14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

Selection of Near-Fault Physics-Based Simulated Earthquake Ground Motions for Seismic Performance Assessment of Structures

Maha Kenawy, Ph.D.
Associate, Exponent, Menlo Park, California, USA

David McCallen, Ph.D.
Professor, Department of Civil and Environmental Engineering, University of Nevada, Reno, USA

ABSTRACT: Selection of earthquake records for seismic performance assessment of near-fault structures is challenging because (1) earthquake-induced ground shaking in the near-fault region is highly sensitive to the fault rupture characteristics, seismic wave propagation patterns, and site conditions, and (2) field recordings of such near-fault shaking are relatively sparse. A common approach to representing the near-fault ground motion in engineering analysis is to explicitly consider records with strong directivity pulses (pulse-type records) in ground motion selection. In this study, we critically examine this approach by using broadband physics-based fault rupture simulations and high-performance computing tools to measure the bias associated with selecting earthquake records for the analysis of near-fault structures. Our findings reveal that the predicted structural demand distributions associated with selected ensembles of unscaled simulated records are not sensitive to the binary record classification based on the presence of strong directivity pulses. These findings support the hypothesis that selection of unscaled site-specific simulated ground motion records based on spectral shape eliminates the need to exclusively consider records classified as pulse-type.

1. INTRODUCTION

Seismic performance assessment of structures typically relies on ground motion models (GMMs) to estimate representative ground motion intensity measures (IMs) from general rupture and site characteristics. The prediction of ground shaking intensity very close to active faults using GMMs may be associated with large uncertainty because of the relative sparsity of the near-fault earthquake records in the available databases (e.g., Ancheta et al. (2014)) and because near-fault ground motion is highly sensitive to the rupture characteristics and site conditions (Kenawy et al., 2021). A common approach to improving the representation of near-fault ground shaking in seismic hazard analysis is to consider the anticipated strong velocity pulses in near-fault records, classify ground motions as either pulse-type or nonpulse based on the amplitude of the largest velocity pulses relative to the rest of the history of ground velocity, and dictate explicit selection of pulse-type ground motions as a portion of the sampled records for nonlinear analysis (Shahi and Baker, 2011; NIST, 2011; Shahi and Baker, 2014). The selection of ground motions with strong velocity pulses for dynamic analysis of near-fault structures is currently recommended in the ASCE 7-16 standard adopted in the U.S., although limited guidance is provided on how to conduct such a selection. The binary classification of near-fault motions into pulse and nonpulse types, and the associated probabilities of encountering strong pulses near the fault, generally rely on empirical procedures developed based on a limited number of historical near-fault records. In addition, the classification of records is typically based on detecting one or two early-arriving velocity pulses that domi-

nate the ground motion time histories, and may misclassify records with multiple strong pulses, such as those resulting from amplification due to sedimentary basins. As a result, the salient characteristics of the damaging near-fault motions may not be appropriately captured by the binary classification into pulse and nonpulse records, and such methodology may fail to incorporate records that are most damaging to a given structure in a nonlinear analysis context, and therefore, underestimate the seismic risks to the structure. Recent studies began to shift away from such an approach, including the study by Zengin and Abrahamson (2021) who proposed a new intensity measure, the maximum instantaneous power $maxIP(T_1)$, for the selection of near-fault records without distinction between pulse-type and nonpulse records. In addition, Tarbali et al. (2019) proposed a methodology for selecting near-fault motions that accounts for the probability of encountering strong pulses into the underlying PSHA, but does not require the classification of records into pulse and non-pulse.

