USGS FINAL TECHNICAL REPORT

AWARD NO. G22AP00380

AWARD TERM: AUGUST 1, 2022 - JULY 31, 2023

Ground Motion Selection for Analysis of Near-Fault Civil Structures using Broadband Physics-Based Earthquake Simulations

Maha Kenawy, Ph.D.

Exponent, Inc.
149 Commonwealth Drive
Menlo Park, CA 94025

Email address: mkenawy@exponent.com
Phone number: 650-862-8484

October 26, 2023

Abstract

The sparsity of near-fault earthquake ground motion records in observational databases poses several challenges to performance-based seismic design of near-fault structures. Simultaneously, earthquake simulations present new opportunities to supplement observational records and advance our understanding of the seismic risks to civil structures in regions proximate to active faults capable of large earthquake events. In this work, broadband physics-based earthquake simulations are used to study the characteristics of near-fault ground motions and their impacts on building structures. The study focuses on two aspects of near-fault ground motion: (1) strong double-sided velocity pulses which are induced by forward rupture directivity effects, and (2) single-sided velocity pulses accompanying the permanent ground offset near the fault, which are known as fling step. The study examines the characteristics and impacts of both of these phenomena on engineering demand parameters, and assesses the need for explicitly incorporating them in the selection of representative ground motion records for nonlinear analysis of near-fault structures. Finally, the performance of simulated earthquake ground motions is compared against real records in a scenario-based seismic analysis context. Based on the findings of the study, guidance on using physics-based earthquake simulations in performance-based seismic engineering applications is provided, and opportunities for future improvements of earthquake simulations and near-fault ground motion selection procedures are highlighted.

Contents

List of Figures

List of Tables

1 Introduction

Performance-based seismic design (PBSD) requires conducting response-history analysis of structures using suites of earthquake records that are representative of the ground shaking intensity at the site of interest. Key prerequisites of the dynamic analysis of structures in PBSD are: (1) defining target intensity levels that reflect the expected ground shaking during the structure's life, and (2) selecting earthquake records that are representative of the target intensity, typically from an available database of historical records (e.g., the NGA-West2 ground motion database (Ancheta et al., 2014)). Implicit in these steps is a reliance on ground motion prediction equations (GMPEs)(Gregor et al., 2014) to estimate representative shaking intensity measures (IMs) from general rupture and site characteristics, and an assumption that those IMs contain the relevant aspects of ground shaking that strongly influence the dynamic response of structures. The analysis of structures located very close to active faults (within 15 km) has highlighted several limitations of this approach. First, ground shaking near the fault was found to be highly sensitive to the rupture characteristics and site conditions (Pitarka et al., 1996; Somerville et al., 1997; Shabestari and Yamazaki, 2003; Karabulut and Bouchon, 2007), including the rupture geometry and slip distribution (source effects), the wave propagation patterns (path effects), and the presence of soft soils that amplify the seismic waves (site effects). Therefore, the prediction of near-fault ground shaking intensity using GMPEs with few rupture-specific and site-specific characteristics may be associated with large uncertainty. Second, because of the unique characteristics of near-fault ground shaking and the high-degree of nonlinear behavior anticipated in near-fault structures, many ground motion IMs that are conventionally used to define target design levels (such as the spectral acceleration at the first-mode period of the structure $SA(T_1)$) have proven deficient in representing the important characteristics of near-fault ground motions (Baker and Cornell, 2004; Luco and Cornell, 2007). Third, because of the sparsity of available earthquake records for large earthquakes at short distances, scaling of the available records to represent the desired intensity level at the site is likely inevitable. The scaling of recorded earthquake time histories, however, has been a subject of investigation because of its potential to alter their important features, especially if distant ground motion records are scaled to represent near-fault records with strong velocity pulses (Grigoriu, 2011; NEHRP Consultants Joint Venture, 2011; Huang et al., 2011).

1.1 Forward rupture directivity effects

Studies have shown that near-fault ground motions may include strong velocity pulses that are highly damaging to civil structures, due to the accumulation of seismic waves as the rupture propagates toward a site, or due to the permanent tectonic displacement of the ground (Housner and Trifunac, 1967; Bertero et al., 1978; Hall et al., 1995). The former is known as forward rupture directivity and is characterized by double-sided velocity pulses in the ground velocity trace (Somerville et al., 1997; Abrahamson, 2000; Somerville, 2003). Examination of the ground motions containing strong pulses (referred to as pulse motions) gained significant research attention over the past two decades. Several researchers have studied the unique characteristics of these recordings, and examined correlations between the ground motion spectral parameters, and the

corresponding strength and ductility demands on structures with varying dynamic properties (Alavi and Krawinkler, 2004; Mavroeidis et al., 2004; MacRae et al., 2001; Iwan et al., 2000; Malhotra, 1999; Gazetas et al., 2008; Tothong and Cornell, 2008; Champion and Liel, 2012; Kohrangi et al., 2019; Li et al., 2020).

Several methods have been proposed to generally improve ground motion selection procedures (Bradley, 2012, 2010) and to explicitly incorporate the effects of near-fault pulse ground motions in the selection process (Shahi and Baker, 2011; Hayden et al., 2014; Shahi and Baker, 2014a; NEHRP Consultants Joint Venture, 2011; Tarbali et al., 2019). Many of the latter group intrinsically rely upon binary classification of near-fault ground motions as either pulse or non-pulse, based on the presence of early-arriving pulses that dominate the ground velocity record. A representative portion of pulse ground motions may be incorporated in the suite of selected earthquake records based on an empirical probability of encountering a strong pulse at a given near-fault location. One example of this classification methodology is the automated algorithm developed by Baker (2007), and later updated in Shahi and Baker (2014a), which utilizes signal processing techniques to detect early-arriving strong pulses in the ground velocity time history. In 2011, the National Institute of Standards and Technology produced a guidance document proposing a procedure for determining the appropriate portion of pulse ground motions to be selected for near-fault engineering analysis (NEHRP Consultants Joint Venture, 2011). Hayden et al. (2014) proposed another automated binary classification algorithm which relies on measures of the energy and the number of significant cycles in the velocity record. The current seismic design standard in the United States, ASCE 7-22 (American Society of Civil Engineers, 2022) recommends computing the likelihood of encountering pulse ground motions near the fault, and requires selecting a representative set of records for the nonlinear time-history analysis of near-fault structures that includes records with directivity pulses. The standard does not explicitly lay out procedures for selecting the ground motions such that they may be reasonably expected to represent the seismic risk to a given structure.

Binary classification of near-fault ground motions into pulse and non-pulse types imposes several challenges: (1) the pulse classification algorithms, and the associated probabilities of encountering strong pulses near the fault, are often developed based on a relatively small number of available near-fault records, and may be associated with large uncertainty; (2) a binary classification of records typically relies on detecting one or two early-arriving pulses that dominate the ground velocity time history. Accordingly, ground motions with multiple strong pulses, such as those resulting from amplification due to sedimentary basins may not be classified as pulse ground motions, although they may impose higher demands on near-fault structures on soft soils compared with ground motions with a single directivity pulse; and (3) studies have shown that the relationship between the frequency content of pulse motions (represented by the period of the dominant pulse T_p) and the demands on structures exhibits a wide dispersion and may vary at different levels of intensity (Zengin and Abrahamson, 2021; Kohrangi et al., 2019; Champion and Liel, 2012). These challenges are amplified by the substantial variability in the shaking intensity in the near-fault region. As a result, the salient characteristics of the damaging near-fault motions may not be appropriately captured by their binary classification into pulse and non-pulse types, and such a methodology may fail to identify records that are most damaging to a given structure. Recent

studies have supported the elimination of such a binary approach, including a study by Zengin and Abrahamson (2021) who analyzed the response of structures to a relatively small number of near-fault records, and proposed a new IM, the maximum instantaneous power $max\ IP(T_1)$, for the selection of near-fault records without distinction between pulse and non-pulse types. In addition, Tarbali et al. (2019) proposed a near-fault record selection methodology that incorporates the probability of encountering strong pulses into the underlying probabilistic seismic hazard analysis (PSHA), but does not require the classification of records into pulse and non-pulse; instead, it relies upon a large group of IMs including the spectral acceleration (SA) at multiple periods, cumulative velocity and significant duration measures. The study showed that the proposed procedure may eliminate the need for ad-hoc classification based on the pulse characteristics of records, but it remains dependent on estimating the probability of encountering ground motions with directivity pulses in the near fault region (and the distribution of T_p at the site), because such motions are not appropriately considered in the conventional PSHA.

1.2 Fling step effects

In addition to the challenges associated with incorporating the effects of directivity pulses in seismic analysis of structures, the characteristics and effects of the permanent tectonic displacement (known as fling step) are less well-understood. Fling is characterized by a long-period single-sided velocity pulse and a corresponding non-zero displacement at the end of ground shaking. Fling effects were largely overlooked until the 1999 Chi-Chi, Taiwan and 1999 Kocaeli, Turkey earthquakes, whose ground motion records included fling-type pulses and large peak velocities, which were distinguished from the known forward directivity pulses (Bolt, 2002). Until recently, fling pulses and the corresponding permanent displacement offsets were documented for only a handful of earthquakes (Burks and Baker, 2016).

The characteristics and effects of fling on civil structures are difficult to characterize because most available ground motion records are processed such that the static offset is removed to eliminate low frequency noise and errors due to movement of the recording equipment. Special processing techniques can be used to preserve the static offset in ground motion records, such as baseline correction (Boore and Bommer, 2005). However, the resulting fling magnitude is highly dependent on the choice of the baseline, and may contain large uncertainty if the magnitude of the permanent offset is not known by other means, for example using GPS measurements (Boore and Bommer, 2005; Akkar and Boore, 2009). Consequently, a small number of historic earthquake records, in combination with simulated ground motions, were typically used to study the characteristics of fling and develop predictive models for seismic hazard analysis, such as those by Kamai et al. (2014) and Burks and Baker (2016). More recently, an extended baseline correction technique (eBASCO) was developed for processing ground motion records to extract the permanent offsets (D'Amico et al., 2018) and used to construct a large database of near-fault ground motion records with permanent offset values corresponding to a wide range of moment magnitudes from many regions around the world (Sgobba et al., 2021). This database was also the basis for developing a new predictive fling model (Schiappapietra et al., 2022) which describes the distance scaling of the magnitude of the permanent displacement.

A number of researchers have studied the effect of fling pulses on single-degree-of-freedom (SDOF) or multi-degree-of-freedom (MDOF) systems using a small number of earthquake records which were specially processed to preserve permanent offsets (Garini et al., 2011; Ventura et al., 2011; Kalkan and Kunnath, 2006; Calugaru and Panagiotou, 2012; Liossatou and Fardis, 2016; Daei et al., 2023; Burks and Baker, 2014; Hamidi et al., 2020). The majority of studies focused on understanding how the processing of the ground motion records to remove the permanent offset impacts the seismic demands on SDOF or MDOF systems, and they obtained somewhat mixed results. However, the findings of previous research generally aligned with the accepted practice of ignoring the static offset associated with fling, because it has a negligible effect on elastic response spectra at periods of engineering interest (Boore, 2001; Akkar and Boore, 2009). Because of the small number of fling records used in those studies and their different spectral characteristics, the findings may have been dominated by the record-to-record variability, instead of highlighting significant trends indicating the relative importance of considering fling effects in performance-based seismic design of structures. Incorporating the effects of fling into seismic hazard analysis and improving fault-displacement hazard analysis are areas of ongoing research (Dalguer et al., 2021; Hisada and Tanaka, 2021). For example, Bayless and Abrahamson (2023) recently proposed a method to develop earthquake time histories for engineering applications which contain a fling step with a specified duration and amplitude, and match a target response spectrum. However, to the author's knowledge, no systematic studies were performed to assess: (1) whether the intensity characteristics of "naturally" occurring fling-containing ground motion records differ from non-fling records corresponding to the same causal parameters (magnitude, distance, site conditions), and (2) whether those characteristics should be taken into account when creating target seismic hazard spectra, and selecting records for the analysis of structures at near-fault locations which may experience substantial fling pulses.

