An Improved Algorithm for Selecting Ground Motions to Match a Conditional Spectrum

Jack W. Baker and Cynthia Lee

Department of Civil & Environmental Engineering, Stanford University, USA

This paper describes an algorithm to efficiently select ground motions from a database while matching a target mean, variance and correlations of response spectral values at a range of periods. The approach improves an earlier algorithm by Jayaram et al. (2011). Key steps in the process are to screen a ground motion database for suitable motions, statistically simulate response spectra from a target distribution, find motions whose spectra match each statistically simulated response spectrum, and then perform an optimization to further improve the consistency of the selected motions with the target distribution. These steps are discussed in detail, and the computational expense of the algorithm is evaluated. A brief example selection exercise is performed, to illustrate the type of results that can be obtained. Source code for the algorithm has been provided, along with metadata for several popular databases of recorded and simulated ground motions, which should facilitate a variety of exploratory and research studies.

1 Introduction

Selection of ground motions is a topic of great interest as dynamic structural analysis, which requires ground motions as inputs, grows more prevalent (Katsanos et al. 2010; NIST 2011). This selection typically involves searching a ground motion database to find time series produced under appropriate seismological conditions (e.g., earthquake magnitude and source-to-site distance), and that have appropriate response spectral values. In some cases, ground motions are selected based on their individual match to a target spectrum; that is, an optimal set of ground motions would have spectra that all perfectly match the target spectrum. In other cases, however, it is important that the ground motions have variability in response spectra that accurately represents target distributions from predictive models (e.g., Kramer and Mitchell 2006; Lin et al. 2013b). As such, a number of algorithms have been proposed to select ground motions with some form of specified response spectral variability (Bradley 2012; Ha and Han 2016a; b; Jayaram et al. 2011; Kottke and Rathje 2008; Wang 2011). Among those algorithms, only

Bradley, Ha and Han (2016b) and Jayaram et al. include two features of interest here: accounting for correlations among spectral parameters and conditioning on a target spectral acceleration amplitude.

Traditional practice in active seismic regions has been to search databases of ground motion recordings, but simulated ground motions are receiving increased use. Further, there is a need for comparative research studies where recorded and simulated motions are selected in a comparable manner and their relative impacts on structures are evaluated (e.g., Galasso et al. 2013; Iervolino et al. 2010). In recognition of these trends, data facilitating searches of several popular libraries of recorded and simulated ground motions are provided with this algorithm. A second trend in ground motion libraries is that they are rapidly growing larger (several databases discussed below have more than 10,000 ground motions), making the computational efficiency of search algorithms more important.

This manuscript describes an updated version of the algorithm proposed by Jayaram et al. (2011), also utilizing aspects of Bradley (2012). Relative to the Jayaram et al. algorithm, the range of selection options has been broadened and the numerical implementation has been improved to both reduce runtime and improve the statistics of the resulting selected motions. Improvements relative to the previous algorithm are noted below, and improvements in numerical efficiency are also reported.

2 Target Response Spectra

2.1 Types of Spectral Targets

Before discussing the ground motion selection procedure, we first introduce some relevant terminology and concepts related to response spectra as targets for ground motion selection. Ground motion models (GMMs) (e.g., Boore et al. 2014) provide the mean and standard deviation of logarithmic spectral acceleration (*Sa*) at a given period, denoted here as $\mu_{\ln Sa}(Rup,T)$ and $\sigma_{\ln Sa}(Rup,T)$, respectively. With this notation, μ denotes a mean, and σ denotes a standard deviation, of the variable noted in subscript. *Rup* denotes the rupture scenario (defined by the earthquake's magnitude, distance, rupture mechanism, and other parameters necessary to evaluate a given GMM) and *T* denotes the spectral acceleration period. The GMM prediction also generally depends upon one or more parameters defining site conditions such as average shear-wave velocity over the top 30 m of the site (V_{s30}), but that explicit dependence is omitted from this notation for brevity. Some GMMs (e.g., Abrahamson et al. 2014) also provide correlation coefficients for log spectral accelerations at pairs of periods, denoted here as $\rho(T_i, T_j)$. If not provided by the GMM, the correlation coefficients can be obtained from a supplemental model (e.g., Baker and Jayaram 2008).

With the above inputs, we define an "Unconditional Spectrum" as the probability distribution of a response spectrum, given a rupture scenario. The distribution of log spectral acceleration values at multiple periods, given a rupture, is well represented by a multivariate normal distribution (Jayaram and Baker 2008), which is fully specified by the mean and covariance matrix for ln*Sa* values

$$\mathbf{M} = [\mu_{\ln Sa}(Rup, T_1)\ \mu_{\ln Sa}(Rup, T_2)\ \ldots\ \mu_{\ln Sa}(Rup, T_p)]^T \tag{1}$$

$$\Sigma = \begin{bmatrix} \sigma_{T_1}^2 & \sigma_{T_1,T_2} & \cdots & \sigma_{T_1,T_p} \\ \sigma_{T_2,T_1} & \sigma_{T_2}^2 & & \vdots \\ \vdots & & \ddots & \\ \sigma_{T_p,T_1} & \cdots & & \sigma_{T_p}^2 \end{bmatrix} \tag{2}$$

where $\mathbf{M}$ is a vector of mean values of ln*Sa* at p periods of interest, superscript T denotes a matrix transpose, and Σ is the covariance matrix for ln*Sa* at these same periods. In equation 2 we adopt abbreviated notation, $\sigma_{T_i,T_j} = \rho(T_i, T_j)\sigma_{\ln Sa}(Rup, T_i)\sigma_{\ln Sa}(Rup, T_j)$, to denote the covariance of ln*Sa* at periods T_i and T_j (and $\sigma_{T_i}^2 = \sigma_{\ln Sa}(Rup, T_i)^2$ is the variance at period T_i).

The "Unconditional" terminology is used here to emphasize the lack of conditioning on a spectral value, for consistency with the use of the term "Conditional" in the following two definitions. An example Unconditional Spectrum is illustrated in Figure 1a. The mean value from equation 1, and the standard deviations embedded in equation 2, are plotted in Figure 1a; the period-to-period correlation embedded in equation 2 is apparent in the ground motion spectra plotted in the figure, which are 'bumpy' (reflecting a lack of perfect correlation) but do vary with some continuity from period to period.