The use of three-dimensional physics-based fault rupture simulations has the potential not only to supplement the available datasets of historical earthquake records, but also to allow the use of site-specific records that closely match the properties of rupture and shaking at the design site; a luxury that is usually not feasible using the recorded field data. The advent of physics-based earthquake simulations may improve the prediction of the seismic demands on near-fault structures, in addition to alleviating other issues such as the need for scaling the relatively limited dataset of historical near-fault records, and the need for explicit consideration of the waveform features of near-fault records. In this study, we critically examine the need for binary classification of near-fault records based on velocity pulse features in the selection of ground motions for structural analysis of near-fault structures. In contrast to previous studies which mostly consider a limited dataset of recorded near-fault ground motion, and use amplitude scaling to match target intensity measures, we use an extensive database of unaltered (unscaled) broadband physics-based simulated ground motions to examine the bias associated with selecting records for the analysis of near-fault structures. We use the Generalized Conditional Intensity Measure (GCIM) approach (Bradley, 2010, 2012), and incorporate an additional step that allows us to compare the seismic demands associated with the selected records to a known reference condition. We use this procedure to answer the question: how well can a selected ensemble of earthquake records represent the important characteristics of near-fault shaking, with and without explicit consideration of the velocity pulse features?

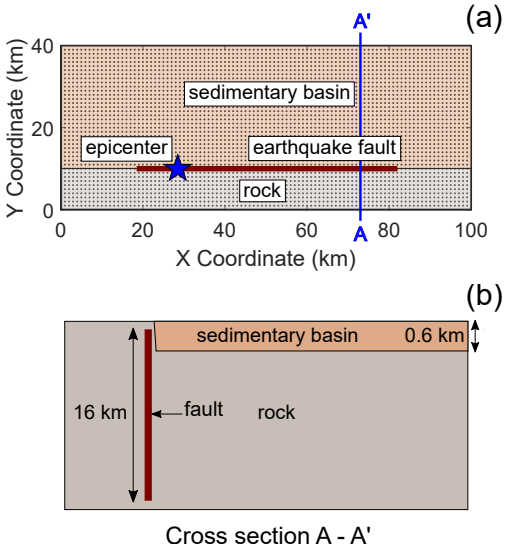

Figure 1: (a) Plan view of the computational domain showing the fault trace, epicenter, surface stations, and relevant geologic features; (b) cross-sectional view showing the basin soil and rock.

2. FAULT RUPTURE AND BUILDING SIMULATIONS

We use fully deterministic broadband (0-5 Hz) physics-based three-dimensional fault rupture realizations representing a M7.0 strike-slip earthquake scenario. The simulations were conducted as part of the the EQSIM (Earthquake Simulation Framework for Physics-Based Fault-to-Structure Simulations) application development under the U.S. Department of Energy Exascale Computing Project (McCallen et al., 2021) using the CORI machine at the National Energy Research Scientific Computing Center. The ground motion acceleration time series were generated using the finite difference

code SW4 (Seismic Waves, fourth order) developed at Lawrence Livermore National Laboratory (Sjögreen and Petersson, 2012).

Table 1: Hypocenter and slip patch locations for the earthquake fault rupture realizations

Rupture Realization	Hypocenter Location	Slip Patch Location
G93	Left	Right - surface
G95	Left	None
G96	Center	None
G97	Center	Right - surface
G98	Left	Right - deep
G99	Center	Right - deep

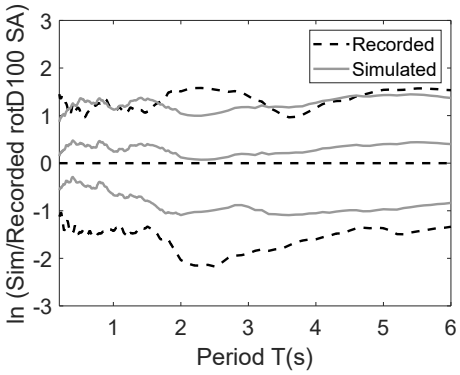

Figure 2: Natural logarithm of the median, 5th percentile and 95th percentile maximum-direction spectral acceleration of the recorded and simulated ground motions within 10 km JB distance normalized by the median spectrum of the recorded dataset.