1.3 Objectives

Physics-based fault rupture simulations have the potential to enrich the available limited dataset of near-fault ground motion records, and enable the use of site-specific records that closely match the shaking intensity at the design site in an engineering analysis. Three-dimensional, physics-based ground motion simulation is an attractive methodology to study the properties and impacts of near-fault ground motion, because it can take into account the fault geometry, site conditions and the physics of wave propagation. Physics-based earthquake simulations also present a promising tool to investigate the characteristics and effects of fling ground motion records because they inherently contain the permanent ground offset without the need for processing. Such simulations have evolved rapidly in recent years due to the advances in understanding earthquake rupture processes and high-performance computing technologies (Dreger et al., 2015; Rodgers et al., 2019; Mccallen et al., 2020; Paolucci et al., 2021), and have been used in several studies to analyze the response of structures to earthquake events (Burks et al., 2015; Marafi et al., 2019; Bijelić et al., 2019a,b; Kenawy et al., 2021), and the effects of fling pulses on structures (Burks and Baker, 2016; Reddy et al., 2021).

This study uses three-dimensional (3D) high-resolution physics-based fault rupture simulations

representative of shallow crustal earthquakes to examine the seismic demands on near-fault structures, and identify the characteristics of near-fault ground motions that should be taken into account for earthquake record selection in engineering analysis applications. The study specifically focuses on interrogating the influence of directivity pulses and fling pulses on the structural demands on building structures with different dynamic characteristics, and the need for considering both phenomena in the context of scenario-based earthquake record selection. Finally, the performance of the simulated ground motions, when used to replace real records for engineering analysis or to establish the target hazard spectra, is compared against the performance of real records and empirical ground motion models. Based on the findings of the study, recommendations are made regarding the selection of earthquake records for the nonlinear analysis of near-fault structures, and for using broadband physics-based simulated ground motions in such analysis contexts.

2 Earthquake Simulations and Structural Analysis Models

2.1 Fault rupture simulations

Nine fully deterministic broadband (0-5 Hz) physics-based three-dimensional fault rupture realizations representing a M7.0 strike-slip earthquake scenario are used in this study, and are listed in Table 1. The earthquake simulations were performed by Arben Pitarka at Lawrence Livermore National Laboratory (LLNL). Six of these simulations (Realizations A through F) were generated under 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., 2021a,b) using the CORI machine at the National Energy Research Scientific Computing Center, and were described in detail in Kenawy et al. (2023). Realizations G through I were generated using the Quartz computerat LLNL and were funded by the LLNL Computing Grand Challenge 16 "Broadband Earthquake Ground Motion Simulations on LLNL's Next Generation HPC Systems". The ground motion acceleration time series were generated using the finite difference code SW4 (Seismic Waves, fourth order) developed at LLNL (Sjögreen and Petersson, 2012), which simulates seismic waves using 3D earth models on distributed-memory parallel computers.

Table 1: Characteristics of the fault rupture realizations and site conditions

Rupture Realization	Hypocenter Location	Slip Patch Location	Rupture Depth (m)	Rock V_{s30} (m/s)	Basin V_{s30} (m/s)	Basin Depth (m)
A	Left	Right - surface	200	500	380	600
B	Left	None	200	500	380	600
C	Center	None	200	500	380	600
D	Center	Right - surface	200	500	380	600
E	Left	Right - deep	200	500	380	600
F	Center	Right - deep	200	500	380	600
G	Left	Right - surface	200	500	380	1200
H	Left	Right - surface	20	500	380	600
I	Left	Right - surface	20	1200	380	600

The kinematic earthquake rupture modeling technique of Graves and Pitarka (GP) was used to simulate the fault rupture processes (Graves and Pitarka, 2016; Pitarka et al., 2019); this technique has been validated in simulations of recorded earthquakes (Pitarka et al., 2022; Graves and Pitarka, 2018; Pitarka et al., 2019). The GP rupture generator uses spatial and temporal kinematic rupture parameters that are calibrated using observed rupture kinematics and physical constraints from dynamic rupture modeling, and the resulting rupture models allow for reproducing observed characteristics of near-fault ground motion on a broad frequency range. Further details about the rupture model and generated ground motion can be found in Graves and Pitarka (2016), Pitarka et al. (2019) and Rodgers et al. (2019).

The strike-slip fault ruptures were simulated with a fault length of 62.6 km and fault width of 16 km. The rupture velocity is set to 80% of the local shear wave velocity, which is consistent with observed rupture velocity values for shallow crustal earthquakes on mature faults. The fault dip angle is approximately 90 degrees, and the average rake angle was set to 0 degree with small spatially correlated random perturbations, computed following the GP method. The rupture realizations represent common shallow crustal earthquakes with a predominantly strike-slip mechanism. The selected realizations have different combinations of hypocenter locations (toward the left or center of the fault), slip distribution characteristics, depth to the top of rupture, and soil conditions, as listed in Table 1. Seven realizations contain rupture asperities or patches of large slip placed in predetermined locations along the fault (this technique is referred to as a hybrid rupture because it contains stochastic and deterministic slip distribution components), whereas the remaining two follow a purely stochastic distribution of the fault slip. Ruptures with localized areas of shallow large slip can result in large ground motion amplitudes at sites near the slip patches (Pitarka et al., 2019); therefore, varying the presence, location and depth of large slip patches in the fault rupture simulations allows for studying the variability in the characteristics of near-fault ground motions and corresponding structural demands. In addition, examining the influence of slip patches on the ground motion characteristics serves to provide guidance for improving future fault rupture simulations. The depth below the ground surface to the top of the fault was set to either 20 m or 200 m, as designated in the table. The very shallow-depth ruptures allow for studying the effects of relatively large permanent offsets in the near-fault region.

The computational domain used in this study spans 100 km by 40 km by 30 km, and incorporates three-dimensional geological features and a flat earth surface. The domain has a minimum grid spacing of 8 m to achieve the numerical accuracy required to model the wave propagation up to a target frequency of 5 Hz. The area on one side of the fault consists of a shallow sedimentary basin with an approximately 90-degree basin edge dip angle, and the rest of the domain consists of rock. The velocity structure in the domain is described in Kenawy et al. (2023), and corresponds to a 600-m deep basin layer with average shear-wave speed in the top 30 meters V_{s30} = 380 m/s, and V_{s30} = 500 m/s in the rock region. This velocity model represents realizations A through F, and realization H. Variations to the velocity structure were introduced in realization G and I as described in Table 1. 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 strike-normal (SN) component, a horizontal strike-parallel (SP) component, and a vertical component. It is noted that the ground motion simulations do not account for nonlinear site response. Studies in recent years have discussed the significance of and potential improvements of incorporating nonlinear site effects in physics-based ground motion simulations (Jeong and Bradley, 2017; de la Torre et al., 2020; Bradley, 2019). However, the inclusion of realistic nonlinear models in 3D wave propagation techniques is still the subject of research and is outside the scope of this study.

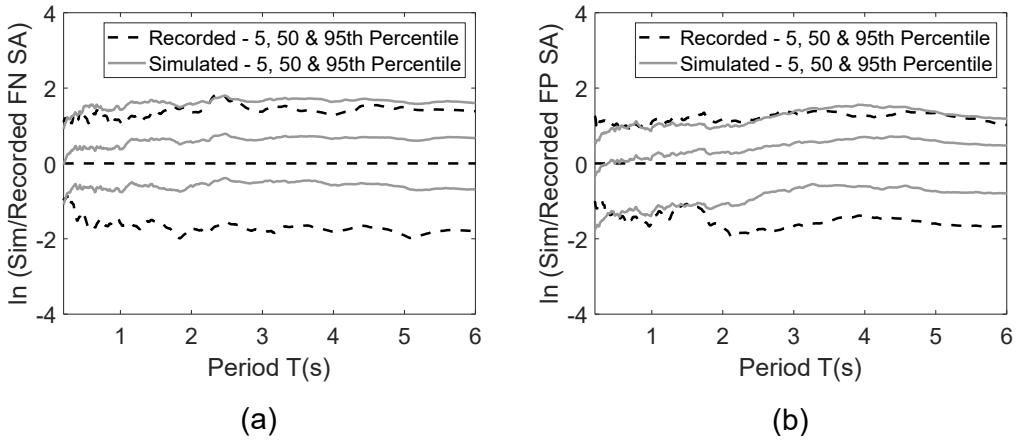

Figure 1: Median, 5th percentile and 95th percentile pseudo spectral acceleration of recorded and simulated ground motions within a 15 km JB distance: (a) strike-normal component (equivalent to fault-normal (FN)); (b) strike-parallel component (equivalent to fault-parallel (FP)).

2.2 Comparison between near-fault simulated and recorded ground motions

The produced database of simulated records used in this study consists of 69,498 horizontal records (where each record includes a SN and SP component) from the nine rupture realizations. The simulated ground motions do not represent a specific historical earthquake; rather, they are intended to produce records that are generally representative of shallow crustal earthquake ground motion records. The characteristics of near-fault subset of this database were extensively studied and compared against those of near-fault recorded ground motion observations in Kenawy et al. (2023), revealing that the underlying idealized rupture model and velocity structure are capable of representing the salient characteristics of near-fault field observations.

The dataset of recorded near-fault ground motions is obtained from the PEER NGA-West2 database (Ancheta et al., 2014). It consists of 94 records near-fault records, which were obtained by filtering the database for the following parameters: earthquake magnitudes between 6.7 and 7.4; Joyner-Boore (JB) distance between 0 and 15 km, and V_{s30} between 200 and 700 m/s. The range of magnitudes was selected such that the resulting records have a mean magnitude of 6.95, as close to the M7.0 simulated scenario as reasonably possible. The NGA-West2 records used in this study are provided in Appendix 1 in Kenawy et al. (2023). The as-recorded horizontal component accelerations were rotated to their SN and SP orientations based on the fault orientations extracted from the database flat files.

Of relevance to this study is the consistency of the spectral characteristics of the simulated ground motions with the recorded ground motions, which is illustrated in Figure 1 for both the SN and SP components. The figure shows the median, 5th and 95th percentile spectra associated with a near-fault subset of the simulated ground motions along with the NGA-West2 dataset of recordings; this subset was sampled from the database of simulated records such that the causal

parameters of both datasets are consistent, as described in Kenawy et al. (2023). The NGA-West2 dataset includes records from shallow crustal earthquakes around the world with varying rupture mechanisms. Therefore, those records have a wider variability than the simulated ground motions, which is evidenced by the spread of the 5th and 95th percentile spectra of Figure 1 around the median spectrum. The median spectrum of the SN component of the simulated motions is consistently higher than that of the recorded motions, although the 95th percentile spectra of the two datasets are in reasonable agreement. The notable difference between the 5th percentile spectra of both datasets suggests that the ground motion simulations estimate somewhat higher shaking intensities compared with the near-fault field records, especially in the SN component. In contrast, the 5th, 50th, and 95th percentile SP acceleration spectra of the simulated ground motions appear to be consistent with those of the recorded ground motions across a broad range of periods. Above a period of 2 s, the SP median and 5th percentile spectra of the simulations are notably higher than those of the field recordings. However, for very short periods (between 0.2 and 0.3 s in this case), the SAs of the SP simulated motions tend to be smaller than those of the recorded motions. Overall, the trends suggest reasonable consistency between the spectral characteristics of the simulated and recorded ground motions. In addition to the spectral shapes, the distance scaling of spectral accelerations and ground velocity, significant durations, velocity pulse features, and dominant pulse periods of the simulated ground motions were shown to be reasonably consistent with those of the recorded ground motions (Kenawy et al., 2023).

