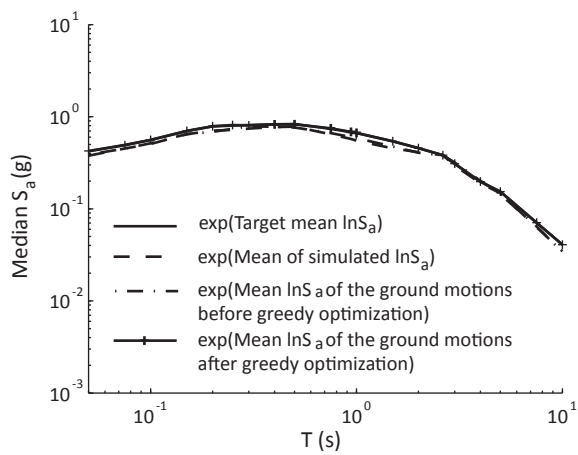
**Figure 7.** (a) Response spectra of 40 ground motions selected using the greedy selection and optimization techniques (b) Response spectrum mean (c) Response spectrum standard deviation.

## REFERENCES

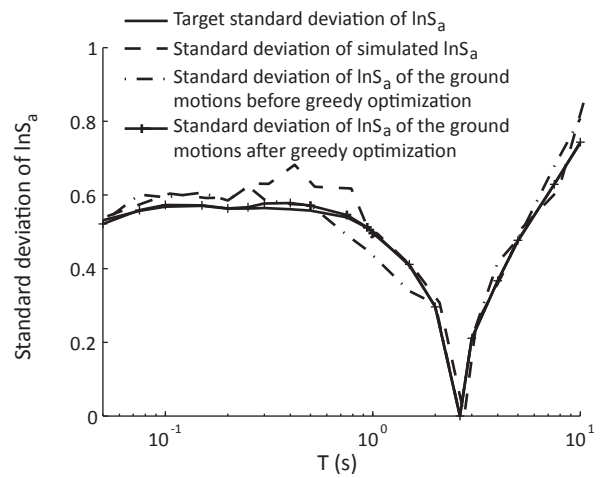
- ASCE, 2005. *Minimum design loads for buildings and other structures*. ASCE 7-05, American Society of Civil Engineers/Structural Engineering Institute, Reston, VA.
- ATC-58, 2009. *Guidelines for Seismic Performance Assessment of Buildings ATC-58 50% Draft*, The Applied Technology Council, Redwood City, CA.
- Baker, J.W. and Cornell, C.A., 2006. Spectral shape, epsilon and record selection, *Earthquake Engineering & Structural Dynamics* **35**, 1077-1095.
- Baker, J.W., 2010. The conditional mean spectrum: A tool for ground motion selection, *ASCE Journal of Structural Engineering*, *in press*.
- Baker, J.W. and Jayaram, N., 2008. Correlation of spectral acceleration values from NGA ground motion models, *Earthquake Spectra* **24**, 299–317.
- Beyer, K. and Bommer, J.J., 2007. Selection and scaling of real accelerograms for bi-directional loading: A review of current practice and code provisions, *Journal of Earthquake Engineering* **11**, 13–45.
- Buratti, N., Stafford, P.J., and Bommer, J.J., 2010. Earthquake accelerogram selection and scaling