Simulation of the fault rupture processes relies on the kinematic earthquake rupture modeling technique of Graves and Pitarka; this technique has been validated in simulations of recorded earthquakes (Graves and Pitarka, 2018; Pitarka et al., 2019). The rupture realizations represent shallow crustal earthquakes with a predominantly strike-slip mechanism. The strike-slip fault ruptures were simulated with a fault length of 62.6 km, and fault width of 16 km. The selected realizations have different combinations of hypocenter locations (toward the left or center of the fault), and different slip distribution characteristics, as listed in table 1. With respect to the latter parameter, a subset of the realizations (G93, G97, G98 and G99) contains rupture asperities or patches of high slip deterministically placed in locations along the fault, whereas the remaining realizations (G95 and G96) follow a purely stochastic distribution of the fault slip. The importance of including these rupture asperities is discussed elsewhere (Pitarka et al., 2019; Kenawy et al., 2021). As listed in the table, two different variations of the depth of the large-slip patches are present in the simulations. The computational domain (shown in figure 1) spans 100 km by 40 km by 30 km, with a minimum grid spacing of 8 m, and incorporates three-dimensional geological features and a flat earth surface. The domain contains a sedimentary basin region with an average shear-wave velocity in the top 30 meters (V_{s-30}) equal to 380 m/s, and a rock region with V_{s-30} equal to 500 m/s. Further details about the rupture model and simulation domain are provided in Kenawy et al. (2021). The surface of the domain contains 3,861 stations - spaced at 1 km intervals - at which three ground acceleration history components are generated: a horizontal fault-normal (FN) component, a horizontal fault-parallel (FP) component, and a vertical component.

Two reinforced concrete special moment-frame buildings were designed for seismic category E: a 3-story building and a 12-story building, following modern seismic provisions in the United States. The seismic design is based on the risk-targeted maps for a site that is about two kilometers away from the Hayward fault in Berkeley, California with site class C. Two-dimensional models of the building frames were created using the structural analysis platform Opensees (McKenna et al., 2000). Further details about the structural design and simulation models are available in Kenawy et al. (2021) and Kenawy and McCallen (2021).

3. CHARACTERISTICS OF SIMULATED GROUND MOTIONS

The database of simulated ground motions consists of 23,166 records from the six rupture realizations previously described, all representing a magnitude 7.0 pure strike-slip earthquake scenario. Within a Joyner-Boore (JB) distance of 10 km, the data set includes 9,456 records with an average JB dis-

tance of 5.3 km, and an average V_{s30} of 434.4 m/s. This near-fault subset of simulated ground motions is not intended to replicate a particular recorded earthquake, yet it is expected to reasonably represent the general characteristics of recorded near-fault shaking. Figure 2 displays the natural logarithm of the median, 5th percentile and 95th percentile of the orientation-independent maximum direction (rotD100) pseudo acceleration spectra of the simulated ground motions in addition to an available dataset of field-recorded motions within approximately 10 km JB distance of various ruptures. The orientation-independent measure was selected for the comparison to eliminate the uncertainties associated with rotating the as-recorded ground motions to their FN and FP components. The recorded dataset consists of 84 records which were obtained from the PEER NGA-West2 ground motion database (Ancheta et al., 2014) using the following parameters: magnitude between 6.5 and 7.5 with a mean of 6.84; JB distance between 0 and 11 km with a mean of 4.2 km, and V_{s30} less than 1100 m/s with an average of 390.9 m/s. The NGA dataset includes records from many shallow crustal earthquakes in various regions across the world and different rupture mechanisms. Therefore, the records of this dataset have wider variability than the simulated motions, which is seen in the spectra of figure 2. The figure shows that the median spectrum of the simulated ground motions is consistently higher than that of the recorded motions, although the 95th percentile spectra of the two datasets are in reasonable agreement. The significant difference between the 5th percentile spectrum of both datasets suggests that the ground motion simulations estimate somewhat higher shaking intensities compared to the existing database within 10 km of the fault.