2.3 Structural analysis models

Three RC special moment frame buildings were designed for seismic category E: a 3-story building, a 12-story building and a 20-story building, following the seismic provisions of ASCE 7-16 (American Society of Civil Engineers, 2017) and the RC provisions in ACI 318-14 (American Concrete Institute, 2014). 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. The first-mode period of the 3-story, 12-story and 20-story buildings are 0.75 s, 2.23 s and 3.45 s, respectively. Two-dimensional structural simulation models of representative building frames were created using the structural analysis platform Opensees (McKenna et al., 2000). The structural components are simulated using the lumped plasticity modeling approach (Giberson, 1967), in which the inelastic behavior of the structural members is assumed to be concentrated at the member ends only. Capturing the deterioration effects of the structural members is important for realistic analysis of near-fault structures (Champion and Liel, 2012), particularly because yielding leads to elongation of the building period that may change its sensitivity to near-fault velocity pulses. In the analysis models created in this study, springs at both ends of each member cumulatively simulate the nonlinear effects due to strength and stiffness degradation of concrete and steel, steel rebar buckling and fracture, and bond slip between steel rebar and the surrounding concrete. The modified Ibarra-Medina-Krawinkler (Ibarra et al., 2005) material model with peak-oriented hysteretic response is used to represent the constitutive behavior of the nonlinear springs, and the model parameters were computed using the prediction equations in Haselton et al. (2016). Further details about the structural design and simulation models are available in Kenawy et al. (2021) and

Kenawy and McCallen (2021).

Each two-dimensional building model is replicated at all 3,861 stations of each rupture realization, and is subjected to the SN and SP horizontal components separately. In this study, the following response quantities are considered: the maximum interstory drift (MIDR) ratio, the residual interstory drift ratio (RIDR), which is defined as the largest interstory drift at the end of the ground shaking, and the maximum floor acceleration (MFA), which is defined as the largest floor acceleration across the time history of the shaking. In addition, the distribution of the peak interstory drift along the height of the building (IDR envelope) is examined in some cases.

3 Consideration of Directivity Pulse Features in the Analysis of Near-Fault Structures

3.1 Sensitivity of the structural demands to the classification of records as pulse/non-pulse

To examine the need for classifying near-fault ground motion records based on pulse features for structural engineering analysis, the sensitivity of the relationship between the near-fault ground motion intensity and the MIDR to the classification of records based on the presence of directivity pulses is analyzed. The following ground motion IMs are considered:

- The spectral acceleration at the first-mode period of the structure
- The peak ground velocity of the record
- The parameter SA_{avg} which is defined as the geometric mean of the spectral acceleration over the period range 0.5 T_1 - 2 T_1. SA_{avg} is used as a proxy for the spectral shape of the ground motion record, and was found in a number of studies to be important for measuring the structural damage potential of a record (Eads et al., 2015; Kohrangi et al., 2019). The period range for SA_{avg} was selected to account for both the presence of higher mode contributions, and the elongation of the structure's first-mode period as it deforms into the inelastic regime.
- The maximum instantaneous power at the first-mode period $max\ IP(T_1)$ as defined in Zengin and Abrahamson (2021). The parameter $max\ IP(T_1)$ was computed after bandpass-filtering the ground velocity records with the limits (0.2 - 3 T_1) using a four-pole butterworth filter.
- The significant duration of the record D_{s595}, which is defined as the duration between 5% and 95% of Arias intensity.

The sensitivity of the IM-MIDR relationships to the classification of the underlying records as pulse/non-pulse was first examined using linear regression models. Regression models of the form $ln(MIDR) = a + b * ln(IM) + c * Pulse\ Class$ were used, where the $Pulse\ Class$ is an indicator categorical variable that takes on the value 1 or 0, and a, b and c are the regression coefficients. Interaction terms between the IM and $Pulse\ Class$ were considered. The significance of the pulse classification as a predictor variable was judged based on the associated p-value, representing the probability that $Pulse\ Class$ is not a statistically significant predictor of the MIDR, given the IM in question. The pulse classification was found to be a statistically significant predictor of the MIDR regardless of the employed IM (across the IM types listed above). To avoid dependencies on the functional form and to examine more complex combinations of IMs, the analysis was repeated in a non-parametric form using a tree-based ensemble learning regression model. Specifically, Random Forest (RF) regression (Breiman, 2001, 2004) was used because of its flexibility, ability to handle nonlinear relationships, and insensitivity to noise variables (Biau, 2012). RF models combine

predictions from different decision trees to make the final model predictions. Studies in recent years have shown that RF estimators which average predictions over trees built on sub-samples of the training dataset are asymptotically normal under certain regularity conditions (Mentch and Hooker, 2016), opening the door to applying traditional statistical inference methods to the models, and expanding their utility in applied research contexts. The recently developed methodology by Coleman et al. (2022) to conduct hypothesis testing for the RF model, in an analogous fashion to conventional F-tests, was applied. The testing methodology entails creating two different RF models: one with all the predictor variables (original model), and one with a permuted version of the predictor variable being tested (reduced model), and the difference between the mean squared error (MSE) of the predictions of both RF models is evaluated at the test data points. Then, the predictions are permuted between trees belonging to both the original and reduced models, and the differences in the MSE are re-evaluated. The purpose of this permutation is to assess whether the variable in question adds any predictive power to the model beyond adding additional randomness. The permutations are performed B times, where B is chosen to be equal to 500 in our study. Then, the following null hypothesis is tested:

$$H_0 : E(MSE(\tau)) = E(MSE_R(\tau)) \tag{1}$$

where $E(MSE(\tau))$ and $E(MSE_R(\tau))$ are the expected values of the mean squared errors corresponding to the full and reduced models, respectively, and evaluated over the test dataset τ. This hypothesis is tested against the alternative hypothesis that the error produced by the full model is significantly smaller than the reduced model. The p-value of the test is based on comparing the MSE difference between the full and reduced models to the MSE differences corresponding to the various model permutations. This testing methodology attains the Type I error rate and has been shown to perform well in the context of high-dimensional feature spaces, and in the presence of highly correlated variables (Coleman et al., 2022).

This testing procedure is conducted for RF models in which the predicted variable is the MIDR, and the predictive features consist of a single IM or a combination of IMs in addition to the pulse classification of the underlying records. The significance of the pulse classification as a predictor variable is tested in each case. An effort is made to minimize the number of correlated variables in each RF model, to reduce any potential impacts on predicting the importance of the pulse classification variable. However, it is important to recognize that we use combinations of spectral accelerations at different periods, which are known to be correlated. The dataset of near-fault ground motions (within a 15 km JB distance) is used in the analysis, and consists of 26,328 data points per structure, where 21% of the ground motions in the dataset are classified as pulse records. 70% of the dataset was used for training, and the remaining 30% for testing (evaluating the MSE). Because the portion of pulse ground motions is relatively small in the dataset, the RF prediction and testing procedure was repeated ten times, where the training and testing data points were selected randomly from the dataset, and reported the average p-value associated with each model. While different applications of the model and testing procedure yielded slightly different p-values, the statistical significance trends remained stable. The R package RFtest was used to conduct the permutation and hypothesis testing with the following parameters which were selected to balance accuracy and computational effort, while adhering to the recommendations in Coleman et al.

(2022): the number of RF trees in each model was selected to be equal to 200; the size of the subsamples used for the training is equal to $n^{0.5}$ where n is the number of observations in the training dataset. The RF models and average p-values for the pulse classification parameter are listed in Table 2 for the 3-story and 12-story buildings.

Table 2: Average p-values for the significance of the last predictor variable (the pulse classification or the dominant pulse period) in Random Forest regression models describing IM-MIDR relationships

Model #	Predictor Variables	3-story Building	12-story Building
1	$ln(SA(T_1))$, $Pulse\ Class$	0.002	0.002
2	$ln(PGV)$, $Pulse\ Class$	0.002	0.268
3	$ln(SA_{avg})$, $Pulse\ Class$	0.003	0.002
4	$ln(max\ IP(T_1))$, $Pulse\ Class$	0.244	0.012
5	$ln(SA(T_1))$, D_{s595}, $Pulse\ Class$	0.002	0.004
6	$ln(SA(T_1))$, $ln(SA(0.5T_1))$, $ln(SA(1.5T_1))$, $ln(SA(2T_1))$, $Pulse\ Class$	0.120	0.204
7	$ln(SA(T_1))$, T_p	0.002	0.002
8	$ln(PGV)$, T_p	0.002	0.002
9	$ln(SA_{avg})$, T_p	0.002	0.002
10	$ln(max\ IP(T_1))$, T_p	0.002	0.002
11	$ln(SA(T_1))$, D_{s595}, T_p	0.002	0.002
12	$ln(SA(T_1))$, $ln(SA(0.5T_1))$, $ln(SA(1.5T_1))$, $ln(SA(2T_1))$, T_p	0.077	0.003
13	$ln(SA(T_1))$, $ln(SA(0.25T_1))$, $ln(SA(0.5T_1))$, $ln(SA(1.5T_1))$, T_p	0.080	0.108

Based on the statistical analysis, the hypothesis that the MIDR imposed by ground motion records with the same $SA(T_1)$ is sensitive to the pulse classification of the record is not rejected, in agreement with the inference based on the linear regression models. This is reflected in the low p-values ($<<$ 0.05) associated with model #1 which consists of $ln(SA(T_1))$ and $Pulse\ Class$. The same was true in the case of model #3 in which $ln(SA_{avg})$ was the predictor. Since pulse records tend to have shorter significant shaking durations, a model which uses both $ln(SA(T_1))$ and D_{s595} as predictors was also tested (model #5); the predicted MIDR by this model remained sensitive to the pulse classification of the underlying records. On the other hand, when $ln(max\ IP(T_1))$ was solely used as a predictor, the MIDR was not sensitive to the pulse classification of the records (p-value $>$ 0.05) for the 3-story building. This result suggests that using $ln(max\ IP(T_1))$ as an IM may capture the effects of directivity pulse features in a ground motion record, although linear regression indicated that the correlation between $ln(max\ IP(T_1))$ and $ln(MIDR)$ is generally lower than the correlation between $ln(SA(T_1))$ and $ln(MIDR)$. Overall, the results support the use of the near-fault ground motion selection approach based on a combination of $ln(SA(T_1))$ and $ln(max\ IP(T_1))$ proposed by Zengin and Abrahamson (2021), especially for relatively stiff buildings.