The "Conditional Mean Spectrum" (CMS) quantifies mean log spectral acceleration values of a ground motion, conditional on a spectral value at a conditioning period and a rupture scenario

$$\mu_{\ln Sa(T_i)|\ln Sa(T^*)} = \mu_{\ln Sa}(Rup, T_i) + \rho(T_i, T^*)\varepsilon(T^*)\sigma_{\ln Sa}(Rup, T_i) \tag{3}$$

where T^* denotes the conditioning period and $\varepsilon(T^*)$ is a residual quantifying the difference between the conditioning *Sa* value ($Sa(T^*)$) and its mean value given the considered rupture

$$\varepsilon(T^*) = \frac{\ln Sa(T^*) - \mu_{\ln Sa}(Rup,T^*)}{\sigma_{\ln Sa}(Rup,T^*)} \tag{4}$$

A CMS, as calculated using equation 3 and a conditioning period of 1.5s, is illustrated in Figure 1b. The term CMS was introduced by Baker and Cornell (2006), and further background is provided in Baker (2011). It is becoming more commonly used to select ground motions for dynamic analysis in several design guidelines (e.g., BSSC 2015; FEMA 2012; TBI Guidelines Working Group 2010).

The "Conditional Spectrum" (CS) is the probability distribution of log spectral acceleration values, conditional on a spectral value at a conditioning period and on a rupture scenario. Unlike the CMS, this spectrum quantifies variability in spectral values at periods other than the conditioning period. If we assume that the distribution is multivariate normal (which is generally reasonable), then the Conditional Spectrum is fully described by conditional means, and a conditional covariance matrix. The conditional means are given by equation 3 and the conditional covariance matrix is

$$\Sigma_{cond} = \Sigma - \frac{\Sigma_{cross}\Sigma_{cross}^{T}}{\sigma_{\ln Sa}(Rup,T^*)^2} \tag{5}$$

where Σ is the covariance matrix from equation 2 and Σ_{cross} is a *p* x 1 matrix of covariances between $\ln Sa(T_i)$ and $\ln Sa(T^*)$. Visually we can represent this distribution by plotting the mean and +/- one or two standard deviations around the mean, as in Figure 1b. The Conditional Spectrum terminology was coined by Abrahamson and Al Atik (2010), but they represented the CS distribution directly by realizations of the spectra rather than an analytical distribution. The Conditional Mean Spectrum was more popular than the Conditional Spectrum prior to approximately 2010, in large part because there was no simple way to select ground motions matching a Conditional Spectrum—a situation rectified by this manuscript and its predecessor algorithms.

Bradley (2010) extended the CS to consider ground motion parameters other than response spectra, and to consider a more general situation where more than one rupture scenario may contribute to occurrence of ground motions with the target amplitude. Bradley refers to the resulting distribution and selection procedure as a Generalized Conditional Intensity Measure (GCIM) approach. This paper focuses on response spectra and single rupture scenarios for simplicity, but the algorithm could, in principle, be generalized by defining equations 3 and 5 to refer to a general vector of intensity measures and to reflect the impact of multiple rupture scenarios on the target means and covariances. Equations 3 and 5 can also be revised to account for the use of multiple GMMs, consistent with current practice in hazard analysis (Lin et al. 2013a).

To complement the above equations, a few observations may provide intuition about the Conditional Spectrum target illustrated in Figure 1b. First, the response spectra “pinch” to a single point at the conditioning period of 1.5s. Since we have specified this amplitude, there is no variability in the spectra at this period. Second, at other periods there is variability in the spectra and that variability tends to be larger at periods further from 1.5s. This is a result of the correlation between spectral values: periods close to 1.5s have spectra highly correlated to $Sa(1.5s)$, so there is relatively little uncertainty in spectra at these nearby periods, while there is larger spectral variability at the (less-correlated) periods far from 1.5s. This pattern in spectral variability is grossly similar to what is observed if one simply scales a set of ground motions so their spectra are equal at some conditioning period, which somewhat confirms the reasonableness of this target. Third, the mean of the conditional spectrum (i.e., the CMS) reflects the expected response spectral shape, and it accounts for both the spectrum associated with the rupture scenario (via the unconditional mean spectrum of equation 1) and the tendency for high-amplitude spectral values to be associated with a peak in the spectrum (via the epsilon value and spectral correlations in equation 3).

It should be intuitive that mean responses obtained from structural analysis are related to the mean amplitude of the input ground motions’ spectra. Further, several studies have shown that considering the full variability in response spectra, rather than only mean values of spectra, can be important for some structural response assessment procedures (e.g., Lin et al. 2013b). This motivates the development of tools like that proposed here to facilitate selection of ground motions matching an Unconditional Spectrum or Conditional Spectrum.

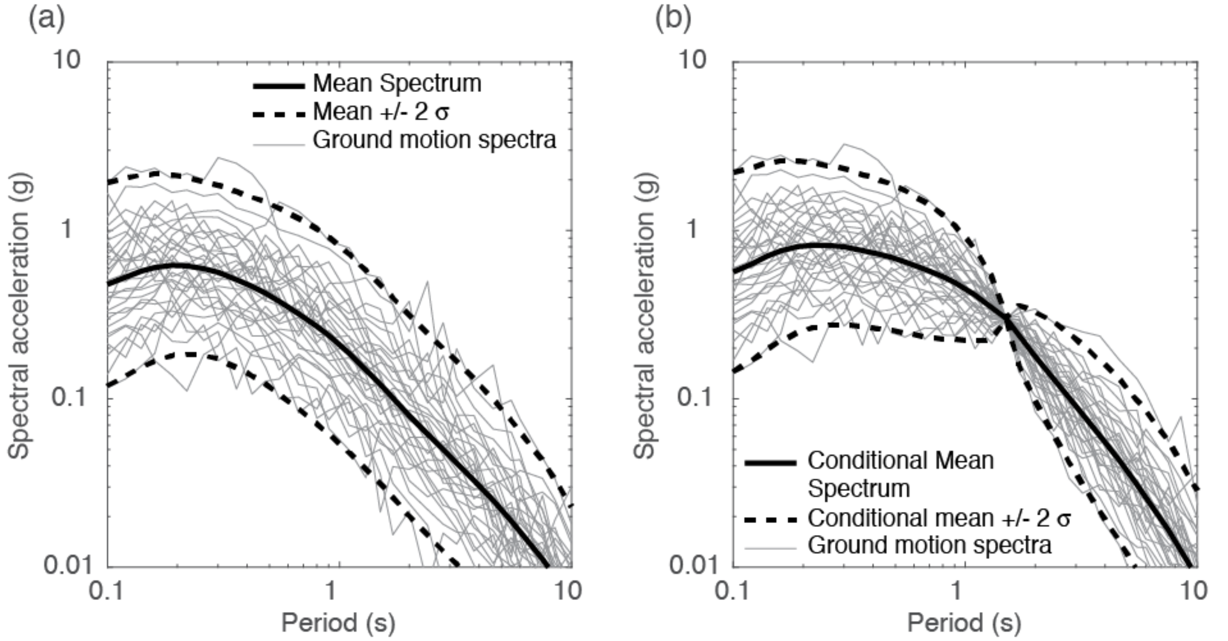

Figure 1. (a) Unconditional Spectrum associated with magnitude = 6.5, distance = 10 km, V_{s30} = 500 m/s and a California strike-slip rupture. (b) Conditional spectrum associated with the same rupture parameters as (a), and with *Sa*(1.5s) = 0.3 g. Response spectra consistent with the target distributions are also shown. Calculations use the GMM of Boore et al. (2014) and the correlation model of Baker and Jayaram (2008).