4. SELECTION OF NEAR-FAULT GROUND MOTIONS FROM A DATABASE OF PHYSICS-BASED SIMULATED RECORDS

We measure the bias associated with selecting simulated earthquake records for near-fault engineering analysis. Specifically, we examine the need for binary classification of near-fault records based on their velocity pulse characteristics in the selection process. We design ground motion selection experiments in which the structural demands associated with an ensemble of selected ground motions are compared against the structural demands associated with reference target conditions based on the fault rupture realizations. We establish reference distributions of ground motion intensity measures consisting of pulse-type records only, and attempt to match those distributions using a "mixed" candidate pool of pulse-type and nonpulse records. The goal of this experiment is to examine the potential degree of bias in the estimated structural demand if: (1) the site at the target distance is dominated by ground motion with strong pulses, and (2) the selection procedure fails to explicitly consider the pulse characteristics. The classification of records as pulse-type or nonpulse was performed using the approach proposed by Shahi and Baker (2014).

The underlying selection experiments are based on the GCIM approach. We use scenario-based ground motion selection, meaning that our target is not a conditional spectrum, but rather is IM distributions obtained directly from multiple fault rupture realizations of a single earthquake scenario of interest (Tarbali and Bradley, 2015). This approach allows us to measure the bias in the predicted structural demands associated with each record selection experiment relative to the benchmark structural demands corresponding to the target IM distributions. We leverage the large database of simulated records and eliminate the scaling of records, which is conventionally done as part of ground motion selection and matching. Instead, we solve an optimization function in search for raw records that minimize the differences between a data sample representing the target IM distribution, and the final set of selected records. The important components of our selection methodology are summarized as follows:

- We only use the pseudo-acceleration response spectra as the target intensity measure; we match the distribution of twenty spectral ordinates corresponding to the following periods: 0.01, 0.02, 0.03, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.75, 1, 1.5, 2, 3, 4 and 5 seconds. We assign equal weights $w_i$to each

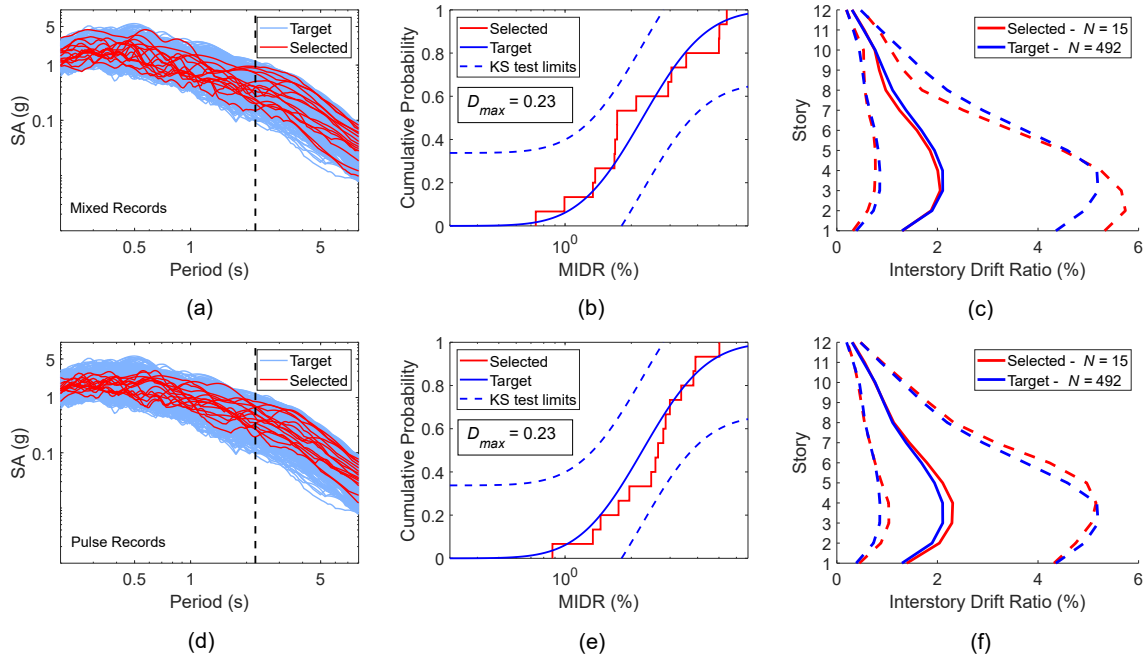

Figure 3: Response spectra, MIDR distribution and IDR envelopes imposed by a selected set of 15 records to match the distributions associated with pulse-type records for a 12-story building at 1 km: (a), (b) and (c) are the results of using a mixed group of records; (d), (e) and (f) are the results of using only pulse-type records. The IDR envelopes represent the median, 5th percentile and 95th percentile envelopes for the target and selected records.

spectral ordinate.