When the spectral accelerations at multiple relevant periods are used as predictors, we find that the predicted MIDR is also not sensitive to the pulse classification of the underlying records (p-values $>$ 0.05). In other words, selecting ground motion records based on the important features of the spectral shape should eliminate the dependence of the MIDR on the classification of the records as pulse/non-pulse. This is an important result because common ground motion selection

methodologies already rely on matching a target spectral shape for the site of interest (Baker, 2011; Bradley, 2010). However, there is an important caveat to this result: while the sensitivity of the MIDR to the pulse classification of the records is significantly reduced when the predictors are spectral shape features across multiple periods, ground motion selection remains susceptible to the bias due to misrepresenting the target spectral shape features, especially for near-fault locations which are not typically well-represented in the empirical data used to develop ground motion prediction models. Improving the representation of the near-fault target spectra has been addressed in the context of PSHA (Abrahamson, 2000; Shahi and Baker, 2011; Almufti et al., 2015; Tarbali et al., 2019). In a seismic analysis based on a particular earthquake scenario of interest, the important spectral shape features would be represented by the site-specific ground motions themselves, as long as the scenario representation contains the appropriate variability. For example, in the ground motion selection experiments presented later, the simulated ground motions are used to construct the target spectra for sites at several distances from the fault.

A representative distribution of the values of the MSE difference corresponding to the permuted models in the case of model #6 (which consists of the spectral accelerations at multiple periods and $Pulse\ Class$) is shown in Figure 2b, and is contrasted with a representative distribution corresponding to model #1 (which consists of $SA(T_1)$ and $Pulse\ Class$ only) in Figure 2a. In each case, the values approximate a normal distribution centered around zero, as expected. The orange line corresponds to the MSE difference between the original full and reduced models, and, in the case of Figure 2b, is not significantly far from the permuted distribution, thus indicating no significant difference between the models including and excluding the pulse classification. The opposite is true for model #1 depicted in Figure 2a. In this study, we used the spectral acceleration at four building-dependent periods as predictors: T_1, 0.5 T_1, 1.5 T_1 and 2 T_1, such that the frequency content corresponding to an elongated effective period after yielding and to higher vibration modes would be represented. Finally, we also note that the use of the peak ground velocity (PGV) as a predictor variable renders the MIDR of the 12-story building insensitive to the pulse classification, but not the 3-story building. This finding qualitatively agrees with empirical evidence which suggests that the PGV is an important IM with respect to the demands on flexible buildings; however, the PGV may not be an appropriate IM for relatively stiff structures such as the 3-story building simulated in this study.

3.1.1 Sensitivity of the structural demands to the dominant directivity pulse period

The seismic demands on structures were found in several studies to be sensitive to the ratio between the period of the dominant pulse of pulse-type ground motions and the first-mode period of the structure T_p/T_1 (Champion and Liel, 2012; Baker and Cornell, 2004; Kohrangi et al., 2019; Zengin and Abrahamson, 2021), such that ground motions with dominant pulses which coincide with the effective vibration period of a given structure may be the most damaging. We studied the sensitivity of the MIDR demands to the period of the dominant pulse in the velocity records of the simulated ground motions. Because we are interested in the demands imposed by single components (either SN or SP), we used the methodology proposed by Baker (2007) to extract the period of the dominant pulse T_p in all the ground motion single-component signals (not only those

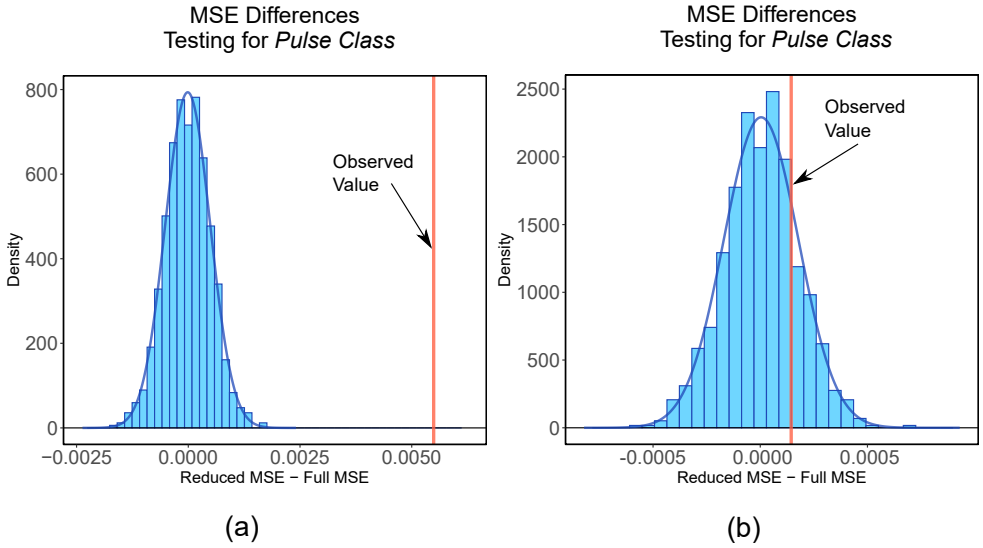

Figure 2: Distribution of the difference in the mean squared error between the full and reduced RF model permutations along with the observed MSE difference between the original full and reduced RF models (with and without $Pulse\ Class$) predicting the 3-story MIDR using (a) the spectral acceleration at T_1 only; and (b) the spectral accelerations at four periods: 0.5 T_1, T_1, 1.5 T_1 and 2 T_1.

classified as pulse) in the database of simulated ground motions; T_p represents the period associated with the maximum Fourier amplitude of the pulse wavelet extracted from the signal, and is illustrated in Figure 3a for an example wavelet.

We first examined the dependence of the residuals of a linear regression model between $ln(SA(T_1))$ and $ln(MIDR)$ on the dominant pulse period across the entire dataset of ground motion records; those residuals are plotted in Figure 3b and c for the 3-story and 12-story buildings, respectively. The plots highlight the range of pulse periods in the simulated ground motion dataset as a ratio of the first-mode period of each building, and emphasize the complexity of the relationship between the MIDR and T_p, yet in each case the regression model detected a statistically significant dependence of the residuals on T_p at the 5% level (p-values $<<$ 0.05). This result indicates a potential bias in the predicted structural demands, if the important pulse frequency characteristics are misrepresented in a set of selected near-fault records for engineering analysis. To more appropriately represent the nonlinear relationship between T_p and $MIDR$, the RF permutation procedure and hypothesis testing described in the preceding section were repeated, but with T_p as the explanatory variable whose significance is tested, instead of the pulse classification. We focused this analysis on the near-fault ground motion dataset, and tested several models consisting of a single IM or combinations of different IMs, including those listed in Table 2. The results indicated that only combinations of structure-specific spectral acceleration ordinates proved effective in reducing the sensitivity of the MIDR to the dominant pulse period. For the 3-story building, T_p was a statistically insignificant predictor of the MIDR when four SA ordinates between 0.25 T_1 and 2 T_1 (model #12 and model #13 in Table 2) were used as predictors. Nonetheless, spectral ordinates at peri-

ods lower than $0.5T_1$ were generally not statistically significant predictors of the 3-story MIDR because the behavior of this building is dominated by first-mode deformations and severe yielding (especially near the fault) leading to period elongation. A representative outcome corresponding to model #12 is shown in Figure 4b. The plot shows the outcome of the permutation and hypothesis testing procedure applied to all the predictor variables in the model. In addition to the significance level of each variable, the figure includes a measure of the variable importance based on the number of standard deviations away that the original MSE difference is from the center of the permutation distribution. In agreement with the expected behavior for this relatively stiff building, the most important spectral ordinates are those between T_1 and 2 T_1, $SA(0.5\ T_1)$ is less important, and T_p is not significant in this case.

The 12-story building, on the other hand, was found to be more sensitive to spectral ordinates at relatively short periods. In model #13 in Table 2, T_p was not statistically significant when four spectral ordinates between 0.25 T_1 and 1.5 T_1 were used as predictors. This is also in agreement with the expected behavior of this relatively flexible building, which is sensitive to higher mode effects, and does not experience substantial yielding. These observations are echoed in the significance level and standard deviation importance shown in Figure 4a for the predictors in model #13 for the 12-story building. To put these observations in context, we used a conventional pushover analysis to compute simple measures of the ductility of both buildings by dividing the peak building drift at a 20% base-shear loss over the building drift value at yielding, and found that the ductility capacity of the 3-story building is approximately twice that of the 12-story building. Overall, the results suggest that using the full spectral shape as the target IM to conduct scenario-based ground motion record selection for analyzing near-fault structures should eliminate the sensitivity of the MIDR demands to the distribution of the dominant pulse periods corresponding to the site of interest.

It is important to reiterate that the MIDR alone may not fully capture the effect of T_p on the structural demands because different T_p values excite different mode shapes and building deformation patterns. Because the interstory drift ratio (IDR) distribution along the building height is especially important for tall buildings, this issue further complicates our understanding of the impact of T_p. We studied the spatial distribution of the IDR along the height of both buildings across different T_p/T_1 ratios, and observed notable sensitivity of the deformed shape of the 12-story building to T_p/T_1, especially in the case of ground motions with pulse periods close to 0.5 T_1. Such ground motions imposed the largest demands on floors 7 through 12, whereas ground motions with T_p equal to or longer than T_1 induced the largest drifts in the bottom floors. In contrast, the deformed shape along the 3-story building height, which is dominated by a first-mode deformation pattern, was not sensitive to the pulse frequency content. The findings related to the 3-story building may be an artifact of the available frequency range in the underlying database of M7.0 rupture realizations. Further examination of the sensitivity of the deformed shape of relatively stiff buildings to shorter pulses that may be associated with smaller-magnitude earthquakes is warranted.

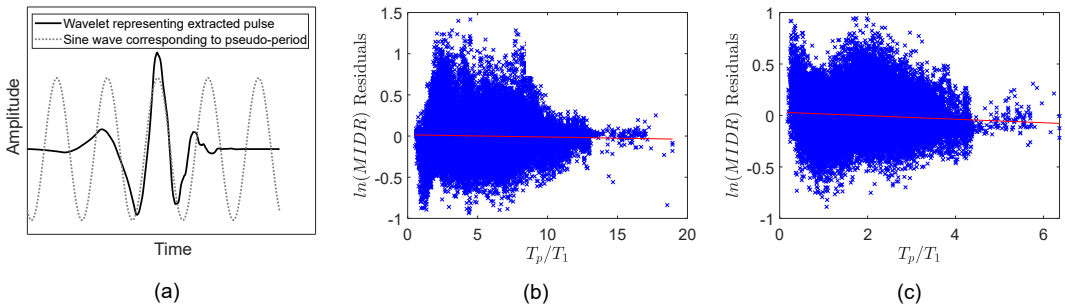

Figure 3: (a) Concept of the dominant pulse period; (b) the 3-story MIDR residuals at different Tp/T_1 values; (c) the 12-story MIDR residuals at different Tp/T_1 values.