2.2 Computational Challenges

With the above methods for quantifying response spectra targets established, here we briefly consider ground motion selection strategies (thinking of the case where one wishes to select *n* ground motions from a database with *m* candidate motions). It is simple to quickly select ground motions to be consistent with a Conditional Mean Spectrum, as only a mean spectrum is relevant: one can simply compute an error metric for each of the *m* candidate motions (e.g., the sum of squared errors between the ground motion's spectrum and the target spectrum, potentially after scaling the motion), and then selects the *n* motions with the smallest error. There may be a benefit to performing a more extensive optimization-based search as discussed below, but the availability of this fast approach makes the optimization-based approach less critical.

It is much more difficult to select ground motions that match a Conditional Spectrum or Unconditional Spectrum target, because the ground motions' spectra should match both a mean target and a covariance matrix. To find ground motions with an appropriate covariance matrix, one cannot evaluate individual candidates but instead must evaluate a set of *n* candidates collectively (that is, it is not possible to determine whether an individual ground motion is a "good" fit to a Conditional Spectrum without knowing the other ground motions it would be paired with). This means that there are *m*-choose-*n* combinations of ground motions to consider—too many to search exhaustively for typical situations where n>10 and m>100. It is this problem that the algorithm below addresses, via heuristics to enable a fast search that consistently produces ground motion sets closely matching the specified targets.

3 Selection Algorithm

The major steps of the proposed ground motion selection process are illustrated in Figure 2. These steps are described in additional detail in the following subsections, which group the steps into distinct conceptual stages.

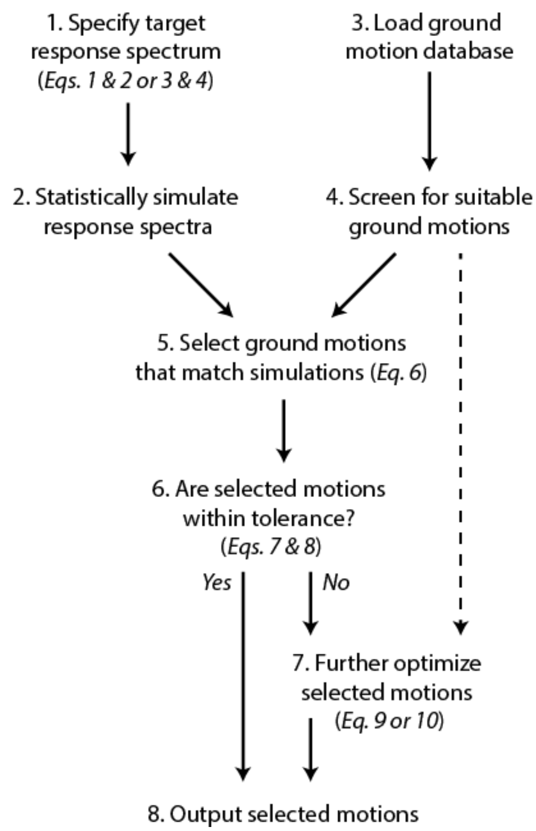

Figure 2. Flow chart of major steps in the ground motion selection process, with relevant equation numbers noted in parantheses. Details for each step are discussed in Section 3.

3.1 Compute Target Spectrum and Statistically Simulate Realizations

The process starts by specifying a target response spectrum (Step 1 in Figure 2). Formally, we specify the mean values of log spectral acceleration, and the covariance matrix for these values. Equations 1 and 2 are utilized if the target is an Unconditional Spectrum, and equations 3 and 5 are utilized if the target is a Conditional Spectrum. The provided software includes a function that computes the required mean and covariance matrix if the user specifies the target earthquake rupture (and the target $\varepsilon(T^*)$ or $Sa(T^*)$ if the target is a Conditional Spectrum). The provided software utilizes the models of Boore et al. (2014) and Baker and Jayaram (2008) to compute the target spectra, but these can be replaced without significant modification to the code if desired.

While the software was developed to solve the problem of selecting ground motions to match a Conditional Spectrum, it can also be adopted for selecting ground motions to match a code spectrum or some other target, by specifying the target spectrum as the mean spectrum, and setting the covariance matrix to consist of all zeros (i.e., specifying that no variability around the target spectrum is desired).

Relative to Jayaram et al. (2011), there are a few updates to the target response spectrum calculation. The main program has been generalized so that a single function can handle Conditional Spectrum and Unconditional Spectrum targets (previously, separate versions of the software were provided for each target type). Additionally, functionality has been provided so that the user can easily match a target $Sa(T^*)$ with a given set of rupture values. Because mean rupture values (rather than the full distribution of possible ruptures considered in a seismic hazard calculation) are often used in these calculations for convenience, the target $Sa(T^*)$ is not necessarily obtained when these mean values are combined with a mean $\varepsilon(T^*)$ value from a hazard deaggregation. An optional calculation now adjusts the $\varepsilon(T^*)$ value so that the conditional mean spectrum matches the target $Sa(T^*)$, as this was seen by Lin et al. (2013) to be a reasonable approximation strategy in many cases.

Step 2 in Figure 2 is to statistically simulate realizations of response spectra from the target distribution. This is done by sampling from a multivariate normal distribution with the target mean and covariance matrices. Since this simulation step is extremely fast, it is performed multiple times and the set of simulations best matching the target spectrum is utilized for the following steps. We note here that the ‘statistically simulated spectra’ in this step are produced by sampling from a probability distribution (e.g., Stein 1987); this is distinct from the ‘simulated ground motions’ discussed in the following section, where are produced by numerical evaluation of equations associated with the earthquake rupture and seismic wave propagation process.

3.2 Specify Candidate Ground Motions

Step 3 of the process in Figure 2 specifies candidate ground motions to select from. Relevant metadata from a candidate ground motion database is loaded, including spectral acceleration values and rupture parameters for each ground motion. The Jayaram et al. (2011) code included metadata for the NGA-West1 database, consisting of 3551 ground motions from 173 earthquakes (Chiou et al. 2008). Here we have added metadata for the NGA-West2 database, which includes 21,539 ground motions from 599 earthquakes (Ancheta et al. 2014). Additionally, we have added metadata for three databases of numerically simulated ground motions produced by a Southern California Earthquake Center (SCEC) project to validate simulations (Goulet et al. 2015). Simulations were produced on the SCEC Broadband Platform using rupture geometries from seven recent California earthquakes. Ground motions from the "EXSIM" (Atkinson and Assatourians 2015), "GP" (Graves and Pitarka 2015) and "SDSU" (Olsen and Takedatsu 2015) simulation algorithms were compiled for use in this software. Each database includes 13,400 ground motions. Because both recorded and simulated ground motion databases are provided in a compatible format, the authors hope that this tool will facilitate further comparative evaluations of similarities and differences in structural demands caused by recorded versus simulated ground motions with comparable response spectra.