- We assume that all the IMs are lognormally distributed. A target lognormal distribution of an IM at distance R is obtained from all the simulated records at distance R. Ideally, such a distribution should incorporate the anticipated variability due to source, path and site effects. Our database represents some of the aleatory variability associated with the hypocenter location and rupture slip distribution, but is in no way comprehensive of all possible conditions. Because we include every location around the fault that is at a distance R, we indirectly incorporate the variability due to rupture directivity effects. Only two different types of soils are represented in our computational domain; therefore, our analysis may not reflect the dispersion expected in the IMs and structural demands associated with more complex spatial variability in the site conditions.

- We use random sampling to draw target points from the target IM distributions, and evaluate all the available ground motion candidates against the target distribution sample using the square of the error between each target point, and all the available candidate points. The only restriction on the candidate pool of simulated records is that they correspond to JB distances between 0 and 15 km. The normalized squared error associated with *n IMs* is:

$$SE^{kj} = \sum_{i=1}^{n} w_i \left(\frac{ln(IM^{k}_{i-tar}) - ln(IM^{j}_{i-cand})}{\sigma_{ln(IM_{i-tar})}} \right)^2 \quad (1)$$

where $ln(IM^{k}_{i-tar})$ is the natural log of the k^{th} point of the target distribution sample of IM i, $ln(IM^{j}_{i-cand})$ is the natural log of the value of IM i associated with candidate ground motion j, $\sigma_{ln(IM_{i-tar})}$ is the standard deviation of the natural log of the target IM i distribution, and w_i is the normalized weight assigned to each IM type i. Equation 1 is used to obtain records that best match the target distribution sample.

- We assess the goodness of fit between the target IM distributions, and the corresponding distributions associated with the selected records using the Kolmogorov–Smirnov test (KS test). We compute a weighted measure of the goodness of fit for all target IMs, following the recommendation of Bradley (2012). This measure is a global residual for the set of selected records and is computed using the weights w_i and the square of the K-S test statistic D_{IM} as follows:

$$R_{IM} = \sum_{i=1}^{n} w_i (D_{IM_i})^2 \tag{2}$$

- We repeat this sampling process and select the best-fit statistical replicate which yields the lowest global residual R_{IM}.

- Because our target IM distributions correspond to a simulated earthquake scenario, we are able to measure the resulting bias in the structural demands–the maximum interstory drift ratio (MIDR) of each building–corresponding to the final set of selected records: Using the KS test, we compare the resulting MIDR distribution against a cumulative distribution function fitted to the reference MIDR values that correspond to the target records at the distance of interest R. In addition to the MIDR, the envelopes of the interstory drift along the height of each building (IDR) associated with the selected and reference distributions are also compared. To compare the IDR envelopes, the peak IDR values at each story were assumed to follow a lognormal distribution. The median, 5th percentile and 95th percentile IDR values were obtained for the selected and reference ground motion groups, and the differences between the percentiles corresponding to both groups were examined.