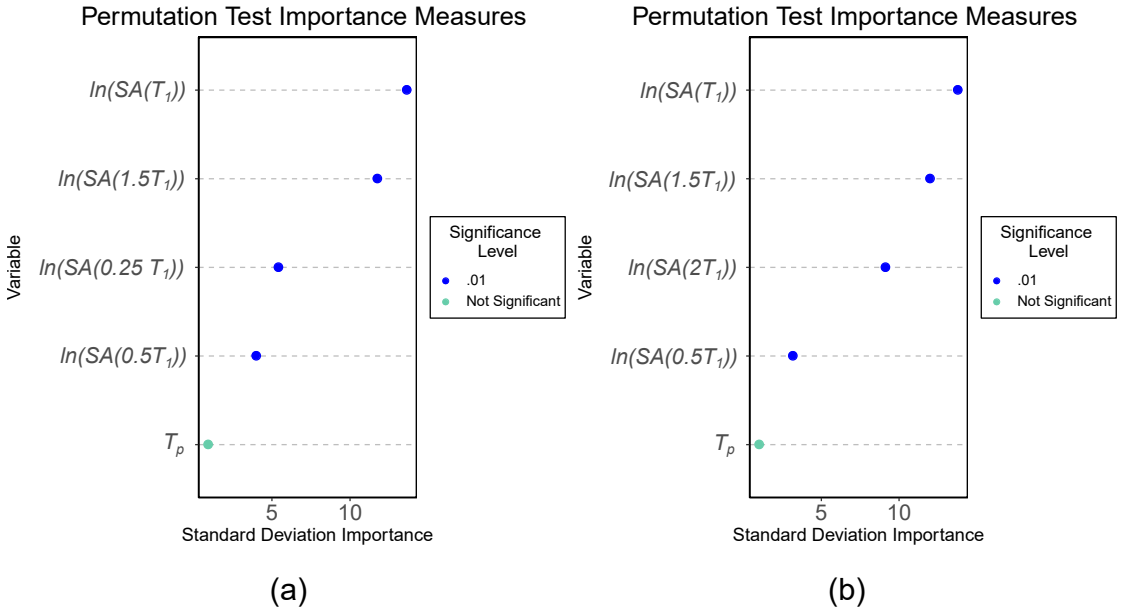

Figure 4: Standard deviation importance and significance level based on permutation tests for the MIDR predictors in two RF models consisting of combinations of spectral acceleration ordinates and the dominant pulse period: (a) 12-story building; (b) 3-story building.

3.2 Representation of directivity effects in ground motion record selection

3.2.1 Selection methodology

We used the Generalized Conditional Intensity Measure (GCIM) approach (Bradley, 2010, 2012) to measure the bias associated with selecting simulated earthquake records for the analysis of near-fault 3-story and 12-story buildings. The GCIM framework allows for flexible and holistic selection of ground motion records to minimize the mismatch between ensembles of records and target distributions of multiple IMs chosen by the analyst. Specifically, we investigated the need for binary classification of near-fault records based on their velocity pulse characteristics in the selection process. We designed ground motion selection experiments in which the structural demands associated with an ensemble of selected records are compared against the demands associated with a reference target distribution based on the fault rupture realizations. We established reference distributions of ground motion IMs consisting of pulse records only, and attempted to represent those distributions using a "mixed" pool of candidate records which did not distinguish between pulse and non-pulse records. The goal of the experiments was to examine the potential degree of bias in the estimated structural demands if: (1) the ground motion at the site of interest is dominated by strong pulses, and (2) the record selection procedure fails to explicitly consider the pulse characteristics.

We use scenario-based ground motion selection (Tarbali and Bradley, 2015), such that our target is not a conditional spectrum, but rather is the IM distributions obtained directly from multiple fault rupture realizations of a single earthquake scenario. We leverage the large database of ground motion simulations 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:

- At a minimum, we use the acceleration response spectra as the target IM; we represent the target distributions of 19 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 spectral ordinate.

- We test the use of target distributions which include the full spectral shape in addition to one of the following IMs: PGV, $max\ IP(T_1)$, D_{s595} and T_p. In each case, we follow the recommendations of Tarbali and Bradley (2015), and use a normalized weight of 0.7 for the SA ordinates, with an equal marginal weight assigned to each ordinate, and a weight of 0.3 assigned to the additional IM.

- We assume that all the IMs are lognormally distributed, although this is not a requirement of the GCIM (Bradley et al., 2015). A target lognormal distribution of an IM at a distance R is obtained from all the simulated records at the same distance. Ideally, such a distribution would 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 points from the target IM distributions, and evaluate the ground motion candidates against the target distribution sample using the square of the error between each target point, and all the available candidate points. To reduce the computational effort, we only evaluate candidate records which correspond to JB distances between 0 and 15 km. The normalized squared error associated with $n\ IMs$ is:

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

  where $ln(IM_{i-tar}^{k})$ is the natural log of the k^{th} point of the target distribution sample of IM i, $ln(IM_{i-cand}^{j})$ 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 i. Equation 2 is used to obtain records that best represent 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 (K-S 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:

$$R_{IM} = \sum_{i=1}^{n} w_i (D_{IM_i})^2 \tag{3}$$

- We repeat this sampling process and select the best-fit statistical replicate which yields the lowest global residual R_{IM}. We use 30 statistical replicates to select 15 ground motions in each experiment. These numbers satisfy and exceed the minimum recommendations by Tarbali and Bradley (2015) who studied the impact of the number of replicates and selected ground motions on the stability of the residuals, and suggested using at least 10 replicates to select 10 ground motions, and at least five replicates to select 20 ground motions.

- Because our target IM distributions correspond to a group of simulated ground motions, we are able to measure the resulting bias in the MIDR imposed by the final set of selected

records: Using the K-S test, we compare the MIDR distribution associated with the selected records against a cumulative distribution function fitted to the reference MIDR values associated with 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.

3.2.2 Ground motion selection results

Figure 5 presents a summary of the bias in the MIDR (K-S test statistic) associated with ground motion selection experiments conducted to represent a target of only pulse records at 1 and 5 km for both the 3-story, 12-story, and 20-story buildings. The x-axis indicates the target IMs used in the selection process, where "SA" is used in reference to the full spectral shape consisting of the SA ordinates at 19 periods. The color coding in the figure indicates differences in the candidate pool of records used in the selection process; either (1) a mixed record pool that does not distinguish between pulse and non-pulse records, or (2) a pool of pulse records only. Each box plot summarizes the distribution of the MIDR K-S test statistic from 30 experiments. The critical value of the K-S test is also shown in each figure to illustrate the acceptable values (i.e., K-S statistic is below the critical value). The analogous plots for the RIDR and MFA demands are shown in Figures 6 and 7, respectively. Based on the results summarized in the figure, the following observations can be made:

- Selecting records from a pool of mixed candidates (pulse and non-pulse) does not appear to introduce additional systematic bias into the predicted structural demands compared to a specialized pool of pulse records, even in the extreme case where the target IM distributions are obtained only from records with strong velocity pulses. This finding is applicable for all buildings and all the near-fault distances we explored. We tested the hypothesis that each pair of samples of MIDR K-S test statistics (representing the bias associated with the use of mixed and pulse candidate records) comes from the same distribution using the two-sample K-S test. The p-values of the tests (shown in Table 3) indicate that there is no statistically significant difference between each pair of samples for any of the target IMs. Similar results were obtained for the majority of the RIDR and the MFA distributions associated with the selected suites of records, as shown in Tables 4 and 5. Based on the majority of the p-values associated with the two-sample K-S test, there is no evidence in the data to suggest that the distributions of the bias associated with each pair of RIDR and MFA demands are significantly different at the 0.05 significance level. The exceptions are the distribution of the RIDR for the 3-story building at 1 km corresponding to records selected based on SA + $maxIP(T_1)$ (p-value = 0.032), and the distribution of the MFA on the 12-story building at 1 km selected based on SA + T_p (p-value = 0.009) where record selection from a mixed pool of records results in lower bias than using a pulse-only pool of records.

- Using an additional IM as a target along with the spectral shape does not appear to offer a systematic reduction of the bias associated with the MIDR, RIDR or MFA demands corresponding to pulse records. The non-parametric Kruskal-Wallis one-way analysis of variance and the Dunnett post-hoc test were used to test whether the medians of each distribution differs significantly from the median of the distribution corresponding to using the spectral accelerations only and selecting records from a pool of pulse-type records only. The results of the test indicate that none of the distributions associated with other IMs or mixed pool of records have a significantly different median, with the exception of using SA + $maxIP(T_1)$ for the 3-story building at 5 km, which appears to reduce the bias associated with the MIDR demands in this case (p-value = 0.031). However, most of the demands imposed on all three buildings at 1 km and at 5 km and selected using any of the IM combinations are associated with acceptable bias (i.e., below the critical value of the K-S test), as shown in Figures 5 through 7. Therefore, the differences between the distributions associated with each IM combination may not be of practical importance.

- In alignment with the results of the preceding section indicating the insensitivity of the MIDR to T_p of records with similar spectral shapes, adding T_p as a target IM does not appear to reduce the bias in the MIDR associated with ensembles of selected records. However, incorporating the T_p as an additional IM (the SA + T_p case) appear to substantially increase the MFA demands on the 12-story building at 1 km, compared to the use of SA alone (Dunnett post-hoc test, p-value = 0.00006). A possible reason for this observation is that any potential reduction in the bias due to incorporating T_p may be offset by an increase in the bias due to reducing the weight of the SA ordinates in the selection procedure.

Table 3: P-Values for the two-sample K-S test on pairs of samples representing the MIDR bias associated with selected ensembles of pulse records or mixed records

	3-story Building		12-story Building		12-story Building	
Target IMs	1 km	5 km	1 km	5 km	1 km	5 km
SA	0.508	0.155	0.508	0.508	0.678	0.241
SA+PGV	0.841	0.954	0.056	0.241	0.508	0.508
SA+$max\ IP(T_1)$	0.678	0.155	0.241	0.508	0.155	0.678
SA+D_{s595}	0.841	0.508	0.358	0.841	0.241	0.358
SA+T_p	0.678	0.508	0.095	0.841	0.954	0.954

Figures 8 and 9 show example outcomes of the lowest-bias ground motion ensembles (out of 15 statistical replicates) selected to represent target IM distributions consisting of pulse records only at a distance of 1 km. The subplots correspond to three separate cases: (a) a target consisting of SA ordinates only, and a mixed pool of candidate records (subplots a), (b) a target consisting of SA ordinates only, and a pool of candidate records which contains pulse records only (subplots b), and (c) a target consisting of SA ordinates and T_p distribution, and a pool of candidate records