An additional improvement in Steps 2 and 3 of the selection process is that the new target computations and ground motion databases utilize both the RotD50 and RotD100 direction-independent metrics of response spectra for multi-component motions (Boore 2010). These metrics are now used often in ground motion models and engineering analysis procedures (Stewart et al. 2011), so their inclusion in the database metadata increases the tool's relevance. Single-component response spectra are also provided so that users can search for single-component motions if desired.

Once database metadata has been loaded, it is screened in Step 4 so that only appropriate ground motions are considered for selection. The current code is set up to allow only ground motions with appropriate values of earthquake magnitude, source-to-site distance, and V_{s30}, but the screening can be easily generalized to consider other properties. These so-called causal parameters are important to screen in order to assure that the considered time series are reasonably consistent with the conditions of interest in ground motion selection, but they should not be screened so aggressively that an insufficient number of candidate motions remain for the next stage of selection (Tarbali and Bradley 2016). The Jayaram et al. (2011) code did not include this screening step, as its objective was to illustrate other aspects of the selection procedure, but the screening has been added here both to improve the quality of the selected motions and to improve the computational cost of the calculation (since motions excluded at this stage need not be considered later for selection).

3.3 Ground Motion Selection

Step 5 of Figure 2 involves selecting ground motions from the database that best match the statistically simulated spectra. For each statistically simulated spectrum and candidate ground motion, the sum of squared errors (*SSE*) is computed

$$SSE = \sum_{j=1}^{P} \left(\ln S_a(T_j) - \ln S_a^{(s)}(T_j) \right)^2 \tag{6}$$

where $\ln S_a(T_j)$ is the log spectral acceleration of the (optionally scaled) candidate ground motion and $\ln S_a^{(s)}(T_j)$ is the $\ln S_a$ of the considered statistically simulated response spectrum. Note that if scaling is not allowed and a target Conditional Spectrum is used, the selected motions will not exactly match the target $Sa(T^*)$, but equation 6 will encourage selection of motions close to the target and the motions may be suitably similar if choosing from a database having ground motions compatible with the target scenario. For each statistically simulated spectrum, the *SSE* is computed for all candidate ground motions that have not already been selected, and the motion with the smallest *SSE* is selected to represent that simulation. The metric of equation 6 is not the only possible selection criterion (e.g., Beyer and Bommer 2007; Buratti et al. 2010), but has been observed to produce satisfactory results; it could easily be modified by a user if desired (e.g., to put varying weights on the squared errors at varying periods).

Simulating spectra from the target distribution (in Step 2), and then searching individual motions to find matches to these simulations, is perhaps the most important step in this algorithm for overcoming the computational cost that would be required to search suites of ground motions instead of individual motions. Its utility is apparent when noting that most prior algorithms to solve this problem have used this approach (Bradley 2012; Jayaram et al. 2011; Wang 2011), though Ha and Han (2016a) recently proposed a non-simulation-based approach that instead uses a limited search of the potential selection combinations.

In Step 6 of Figure 2, the selected suite of motions is evaluated to see whether it is sufficiently close to the target distribution. The maximum percentage mismatch of the mean and standard deviation of the selected motions' spectra, relative to their targets, are calculated

$$Err_{mean} = \max_j \left(\frac{\left| m_{\ln Sa}(T_j) - \mu_{\ln Sa}(T_j) \right|}{\mu_{\ln Sa}(T_j)} \right) \times 100 \tag{7}$$

$$Err_{std} = \max_j \left(\frac{\left| s_{\ln Sa}(T_j) - \sigma_{\ln Sa}(T_j) \right|}{\sigma_{\ln Sa}(T_j)} \right) \times 100 \tag{8}$$

where $m_{\ln Sa}(T_j)$ is the sample mean of the ln*Sa* values of the selected motions at period T_j and $\mu_{\ln Sa}(T_j)$ is the target mean from a GMM (in the case of Unconditional Selection) or equation 5 (in the case of Conditional Selection). Similarly, $s_{\ln Sa}(T_j)$ is the sample standard deviation and $\sigma_{\ln Sa}(T_j)$ is the target standard deviation at period T_j. Finally, $|\ |$ denotes an absolute value. The user can specify a maximum tolerance for the errors defined by equations 7 and 8, and if the errors at this step are less than the tolerance then the selection process is complete.

If the errors are too large, then a finite number of optimization rounds are performed to further improve the selection. Step 7 of Figure 2 involves further optimizing the initial selection if needed. At this stage, the selected set of ground motions are modified by replacing individual ground motions from the set with available motions from the screened database and seeing whether the set is improved in its match to the target response spectrum. There are two objective functions available to the user when performing the optimization.

In the first case, a weighted sum (over all periods of interest) of squared errors in the spectra's mean values and standard deviations are utilized to evaluate goodness of fit, as follows

$$SSE_s = \sum_{j=1}^{p} \left[\left(m_{\ln Sa}(T_j) - \mu_{\ln Sa}(T_j) \right)^2 + w \left(s_{\ln Sa}(T_j) - \sigma_{\ln Sa}(T_j) \right)^2 \right] \tag{9}$$

where SSE_S denotes the sum of squared errors of the set of ground motions, w is a user-defined weight that assigns relative importance to mismatches in mean versus standard deviation values.

In the second case, the *d* statistic from a Komogorov-Smirnov goodness of fit test (KS test) is used as the metric. The *d* statistic at a given period is given by the maximum absolute values of the difference between the target distribution's cumulative distribution function ($F_{\ln Sa(T_j)}(x)$) and the empirical cumulative distribution function ($F^{*}_{\ln Sa(T_j)}(x)$) given by the selected motions' $Sa(T_j)$ values

$$d(T_j) = \max\left(\left| F^{*}_{\ln Sa(T_j)}(x) - F_{\ln Sa(T_j)}(x) \right| \right) \tag{10}$$

Example d values are illustrated in Figure 3 for a single period. In the software, the d values for all periods of interest are summed and used as the error metric. This metric has been successfully utilized by others in ground motion selection problems (Bradley 2012; Chandramohan et al. 2016), and so is incorporated into this algorithm to take advantage of this insight.