5. RESULTS AND DISCUSSION

Figure 3 shows a representative outcome of the selection experiments conducted to match the target distribution of response spectral ordinates associated with a group of pulse-type records using (1) a mixed record pool that does not distinguish between pulse-type and nonpulse records (subplots a through c), and (2) a pool of pulse-type records only (subplots d through f). The figure contains best-fit replicates (across 20 replicates in each case) for a 12-story building at a JB distance of 1 km, and using a selection sample size of 15 records only. Subplots a and d indicate that the selected records from both candidate pools produce response spectra that are consistent with the target response spectra. The MIDR distributions corresponding to the selected and target records for both sets of experiments (subplots b and e) reveal that both selection approaches lead to statistically acceptable ensembles of records. However, the mixed group of records tends to somewhat underestimate the structural demands, whereas the pulse-type group tends to overestimate the structural demands. For example, based on the target MIDR distribution, the cumulative probability of not exceeding an MIDR of 2% is about 45%. This probability is estimated as 52% when mixed records are used, and only 33% when the pulse-type records are used. Several repetitions of the selection experiments support this trend, which may be attributed to the fact that the pulse-type record pool has a larger proportion of higher intensity ground motions. The median, 5th percentile and 95th percentile IDR envelopes along the building height in subplot c and f show remarkable agreement with the target distribution using only 15 records, regardless of whether the underlying pool contains mixed records or pulse-type records only. Figure 4 presents summary statistics for 30 statistical replicates of the ground motion selection process performed for the 3-story and 12-story buildings at a target distance of 1 km to match a target consisting of pulse-type records only. On the x axis, "mixed" refers to the use of a mixed group of ground motions (pulse and nonpulse), and "pulse" refers to the use of a group consisting of pulse-type records only. The critical value of the statistical KS test is shown in each figure illustrating the acceptable values (i.e., KS statistic is below the critical value), and the ratio of acceptable ensembles (out of 30) is listed in table 2 for each building. Based on the results summarized

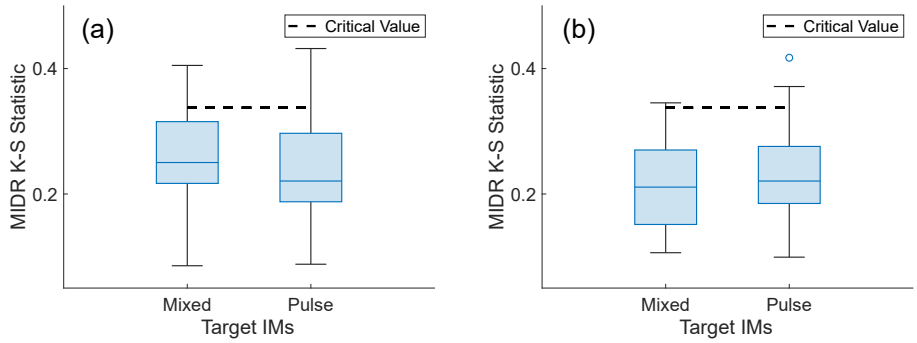

Figure 4: Values of the K-S test statistic of the distribution of MIDR imposed by earthquake records selected to match a target distribution consisting of only pulse-type records for (a) a 3-story building at 1 km; (b) a 12-story building at 1 km. Each box plot represents the statistics of 30 replicates. The number of selected records in each case is 15.

in the figure, the use of mixed pools of candidate records does not appear to introduce additional bias into the predicted structural demands compared to a specialized pool of pulse-type records, even in the extreme case where the entirety of the target is represented by records with strong velocity pulses. Based on the ratios in table 2, the mixed record group appears to perform somewhat better than the pulse-type record group for the 12-story building, but marginally worse for the 3-story building. These structure-specific differences in ground motion record selection will be further explored in future studies.

Table 2: Ratio of ground motion selection replicates producing acceptable MIDR distributions

Building	Mixed records	Pulse-type records
12-story	0.97	0.90
3-story	0.83	0.87

6. Conclusions

The framework presented in this study leverages physics-based earthquake simulations to critically examine procedures that have been adopted to compensate for the sparsity of recorded near-fault field observations, specifically the classification of ground motion records as pulse-type and nonpulse. Based on the results of the study, we do not find justification for the reliance on ad-hoc binary classification of near-fault records into pulse-type and nonpulse records, and the accompanying processes to evaluate the probability of encountering pulse-type motions in the near-fault region. Instead, we demonstrate via a benchmark scenario-based selection approach that restrictions on the underlying candidate pool of near-fault records based on velocity-pulse features do not appear to be important, and eliminating those restrictions does not contribute to statistically significant bias. It is important to note that these results were obtained from the analysis of site-specific simulated earthquake records only. The findings are expected to be applicable to the selection of historical records as well, as long as the matched target response spectra are appropriately representative. However, such an investigation is left for future studies, which should also consider the effects of differing site conditions between the target and selected records. In addition, our database of ground motions consists of a single magnitude 7.0 strike-slip event; therefore, we did not assess the importance of restricting a range of magnitudes or rupture mechanisms for maintaining acceptable bias in the predicted structural demands. We also plan to explore the use of additional target intensity measures in the selection process that may reduce the bias in the predicted near-fault structural demands, such as measures of the record's cumulative energy or significant duration.