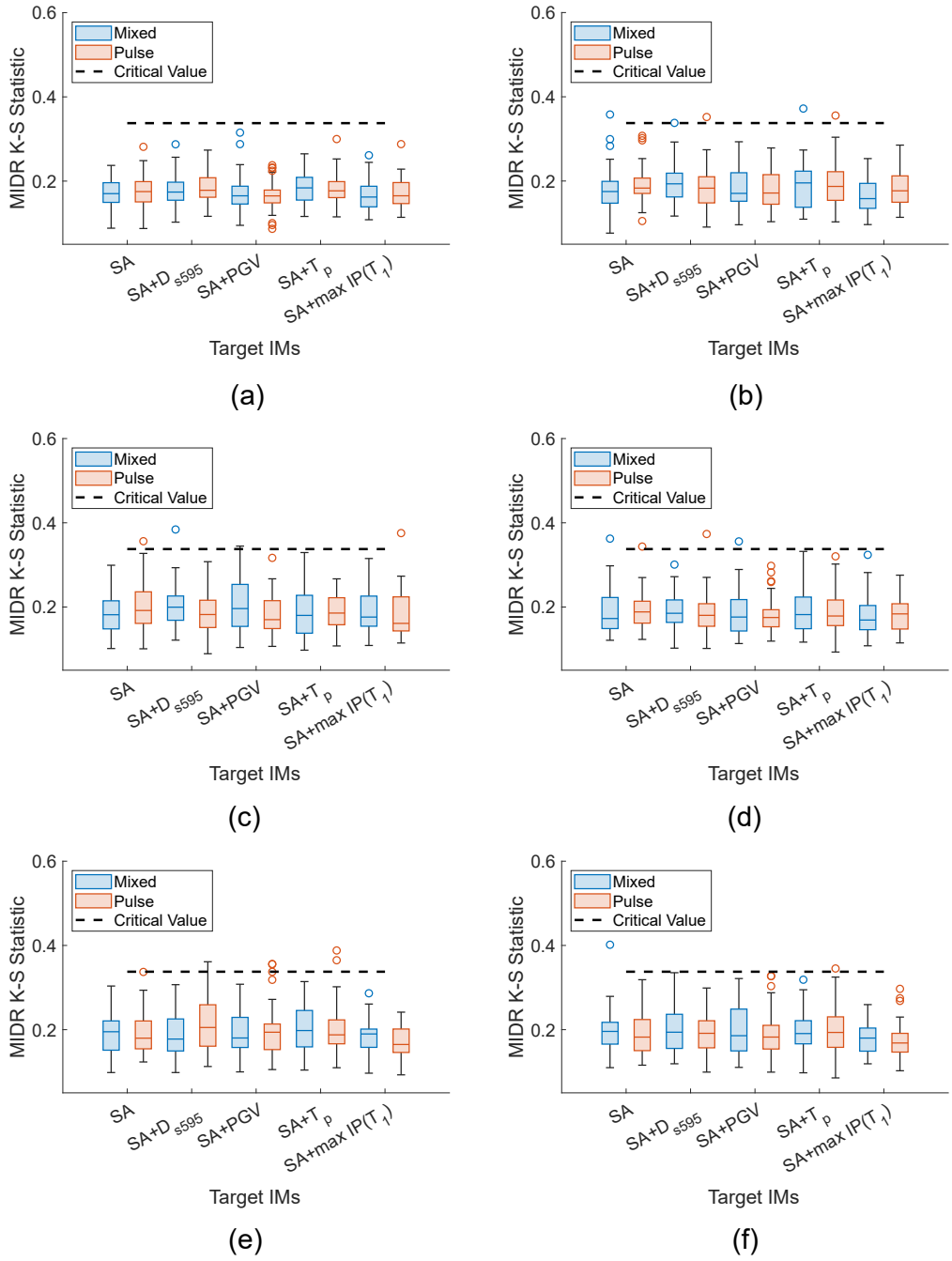

Figure 5: Values of the K-S test statistic of the distribution of MIDR imposed by earthquake records selected to represent a target distribution of only pulse records for: (a) a 3-story building at 1 km; (b) a 3-story building at 5 km; (c) a 12-story building at 1 km; (d) a 12-story building at 5 km; (e) a 20-story building at 1 km; (f) a 20-story building at 5 km. Each box plot represents the statistics of 50 replicates. The number of selected records in each case is 15.

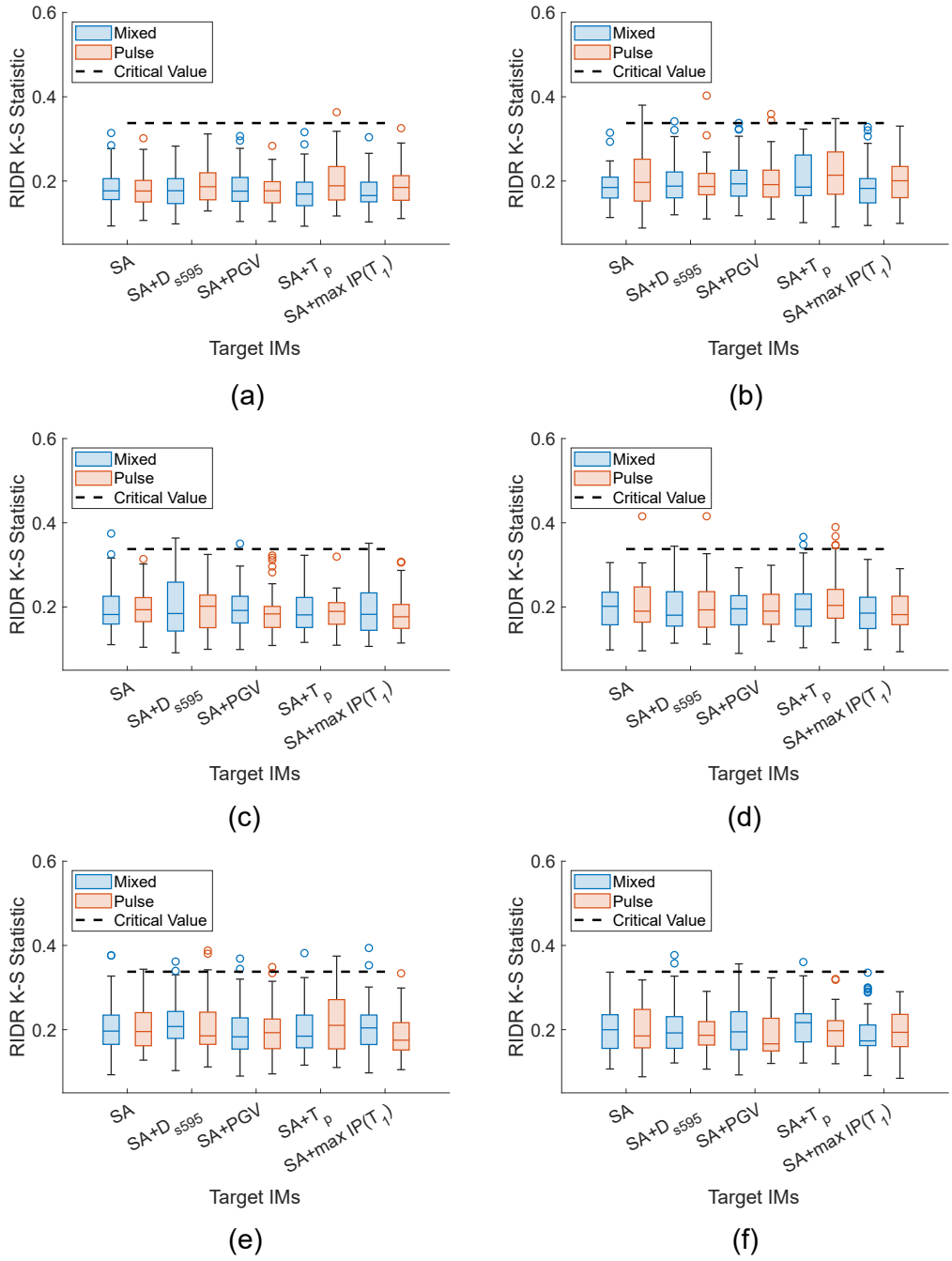

Figure 6: Values of the K-S test statistic of the distribution of RIDR imposed by earthquake records selected to represent a target distribution of only pulse records for: (a) a 3-story building at 1 km; (b) a 3-story building at 5 km; (c) a 12-story building at 1 km; (d) a 12-story building at 5 km; (e) a 20-story building at 1 km; (f) a 20-story building at 5 km. Each box plot represents the statistics of 50 replicates. The number of selected records in each case is 15.

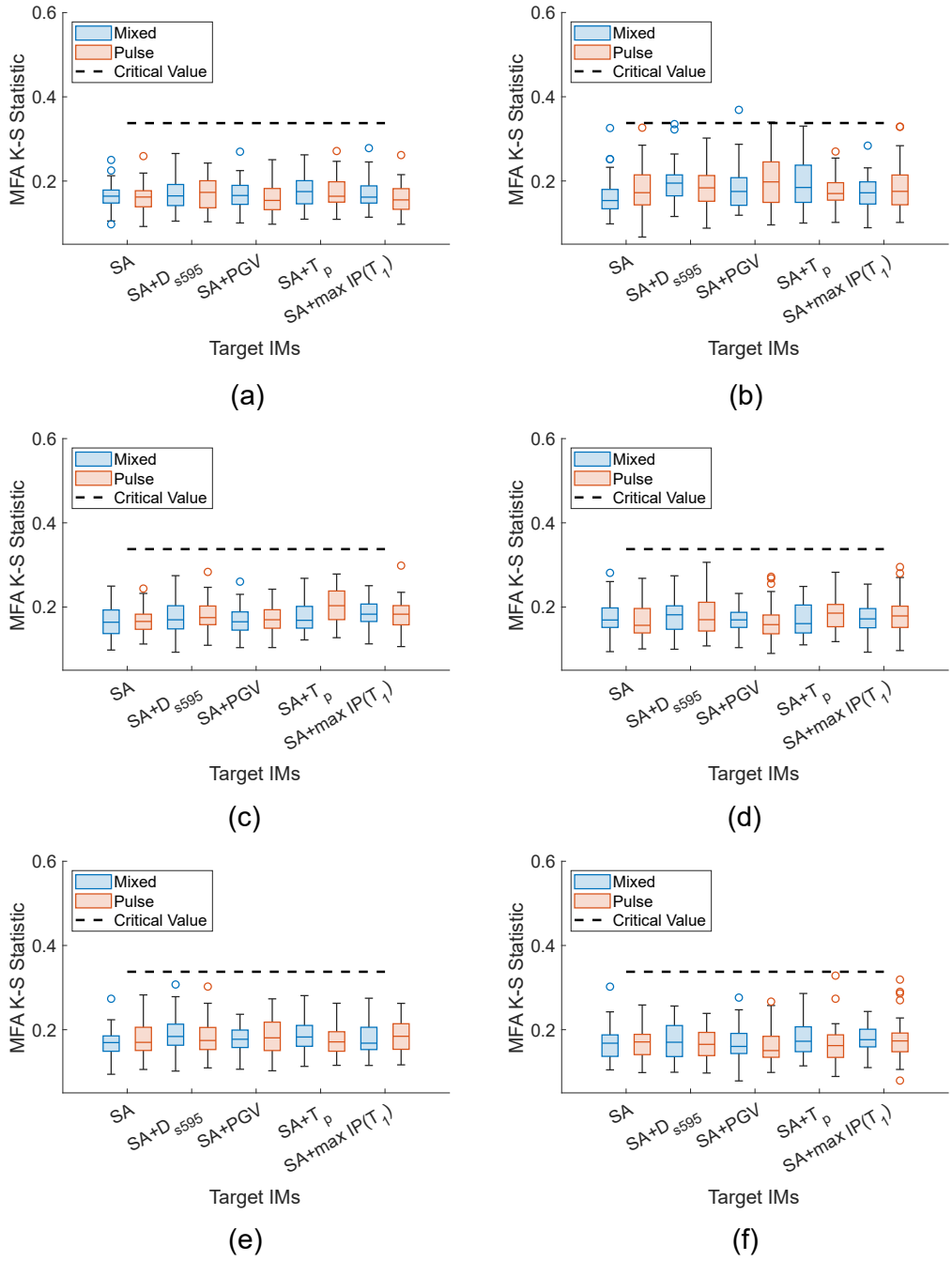

Figure 7: Values of the K-S test statistic of the distribution of MFA imposed by earthquake records selected to represent a target distribution of only pulse records for: (a) a 3-story building at 1 km; (b) a 3-story building at 5 km; (c) a 12-story building at 1 km; (d) a 12-story building at 5 km; (e) a 20-story building at 1 km; (f) a 20-story building at 5 km. Each box plot represents the statistics of 50 replicates. The number of selected records in each case is 15.