- procedures for estimating the distribution of structural response, *ASCE Journal of Structural Engineering*, published online.
- Campbell, K.W. and Bozorgnia, Y., 2008. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10s, *Earthquake Spectra* **24**, 139–171.
- Chiou, B.S.J., Darragh, R.B., Gregor, N.J., and Silva, W.J., 2008. NGA project strong-motion database, *Earthquake Spectra* **24**, 23–44.
- Chopra, A.K., 2007. *Dynamics of structures*, Prentice Hall, Upper Saddle River, NJ.
- FEMA, 2009. *Recommended Methodology for Quantification of Building System Performance and Response Parameters*, FEMA P695A, Prepared for the Federal Emergency Management Agency, The Applied Technology Council, Redwood City, CA.
- Haselton, C.B., Baker, J.W., Bozorgnia, Y., Goulet, C.A., Kalkan, E., Luco, N., Shantz, T.J., Shome, N., Stewart, J.P., Tothong, P., Watson-Lamprey, J.A., and Zareian, F., 2009. *Evaluation of ground motion selection and modification methods: Predicting median interstory drift response of buildings*, Technical report, Report 2009/01, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.
- Haselton, C.B. and Deierlein, G.G., 2007. *Assessing seismic collapse safety of modern reinforced concrete frame buildings*, Technical report, Report 2007/08, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.
- Ibarra, L.F., Medina, R.A., and Krawinkler, H., 2005. Hysteretic models that incorporate strength and stiffness deterioration, *Earthquake Engineering and Structural Dynamics* **34**, 1489–1511.
- Jayaram, N. and Baker, J.W., 2010. Ground-motion selection for PEER Transportation Systems Research Program, in *Proceedings, 7th CUEE and 5th ICEE Joint Conference*, Tokyo, Japan.
- Jayaram, N. and Baker, J.W., 2008. Statistical tests of the joint distribution of spectral acceleration values, *Bulletin of the Seismological Society of America* **98**, 2231–2243.
- Johnson, R.A. and Wichern, D.W., 2007. *Applied Multivariate Statistical Analysis*, Prentice Hall, Upper Saddle River, NJ.
- Katsanos, E.I., 2010. Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering perspective, *Soil Dynamics and Earthquake Engineering*, in press.
- Kottke, A.R. and Rathje, E.M., 2008. A semi-automated procedure for selecting and scaling recorded earthquake motions for dynamic analysis, *Earthquake Spectra* **24**, 911–932.
- Naeim, F., Alimoradi, A., and S. Pezeshk, S., 2004. Selection and scaling of ground motion time histories for structural design using genetic algorithms, *Earthquake Spectra* **20**, 413–426.

- Open System for Earthquake Engineering Simulation (OpenSEES), 2007. Pacific Earthquake Engineering Research Center, University of California, Berkeley, <http://opensees.berkeley.edu/> (last verified 03/29/2010).
- Shantz, T.J., 2006. Selection and scaling of earthquake records for nonlinear dynamic analysis of first model dominated bridge structures, in *Proceedings of the 8th U.S. National Conference on Earthquake Engineering*, San Francisco, CA.
- USGS, 2008. 2008 Interactive Deaggregations (Beta). <http://eqint.cr.usgs.gov/deaggint/2008/> (last verified 03/29/2010).
- Watson-Lamprey, J.A. and Abrahamson, N.A., 2006. Selection of ground motion time series and limits on scaling, *Soil Dynamics and Earthquake Engineering* **26**, 477–482.
- Youngs, R.R., Power, M.S., Wang, G., Makdisi, F.I., and Chin, C.C., 2007. Design ground motion library (DGML) - Tool for selecting time history records for specific engineering applications, in *SMIP Seminar on Utilization of Strong-Motion Data*.