Figure 3 illustrates the implications of these error metrics using a target distribution for Sa(2s), and Sa(2s) values for two hypothetical sets of candidate ground motions. The target and sample means are labeled on each figure, as well as the d(2s) value; sample standard deviations (the third parameter used in error computations) are not shown. In Figure 3a, the candidate motions have nearly the same mean and standard deviation as the target spectrum, and a d(2s) value of 0.14. In Figure 3b the candidate ground motions have a sample mean that is 4% larger than the target mean (and, although not shown graphically, the sample standard deviation closely matches the target). However, the d(2s)=0.07 value in this case is half of that in Figure 3a. So the SSE_S error metric would prefer the motions in Figure 3a while the KS test metric would prefer the motions in Figure 3b (keeping in mind that the selection algorithm sums errors across multiple periods rather than considering only a single period as we have here). An additional relevant factor is that the KS test calculation is somewhat slower than the SSE_S calculation, as will be seen below. Ultimately it is left to the user to select a metric for a given selection case, as both error metrics have merits.

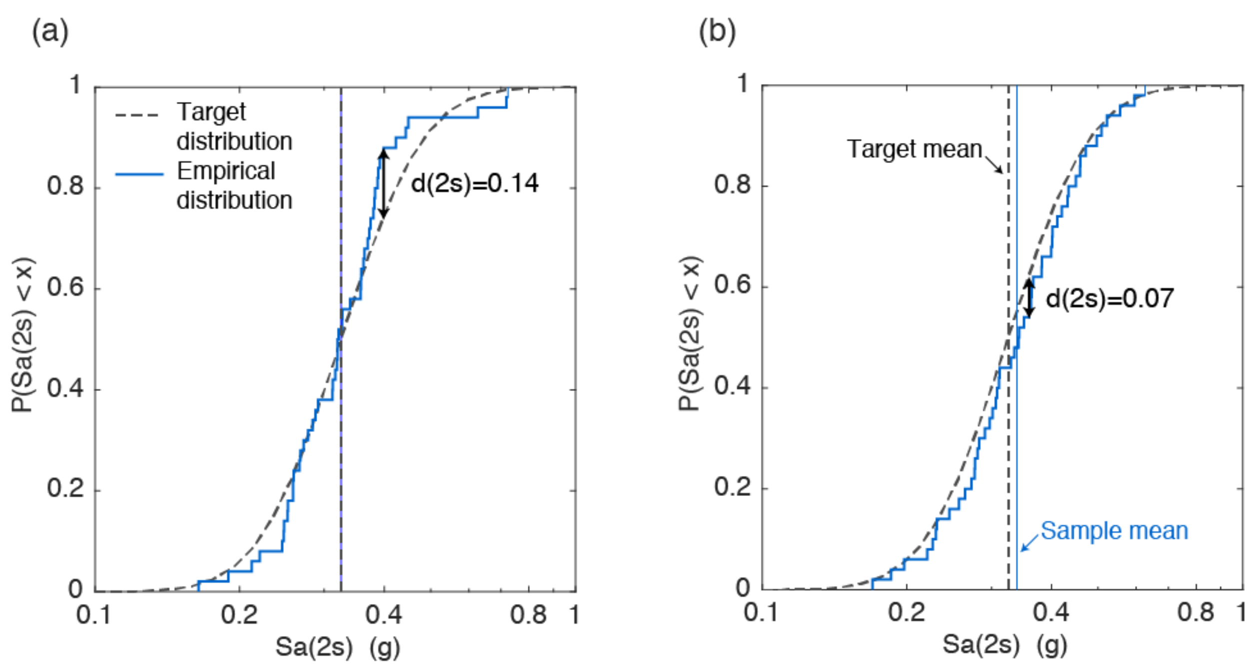

Figure 3. Illustration of spectral acceleration target distributions and error metrics. (a) Selected ground motions whose mean and standard deviation closely match the target. (b) Selected ground motions whose

mean is slightly less than the target mean, but who more closely match the target with regard to the *d* statistic.

This a greedy optimization algorithm, as it searches for only local improvements (i.e., by replacing only one ground motion at a time), and thus misses opportunities for improvements resulting from replacing two or more ground motions that cause an improvement in aggregate (if the improvement was not detectable when replacing them one at a time). While such opportunities surely exist, it is not clear that considering them would result in dramatically improved selection results. Further, this choice to use a greedy optimization approach makes the calculation computationally tractable. The optimization algorithm can terminate early if the error metrics of equations 7 and 8 are within tolerance.

Once a final set of ground motions is determined, in Step 8 of Figure 2 an output file is produced to document the selected ground motions (and scale factors, if scaling was allowed). For the NGA-West1 and simulated ground motions, the ground motion time series are available for direct download over the Internet and the selected motions' URLs are provided in the output so that the time series can be obtained automatically. For the NGA-West2 database, the index numbers of the selected ground motions must be copied into the NGA-West2 search tool in order to download the ground motion files. Additional detail regarding this download process is provided in the software documentation.

4 COMPUTATIONAL EXPENSE

The algorithm's initial record-by-record selection (Step 5) takes time proportional to m x n, where m is the number of candidate ground motions in the database and n is the number of selected ground motions, because the m motions are compared once each to the n simulations. The optimization (Step 7) also takes time proportional to m x n, because the m-n candidates are evaluated as candidates to replace each of the n previously selected motions. This is much better than the m-choose-n computational expense of the exhaustive search discussed in section 2.2. Any calculations within the loops over the m x n candidate evaluations were optimized to limit computational cost to the extent possible.

We further manage computational expense in a few ways. First, we screen the database (Step 4 in Figure 2) to limit the size of m before starting the search for motions to select. While this is conceptually simple, it was not implemented in the Jayaram et al. (2011) software. Second, within the optimization stage, we skip all ground motions that need to be scaled by a larger-thanallowable factor, before proceeding to the more expensive calculation of considering the ground motion as a potential replacement for a currently selected motion. These first two steps often lead to great reductions in the numbers of considered ground motions when typical restrictions on acceptable ground motions and scale factors are used. Third, we stop the optimization early if a selected set of ground motions is sufficiently close to the target spectrum (as evaluated using a user-specified error tolerance). Finally, the optimization step of the algorithm—the most expensive step—is optionally parallelized so that each currently selected ground motion is evaluated in parallel to see if a better alternative can be found.

The computational cost of the algorithm is illustrated in Figure 4, for varying sizes of selected ground motion sets and varying sizes of the database being searched. A few observations can be made. First, the improvements discussed here have greatly reduced the algorithm's run time relative to the previous implementation by Jayaram et al. (2011). A somewhat typical problem of selecting 20 ground motions from a pool of 2000 candidates now requires less than 30 seconds. Second, the cost of the algorithm now scales approximately linearly with the number of selected ground motions (Figure 4a) and the number of searchable ground motions (Figure 4b), as expected based on the discussion above.