Table 4: P-Values for the two-sample K-S test on pairs of samples representing the RIDR bias associated with selected ensembles of pulse records or mixed records

	3-story Building		12-story Building		12-story Building	
Target IMs	1 km	5 km	1 km	5 km	1 km	5 km
SA	0.841	0.095	0.841	0.841	0.954	0.841
SA + PGV	0.678	0.954	0.155	0.678	0.508	0.508
SA+$max\ IP(T_1)$	0.032	0.358	0.841	0.841	0.095	0.241
SA+D_{s595}	0.678	0.678	0.508	0.841	0.056	0.678
SA+T_p	0.095	0.508	0.841	0.508	0.358	0.155

Table 5: P-Values for the two-sample K-S test on pairs of samples representing the MFA bias associated with selected ensembles of pulse records or mixed records

	3-story Building		12-story Building		12-story Building	
Target IMs	1 km	5 km	1 km	5 km	1 km	5 km
SA	0.954	0.056	0.241	0.155	0.241	0.841
SA + PGV	0.095	0.155	0.508	0.155	0.155	0.155
SA+$max\ IP(T_1)$	0.155	0.358	0.508	0.508	0.358	0.358
SA+D_{s595}	0.678	0.508	0.678	0.954	0.508	0.508
SA+T_p	0.954	0.241	0.009	0.155	0.241	0.155

which contains pulse records only (subplots c). The top panels of Figure 8 display the acceleration spectra of the selected records along with those of the target pulse records. The plots show that all three selection approaches generate ground motion records which are relatively consistent with the median, 16th and 84th percentile spectra of the target records. However, incorporating T_p as an additional intensity measure (subplots c) leads to a larger difference between the target and selected spectral accelerations, especially in the short-period range. This is likely the case because less weight is given to the spectral acceleration in case (c), and matching the distribution of T_p (with a normalized weight of 0.3) may favor ground motions with higher long-period spectral energy. As a result, the associated distributions of the 12-story MIDR corresponding to the selected and target records shown in Figure 9 exhibit similar and statistically acceptable bias with respect to the K-S test critical value marked with dashed lines in the plots, and those similarities are echoed in the IDR envelopes shown in the bottom panel of the figure. Specifically, despite T_p not being explicitly considered in the selection process in subplots a and b, the median, 5th percentile and 95th percentile IDR envelopes along the building height are in good agreement with the target distribution, regardless of whether the underlying pool contains mixed records or pulse records only (and even though only 15 records are selected to represent the target). This is due to the fact that the interstory drift ratios on this building are most influenced by the spectral shape of the

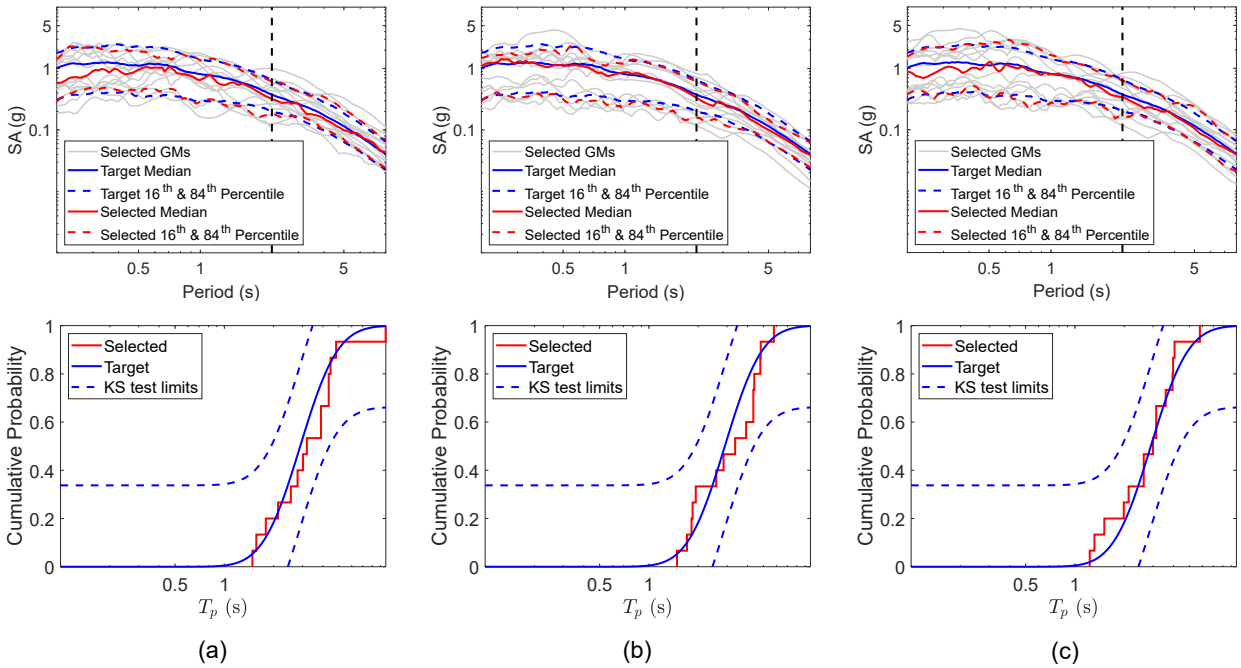

Figure 8: Selected and target ground motion spectra, and the distribution of T_p corresponding to target and selected ground motion records using the following target intensity measures and underlying candidate pools: (a) SA target and a pool of mixed candidate records; (b) SA target and a pool of pulse candidate records; (c) SA and T_p target, and a pool of pulse candidate records.

records near the building's first-mode period, which is well represented in all three cases. However, the MFA demands on the building exhibit larger bias in case (c), because of the lower short-period spectral accelerations in this case. Examining the corresponding distribution of T_p associated with each selection case in the bottom panels of Figure 8 reveals that the distribution of T_p in the target records is reasonably represented by all selection approaches, and is within the acceptable limits of the K-S test, despite not being explicitly considered in the selection processes in cases (a) and (b).

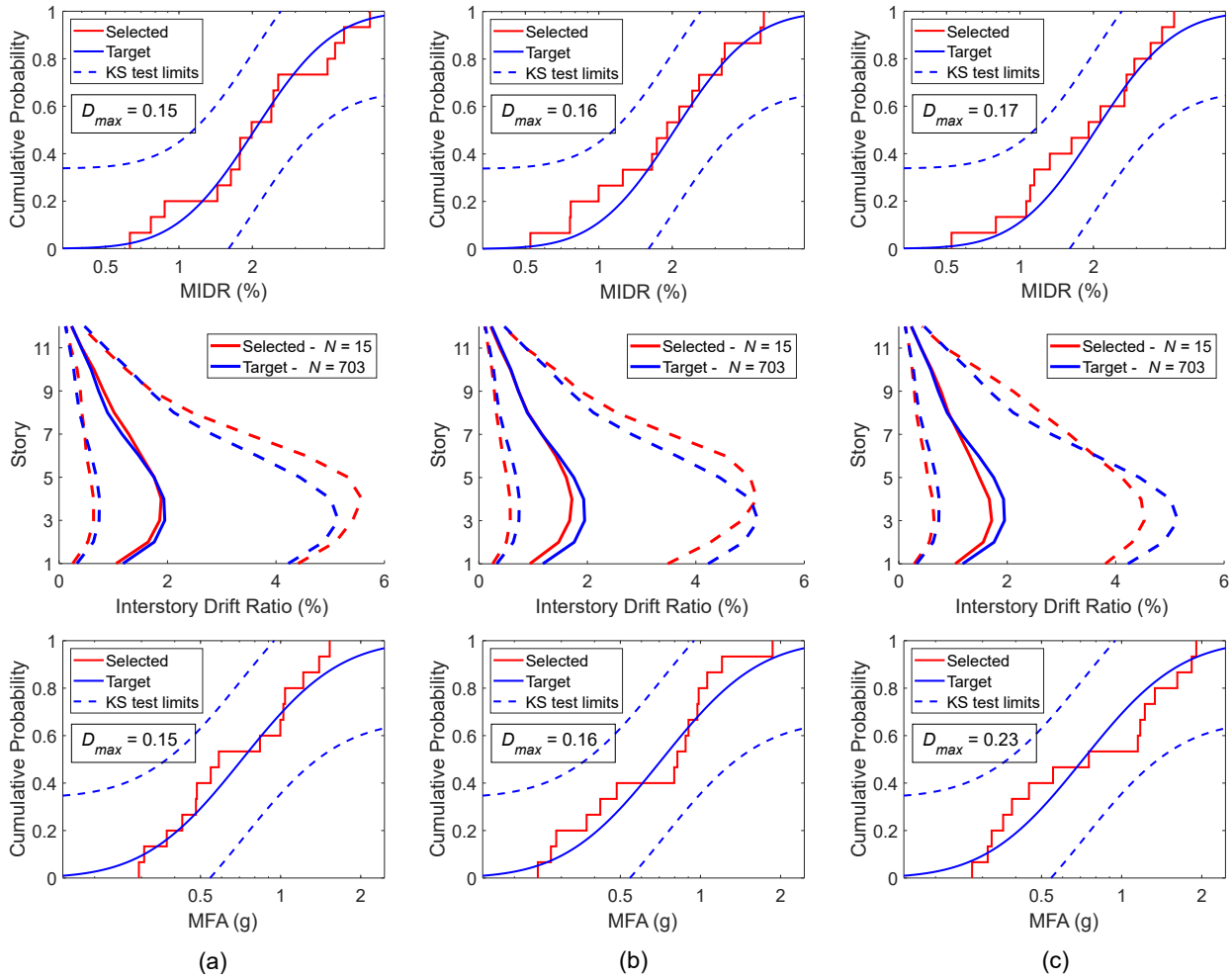

Figure 9: MIDR distributions, IDR envelopes (median, 5th percentile and 95th percentile), and MFA distributions for the 12-story building using the following target intensity measures and underlying candidate pools: (a) SA target and a pool of mixed candidate records; (b) SA target and a pool of pulse candidate records; (c) SA and T_p target, and a pool of pulse candidate records.

4 Consideration of Fling Effects in the Analysis of Near-Fault Structures

4.1 Extraction of fling features

The presence of fling is characterized by a one-sided velocity pulse, and a corresponding permanent offset in the ground displacement time history. The fling properties have been commonly described by three parameters: the static offset at the end of the displacement history D_p, the period of the fling pulse T_p, and the arrival time of the fling pulse t_1. We extracted the three parameters from all the simulated ground motion records using the method recommended by Burks (2014), which fits the following shifted-sine function to the displacement time history:

$$d_{shifted-sine}(t) = \frac{D_p}{2} sin\left[\frac{\pi}{T_p}\left(t - t_1 - \frac{T_p}{2}\right)\right] + \frac{D_p}{2} \tag{4}$$

The parameter D_p is readily available, and the pattern search optimization algorithm was used to extract T_p and t_1 by minimizing the sum of the squared error between the displacement time history, and the function in equation 4. Examples of the ground displacement and corresponding extracted pulses are displayed in Figure 10 for simulated and recorded ground motions. We found that the pattern search algorithm, recommended by Burks (2014), performed generally well for most records with moderate to large permanent offsets. However, in cases with negligible permanent offset in the displacement record, the algorithm did not perform favorably. Specifically, excessively long T_p was predicted by the algorithm in finding the best function fit to the displacement record. Similar observations were made by Burks (2014). In addition, we also observed that the algorithm was not appropriate in cases where the displacement time history was dominated by several pulses, as opposed to a single permanent offset. In these cases, the algorithm also tended to predict very long pulse periods, which in reality correspond to several pulses, not a single dominating pulse. Two examples of such displacement records are shown in Figure 10c and d from the group of recorded and simulated ground motions, respectively.