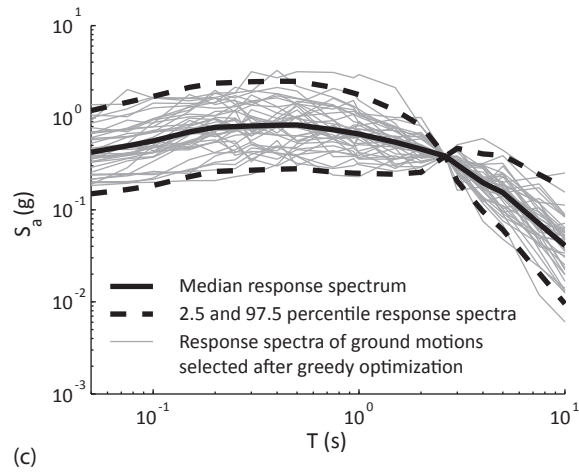
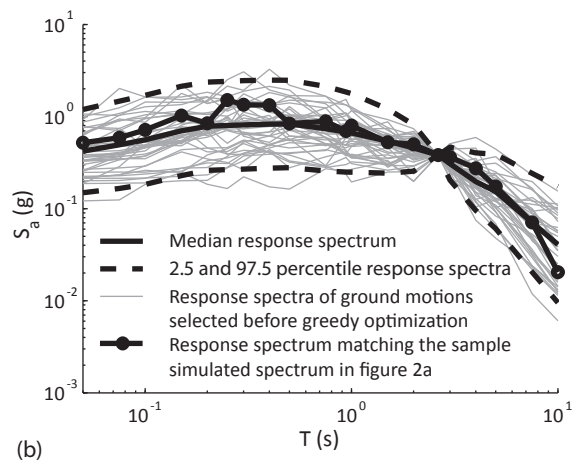
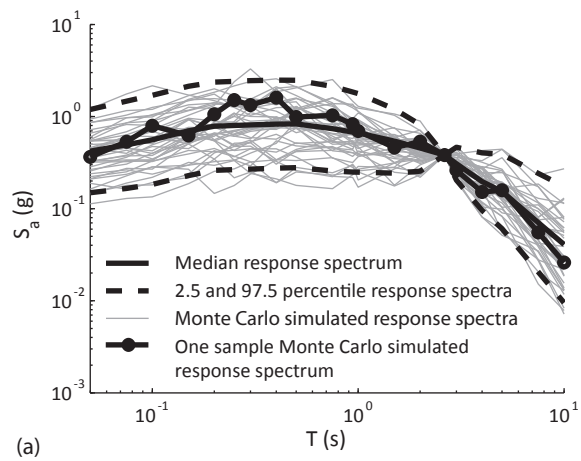


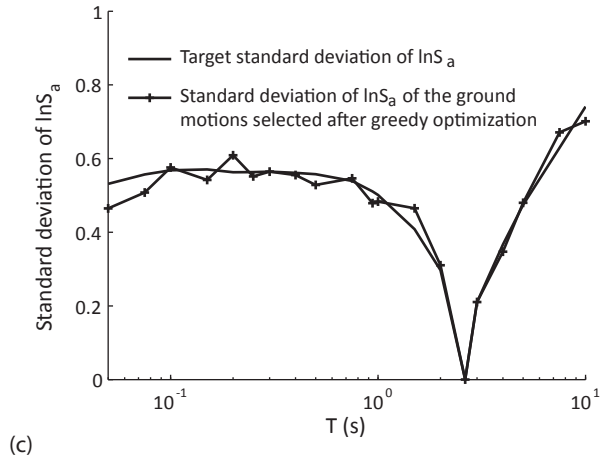
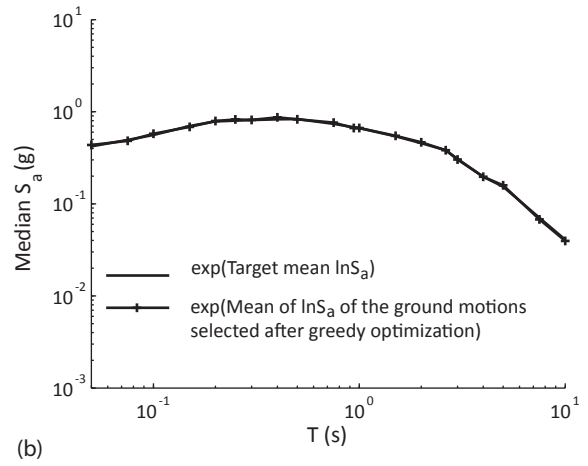
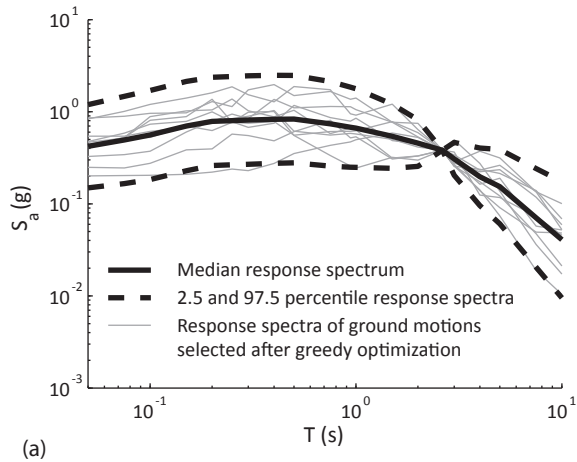
(a)