The Figure 4 results are produced without allowing early termination of the optimization and without parallelized optimization, in order to provide conservative run times, so users may find better performance depending upon their use of these features. The benefits of these features are problem dependent, so no general run time results are provided here. Further speed-up of the algorithm is possible, most obviously by switching to a faster programing language, and by restructuring some code for better numerical performance. We consider the current implementation to be well suited, however, for our goals of providing an educational code with reasonable run times.

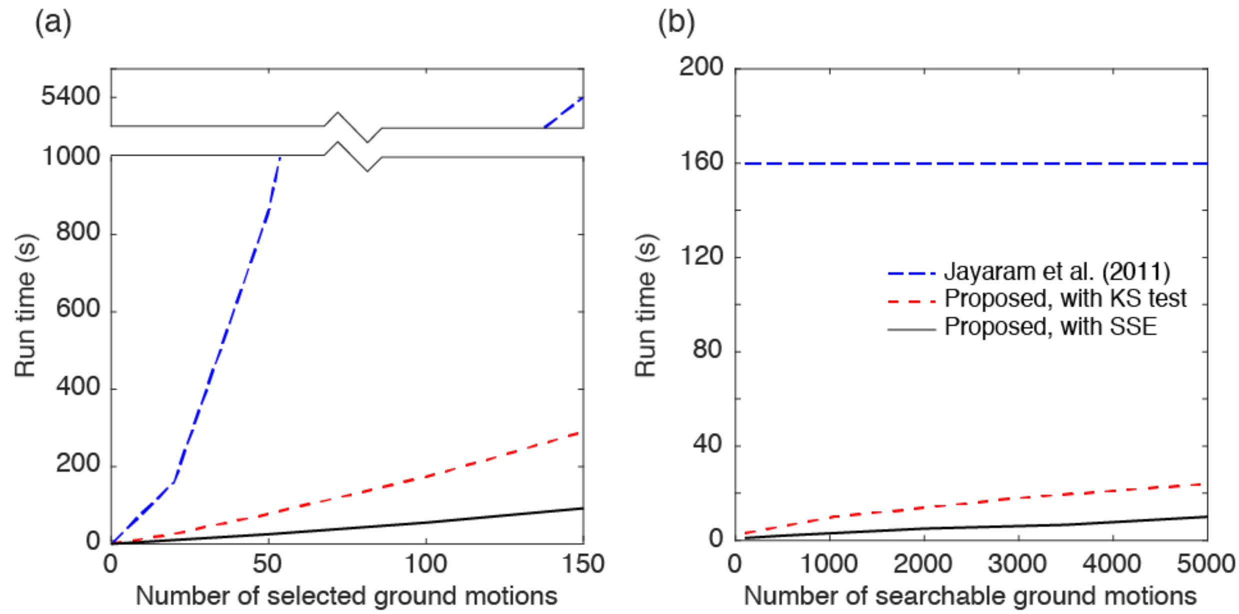

Figure 4. (a) Time to select a given number of ground motions from a database of 5000 motions, for three different selection approaches. (b) Time to select 20 ground motions from a database of 5000, with varying numbers of motions remaining after screening.

5 Example Ground Motion Selection

To illustrate use of the software, selection of simulated and recorded motions is briefly demonstrated, referring to the numbered steps in Figure 2. In step 1, the target spectrum must be specified; the Conditional Spectrum of Figure 1b is used here as the target RotD50 spectrum. Step 2 statistically simulates realizations of spectra from this distribution, and no user choices need to be made. In step 3, the ground motion database must be specified; here we consider two alternatives—the NGA-West1 database of recorded ground motions and the Graves and Pitarka database of simulated motions discussed above. In step 4 these databases are screened for suitable ground motions. In this case, because the target spectrum is associated with a magnitude = 6.5, distance = 10 km event, we restricted selection to consider only ground motions within 50 km of an earthquake with magnitude between 6 and 7. Scaling of ground motions was allowed, but scale factors were limited to a maximum of five. These criteria are somewhat consistent with typical ground motion selection procedures (e.g., NIST 2011), but are used here simply to illustrate the selection process and are not recommended values. With these criteria, the NGA-West1 database had 582 ground motions satisfying the initial screening, and the GP database had 6000.

With the initial criteria and database screening performed, ground motions can then be selected. In step 5, initial ground motions were selected from each database to match simulated spectra—a step which requires no user choices. In step 6, a maximum 10% error in means and standard deviations of the spectral was specified. Neither database satisfied that criterion with the initial selection (the step 6 check), so optimization (step 7) was performed in both cases. The SSE objective function of equation 9 was chosen for the optimization. For the NGA-West1 database, the max errors in mean and standard deviation were 8% and 34%, respectively, before optimization and 6% and 7% after. For the Graves and Pitarka database, the max errors were 41% and 32% before optimization and 29% and 26% after.

In step 8, the final selections of ground motions are output. Figure 5 shows the response spectra of the selected motions. We see that, although the Graves and Pitarka selection had significantly worse error metrics in the previous paragraph, the response spectra plot does not indicate serious deficiencies in general (the periods with poor metrics are the short periods at the left of Figure 5b); these errors would need to be evaluated on a project-specific basis for acceptability. Figure 6 shows example velocity time series for both sets of selected motions. The two sets of selected motions' time series are comparable under a cursory visual inspection. These two sets of ground motions would provide useful inputs for a more detailed study of any subtler differences in the recordings and simulations that lead them to produce different structural responses when used as input to a dynamic analysis problem. Such a study is beyond the scope of this paper but is the type of exercise that the new software features are intended to facilitate.