7. ACKNOWLEDGMENTS

This research was partially supported by the Exascale Computing Project (ECP), Project Number: 17-SC-20-SC, a collaborative effort of two U.S. Department of Energy organizations - the Office of Science and the National Nuclear Security Administration. The fault rupture simulations were conducted by Dr. Arben Pitarka under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Release: LLNL-PROC-802920. This work was also partially supported by the U.S. Geological Survey Earthquake Hazards Program (Award No. G22AP00380).

8. REFERENCES

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., et al. (2014). "Nga-west2 database." *Earthquake Spectra*, 30(3), 989–1005.

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.

Graves, R. and Pitarka, A. (2018). "Validating ground-motion simulations on rough faults in complex 3-d media." Proceedings of the 11th National Conference in Earthquake Engineering, EERI, Los Angeles, California.

Kenawy, M. and McCallen, D. (2021). "Cceer-20-07: Regional-scale seismic risk to reinforced concrete buildings based on physics-based earthquake ground motion simulations." *Report no.*, Center for Civil Engineering Earthquake Research, University of Nevada, Reno.

Kenawy, M., McCallen, D., and Pitarka, A. (2021). "Variability of near-fault seismic risk to reinforced concrete buildings based on high-resolution physics-based ground motion simulations." *Earthquake Engineering & Structural Dynamics*, 50(6), 1713–1733.

McCallen, D., Petersson, A., Rodgers, A., Pitarka, A., Miah, M., Petrone, F., Sjogreen, B., Abrahamson, N., and Tang, H. (2021). "Eqsim—a multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers part i: Computational models and workflow." *Earthquake Spectra*, 37(2), 707–735.

McKenna, F., Fenves, G. L., Scott, M. H., et al. (2000). "Open system for earthquake engineering simulation." *University of California, Berkeley, CA*.

NIST (2011). *Selecting and Scaling Earthquake Ground Motions for Performing Response-History Analyses (NIST GCR 11-917-15)*. National Institute of Standards and Technology.

Pitarka, A., Graves, R., Irikura, K., Miyakoshi, K., and Rodgers, A. (2019). "Kinematic rupture modeling of ground motion from the m7 kumamoto, japan earthquake." *Pure and Applied Geophysics*, 1–23.

Shahi, S. K. and Baker, J. W. (2011). "An empirically calibrated framework for including the effects of near-fault directivity in probabilistic seismic hazard analysis." *Bulletin of the Seismological Society of America*, 101(2), 742–755.

Shahi, S. K. and Baker, J. W. (2014). "An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions." *Bulletin of the Seismological Society of America*, 104(5), 2456–2466.

Sjögreen, B. and Petersson, N. A. (2012). "A fourth order accurate finite difference scheme for the elastic wave equation in second order formulation." *Journal of Scientific Computing*, 52(1), 17–48.

Tarbali, K. and Bradley, B. A. (2015). "Ground motion selection for scenario ruptures using the generalised conditional intensity measure (gcim) method." *Earthquake Engineering & Structural Dynamics*, 44(10), 1601–1621.

Tarbali, K., Bradley, B. A., and Baker, J. W. (2019). "Ground motion selection in the near-fault region considering directivity-induced pulse effects." *Earthquake Spectra*, 35(2), 759–786.

Zengin, E. and Abrahamson, N. A. (2021). "A procedure for matching the near-fault ground motions based on spectral accelerations and instantaneous power." *Earthquake Spectra*, 87552930211014540.