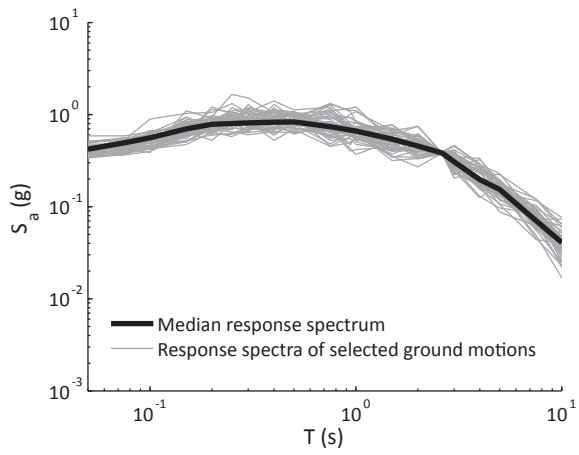


(b)

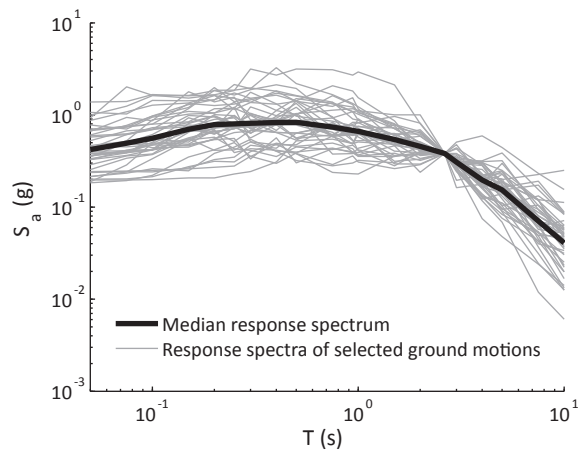




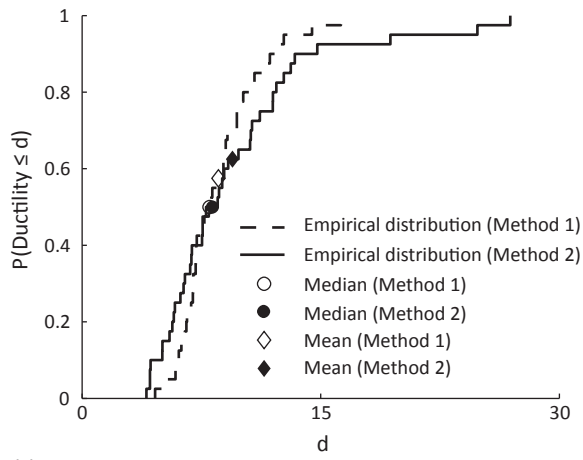




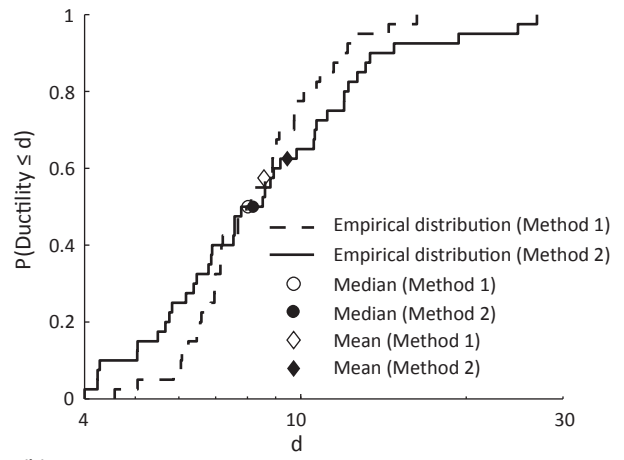
(a)



(b)



(a)



(b)