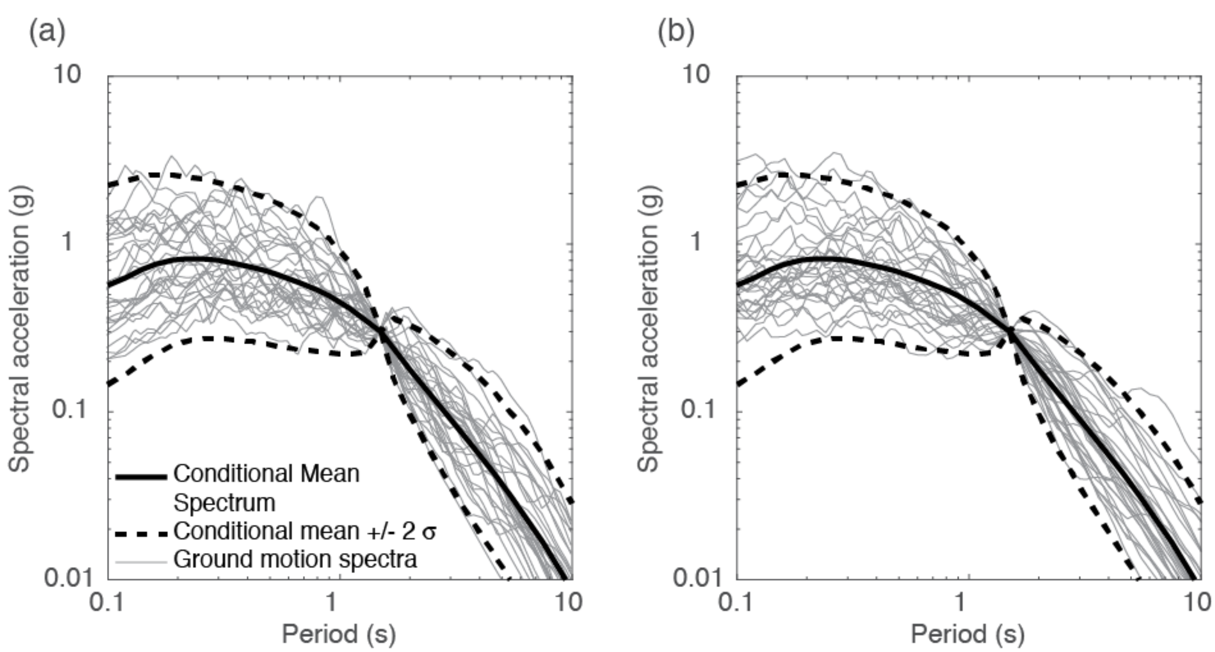

Figure 5. Response spectra of selected ground motions with spectra matching the target Conditional Spectrum of Figure 1b. The mean and mean +/- two standard deviations of the target ln*Sa* distribution are superimposed. (a) NGA-West1 recorded ground motions. (b) Graves and Pitarka simulated ground motions.

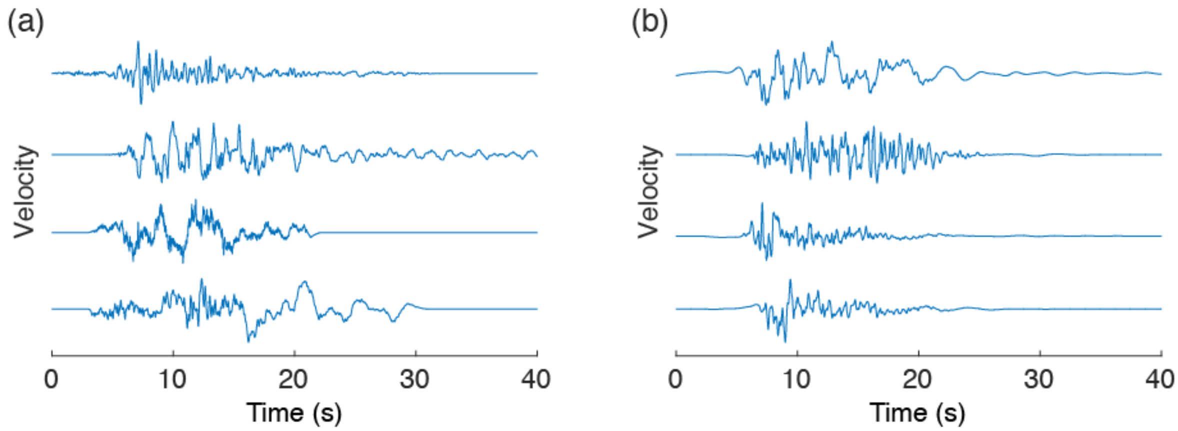

Figure 6. Example velocity time series from the selected motions shown in Figure 5. The time series are scaled to have the same peak ground velocity values for ease of comparison, so no velocity scale is provided. (a) NGA-West1 recorded ground motions. (b) Graves and Pitarka simulated ground motions.

6 Numerical Implementation

Matlab source code for the algorithm is available at https://github.com/bakerjw/CS_Selection. The repository includes metadata and documentation for five ground motion databases discussed above. In addition to the algorithmic features discussed above, this source code also improves upon the previous code from Jayaram et al. (2011) by providing a single general purpose program (the Jayaram et al. code consistent of four separate programs for conditional versus unconditional selection, and single-component versus two-component selection), and better modularization of functions related to the tasks outlined in Figure 2.

7 Conclusions

This paper has presented an efficient algorithm for selecting ground motions from a database that match a target response spectrum distribution (i.e., a Conditional Spectrum or Unconditional Spectrum). The motivation for this work is that when the target spectrum has a distribution, rather than a single value, it is not possible to evaluate individual ground motions for selection without considering them as part of a suite of ground motions that collectively represent the distribution. But evaluating all possible suites of ground motions is impossible when considering large ground motion databases typical in practice today. This algorithm utilizes several heuristics to quickly identify ground motion sets with close match to the target spectrum.

Note that the code can also be easily adapted for the more common situation where the user wants to select ground motions that each closely match a target spectrum (e.g., for closely matching a design spectrum from a building code, with each motion closely matching the target) by specifying the target spectrum as the mean value and setting the target variances to zero. The algorithm's complexity is more than is needed for this simpler application, but it is well-suited for the problem and so a user-defined flag setting the variance to zero is provided in the software.

The key steps in this algorithm are (1) compute a target response spectrum distribution, (2) statistically simulate response spectra from the target distribution, (3) load and screen a database of candidate ground motions, (4) select ground motions from the database that individually match the statistically simulated spectra, (5) make incremental changes to the initially selected ground motion set to further optimize its fit to the target spectrum distribution. The algorithm follows the general structure of a proposal by Jayaram et al. (2011), but incorporates a number of new features that improve its utility and speed.

Data for a number of new ground motion databases are also provided, to allow users to search recently developed catalogs of recorded or simulated ground motion data. Example selection of comparable recorded and simulated ground motions are illustrated, to demonstrate the feasibility of selecting equivalent motions from differing sources. Matlab source code for the algorithm has been provided publically for readers interested in using or modifying the algorithm.

Acknowledgements

We thank Nenad Bijelic, Reagan Chandramohan, Beliz Ugurhan and three anonymous reviewers for helpful advice and comments on this work. This work was supported in part by the State of California through the Transportation Systems Research Program of the Pacific Earthquake Engineering Research Center (PEER).

References

Abrahamson, N. A., and Al Atik, L. (2010). "Scenario Spectra for Design Ground Motions and Risk Calculation." *9th US National and 10th Canadian Conference on Earthquake Engineering*, Toronto, Canada, 12.

Abrahamson, N. A., Silva, W. J., and Kamai, R. (2014). "Summary of the ASK14 Ground Motion Relation for Active Crustal Regions." *Earthquake Spectra*, 30(3), 1025–1055.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. (2014). "NGA-West2 Database." *Earthquake Spectra*, 30(3), 989–1005.

Atkinson, G. M., and Assatourians, K. (2015). "Implementation and validation of EXSIM (a stochastic finite-fault ground-motion simulation algorithm) on the SCEC broadband platform." *Seismological Research Letters*, 86(1), 48–60.

Baker, J. W. (2011). "Conditional Mean Spectrum: Tool for ground motion selection." *Journal of Structural Engineering*, 137(3), 322–331.

Baker, J. W., and Cornell, C. A. (2006). "Spectral shape, epsilon and record selection." *Earthquake Engineering & Structural Dynamics*, 35(9), 1077–1095.

Baker, J. W., and Jayaram, N. (2008). "Correlation of spectral acceleration values from NGA ground motion models." *Earthquake Spectra*, 24(1), 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.

Boore, D. M. (2010). "Orientation-Independent, Nongeometric-Mean Measures of Seismic Intensity from Two Horizontal Components of Motion." *Bulletin of the Seismological Society of America*, 100(4), 1830–1835.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M. (2014). "NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes." *Earthquake Spectra*, 30(3), 1057–1085.

Bradley, B. A. (2010). "A generalized conditional intensity measure approach and holistic ground-motion selection." *Earthquake Engineering & Structural Dynamics*, 39(12), 1321–1342.

Bradley, B. A. (2012). "A ground motion selection algorithm based on the generalized conditional intensity measure approach." *Soil Dynamics and Earthquake Engineering*, 40, 48–61.

BSSC. (2015). *NEHRP Recommended Seismic Provisions for New Buildings and Other Structures*. FEMA P-1050, Building Seismic Safety Council, Washington, D.C., 515p.

Buratti, N., Stafford, P. J., and Bommer, J. J. (2010). "Earthquake Accelerogram Selection and Scaling Procedures for Estimating the Distribution of Drift Response." *Journal of Structural Engineering*, 137(3), 345–357.

Chandramohan, R., Baker, J. W., and Deierlein, G. G. (2016). "Impact of hazard-consistent ground motion duration in structural collapse risk assessment." *Earthquake Engineering and Structural Dynamics*, (in press).

Chiou, B., Darragh, R., Gregor, N., and Silva, W. (2008). "NGA Project Strong-Motion Database." *Earthquake Spectra*, 24(1), 23–44.

FEMA. (2012). *Seismic Performance Assessment of Buildings*. FEMA P-58, Prepared by Applied Technology Council for the Federal Emergency Management Agency.

Galasso, C., Zhong, P., Zareian, F., Iervolino, I., and Graves, R. W. (2013). "Validation of ground-motion simulations for historical events using MDoF systems." *Earthquake Engineering & Structural Dynamics*, 42(9), 1395–1412.

Goulet, C. A., Abrahamson, N. A., Somerville, P. G., and Wooddell, K. E. (2015). "The SCEC Broadband Platform Validation Exercise: Methodology for Code Validation in the Context of Seismic - Hazard Analyses." *Seismological Research Letters*, 86(1), 17–26.

Graves, R., and Pitarka, A. (2015). "Refinements to the Graves and Pitarka (2010) Broadband Ground - Motion Simulation Method." *Seismological Research Letters*, 86(1), 75–80.

Ha, S. J., and Han, S. W. (2016a). "A method for selecting ground motions that considers target response spectrum mean and variance as well as correlation structure." *Journal of Earthquake Engineering*, (in press).

Ha, S. J., and Han, S. W. (2016b). "An efficient method for selecting and scaling ground motions matching target response spectrum mean and variance." *Earthquake Engineering & Structural Dynamics*, 45(8), 1381–1387.

Iervolino, I., De Luca, F., and Cosenza, E. (2010). "Spectral shape-based assessment of SDOF nonlinear response to real, adjusted and artificial accelerograms." *Engineering Structures*, 32(9), 2776–2792.

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(5), 2231–2243.

Jayaram, N., Lin, T., and Baker, J. W. (2011). "A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance." *Earthquake Spectra*, 27(3), 797–815.

Katsanos, E. I., Sextos, A. G., and Manolis, G. D. (2010). "Selection of earthquake ground motion records: A state-of-the-art review from a structural engineering perspective." *Soil Dynamics and Earthquake Engineering*, 30(4), 157–169.

Kottke, A., and Rathje, E. M. (2008). "A Semi-Automated Procedure for Selecting and Scaling Recorded Earthquake Motions for Dynamic Analysis." *Earthquake Spectra*, 24(4), 911–932.

Kramer, S. L., and Mitchell, R. A. (2006). "Ground Motion Intensity Measures for Liquefaction Hazard Evaluation." 22(2), 413–438.

Lin, T., Harmsen, S. C., Baker, J. W., and Luco, N. (2013a). "Conditional Spectrum Computation Incorporating Multiple Causal Earthquakes and Ground Motion Prediction Models." *Bulletin of the Seismological Society of America*, 103(2A), 1103–1116.

Lin, T., Haselton, C. B., and Baker, J. W. (2013b). "Conditional spectrum-based ground motion selection. Part I: Hazard consistency for risk-based assessments." *Earthquake Engineering & Structural Dynamics*, 42(12), 1847–1865.

NIST. (2011). *Selecting and Scaling Earthquake Ground Motions for Performing Response-History Analyses*. NIST GCR 11-917-15, Prepared by the NEHRP Consultants Joint Venture for the National Institute of Standards and Technology, Gaithersburg, Maryland.

Olsen, K., and Takedatsu, R. (2015). "The SDSU Broadband Ground-Motion Generation Module BBtoolbox Version 1.5." *Seismological Research Letters*, 86(1), 81–88.

Stein, M. (1987). "Large Sample Properties of Simulations Using Latin Hypercube Sampling." *Technometrics*, 29(2), 143–151.

Stewart, J. P., Abrahamson, N. A., Atkinson, G. M., Baker, J. W., Boore, D. M., Bozorgnia, Y., Campbell, K. W., Comartin, C. D., Idriss, I. M., Lew, M., Mehrain, M., Moehle, J. P., Naeim, F., and Sabol, T. A. (2011). "Representation of Bi-Directional Ground Motions for Design Spectra in Building Codes." *Earthquake Spectra*, 27(3), 927–937.

Tarbali, K., and Bradley, B. A. (2016). "The effect of causal parameter bounds in PSHA-based ground motion selection." *Earthquake Engineering & Structural Dynamics*, (in press).

TBI Guidelines Working Group. (2010). "Guidelines for performance-based seismic design of tall buildings." *Berkeley: University of California (PEER Report No. 2010/05)*.

Wang, G. (2011). "A ground motion selection and modification method capturing response spectrum characteristics and variability of scenario earthquakes." *Soil Dynamics and Earthquake Engineering*, 31(4), 611–625.