4.2 Comparison against recorded ground motions

We compared the magnitude of fling displacements in the simulated ground motions against the recently published NESS2.0 (NEar-Source Strong-motion) database (Sgobba et al., 2021). This database contains the properties and intensity measures of uniformly processed near-fault strong motions recorded worldwide, and includes the amplitudes of the fling step. The records were processed using the extended baseline correction (eBASCO) processing scheme (D'Amico et al., 2018), which preserves the offset at the end of the displacement traces. The fling displacements of the near-fault simulated ground motions were compared against those of a subset of the NESS2.0 records with moment magnitudes between 6.4 and 7.6 (mean= 6.95; median = 6.90), and for rupture distances less than or equal to 20 km. This dataset includes 194 records from 24 unique events worldwide. The depth to the top of the fault ranges between 0 and 5.2 km.

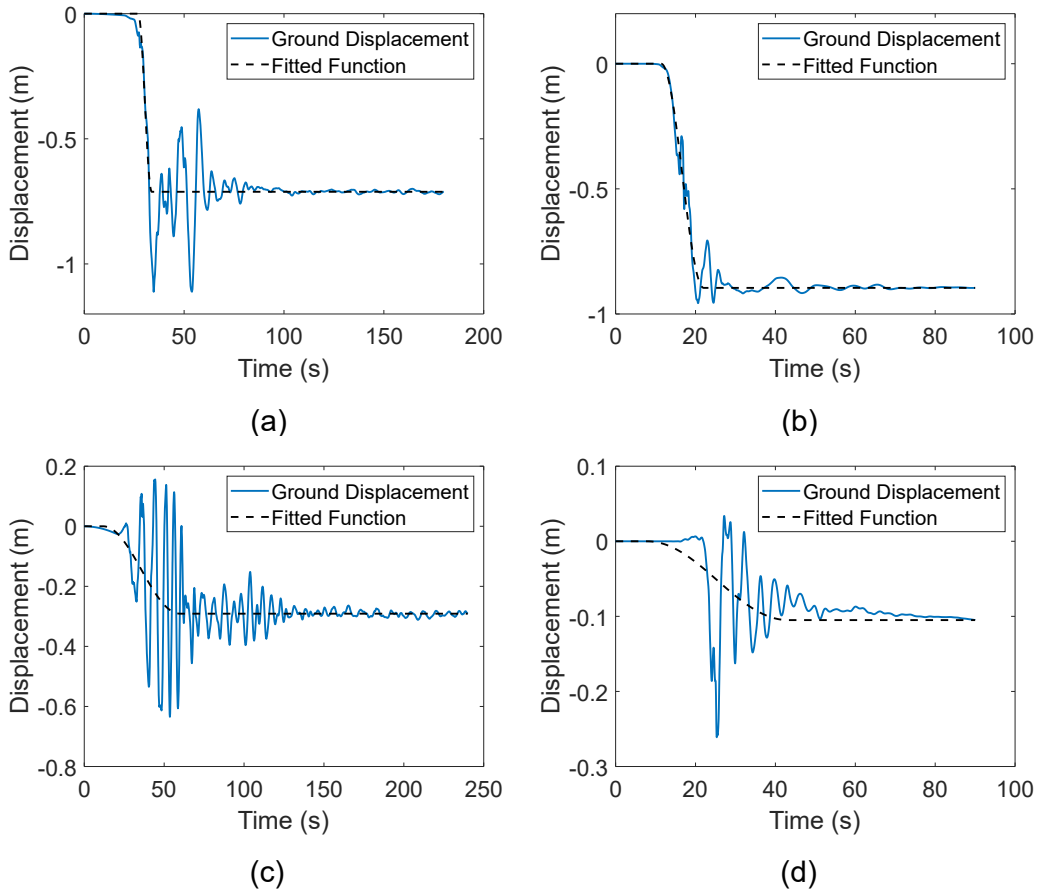

Figure 10: Examples of recorded and simulated ground motions with permanent displacements and the corresponding function fit: (a) Rotated maximum horizontal component of station TCU051 of the 1999 M7.62 Chi-Chi, Taiwan earthquake (R_{rup} = 7.6 km, T_p = 6.9 s - D_p = 0.7 m); (b) Rotated maximum horizontal component of a simulated ground motion from realization H (R_{rup} = 3.0 km; T_p = 11.4 s; D_p = 0.9 m). (c) North-south component of station CHY104 of the 1999 M7.62 Chi-Chi, Taiwan earthquake processed by Burks and Baker (2016) (R_{rup} = 18.0 km, T_p = 48.3 s - D_p = 0.3 m); (d) Rotated maximum horizontal component of a simulated ground motion from realization H (R_{rup} = 10.8 km; T_p = 35.3 s; D_p = 0.11 m). Subplots (a) and (d) show the detected long pulses, whereas (c) and (d) show the function fitting multiple shorter pulses.

Figure 11a shows the data points corresponding to the magnitude of the maximum horizontal permanent offset of the simulated and recorded ground motions for rupture distances less than or equal to 20 km, with color coding indicating the moment magnitude of each recording. The magnitude of the maximum permanent offset is defined as the resultant of the ground displacement at the end of the shaking time history of both horizontal components of the ground motion record. Overall, the rupture simulations appear to produce permanent offsets that are consistent with those of the recorded near-fault ground motions. We note, however, that the M7.0 simulated ground motions tend to have somewhat lower permanent offsets than recorded ground motions corresponding to M7.0 events; this is likely due to the fact that most of the simulated ruptures are buried at 200 m, whereas 75% of the recorded ground motions shown in the figure correspond to ruptures shallower than 200 m. Previous studies have established that fling pulses attenuate in an inversely proportional fashion to the square of the distance (Hisada and Tanaka, 2021); therefore, shallower rupture depths are expected to be associated with larger fling step amplitudes. The figure also shows the kernel moving mean of the simulated ground motions, along with two empirical models: an empirical model developed by Burks and Baker (2014) (referred to as BB-2014) and a recent model developed by Schiappapietra et al. (2022) based on the NESS2.0 database (referred to as SLS-2022). The moving mean of the simulations is in remarkable agreement with the distance scaling predicted by the recent SLS-2022 model for distances less than approximately 5 km. For distances larger than 5 km, the permanent offsets in the rupture simulations attenuate faster than the SLS-2022 model predictions, likely due to the fact that the SLS-2022 empirical model is developed based on a database where 51% of the records correspond to ruptures shallower than 200 m. The moving mean of the simulations is in excellent agreement with the distance scaling predicted by the BB-2014 model for distances longer than 5 km. It is worth noting, however, that the BB-2014 model is based on a smaller number of recorded ground motions compared with the SLS-2022 model, and was developed based on a single-component fling magnitudes (SN or SP), whereas the SLS-2022 model predicts the maximum horizontal fling amplitudes, which more appropriately represent the plotted individual data points.

The NESS2.0 database does not contain information regarding the fling pulse periods, but Burks and Baker (2016) extracted T_p from a small number of recorded ground motions, which they processed to preserve permanent offset in the displacement time history (Burks and Baker, 2015). The Burks-Baker database of near-fault recorded ground motions with magnitudes relevant to our study consists of 42 records, 40 of which represent different station recordings from the Chi-Chi Taiwan earthquake, and a single station from the Kocaeli, Turkey earthquake, and the Darfield, New Zealand earthquake. The Burks-Baker database includes a single horizontal component for each station, and two orthogonal components in some limited cases. To conduct a consistent comparison between the fling characteristics of both simulated and recorded ground motions, we rotate each ground motion record to the orientation that coincides with the maximum amplitude of the ground displacement at the end of the time history. In cases where only a single horizontal component is provided in the Burks-Baker database, that component was assumed to coincide with the direction of the maximum static offset. Effectively, this means that we assume that the static offset in the orthogonal and unavailable direction is negligible. An additional data point representing the permanent offset associated with the Lucerne station of the Landers earthquake

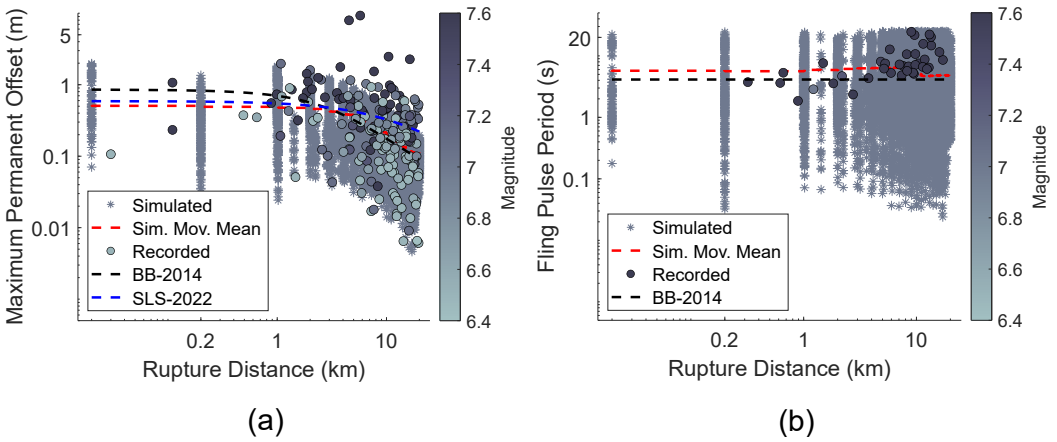

Figure 11: (a) Maximum permanent offset in simulated and recorded ground motions; (b) Period of the fling pulse in the direction of maximum offset extracted from simulated and recorded ground motions. The dashed red line represents the kernel moving mean of the simulated ground motions, the dashed black line represents the predictive model of Burks and Baker (2016), and the dashed blue line represents the predictive model of Schiappapietra et al. (2022).

was obtained from the Cosmos Virtual Database Center (Archuleta et al., 2005). The complete dataset consists of ground motions recorded at rupture distances between 0 and 20 km, and spans moment magnitudes between 7.0 and 7.62, but it is primarily dominated by the 7.62 Chi Chi earthquake. V_{s30} for the recorded ground motions range between 277.5 m/s and 1369 m/s, with a mean value of 510.4 m/s, and a median value of 476.1 m/s.

Figure 11b displays the extracted fling pulse periods from this group of recorded ground motions, along with the simulated ground motions, for rupture distances shorter than or equal to 20 km. Based on the issues discussed in the preceding section, we excluded any ground motions with extracted pulse periods larger than 40 seconds, since such periods correspond to multiple pulses as opposed to a single pulse, and, in most cases, correspond to negligible permanent offsets. The figure highlights the small number of available fling pulse periods from recorded ground motions, making it difficult to conduct a consistent comparison against the simulations. The figure also displays the kernel moving mean of the pulse periods of the simulated ground motions, along with the mean fling pulse period predicted by the empirical model of BB-2014. The plot shows that a large number of the simulated ground motions have shorter fling pulses than the recorded ground motions. However, the kernel moving mean suggests that the fling pulses in the simulated ground motions are slightly longer than the predictions of the empirical model. Finally, in agreement with both the field observations and empirical model, the fling pulse periods of the simulated ground motions do not exhibit a clear attenuating trend with distance, and appear to be constant for the M7.0 rupture, with wide variability around the mean value.

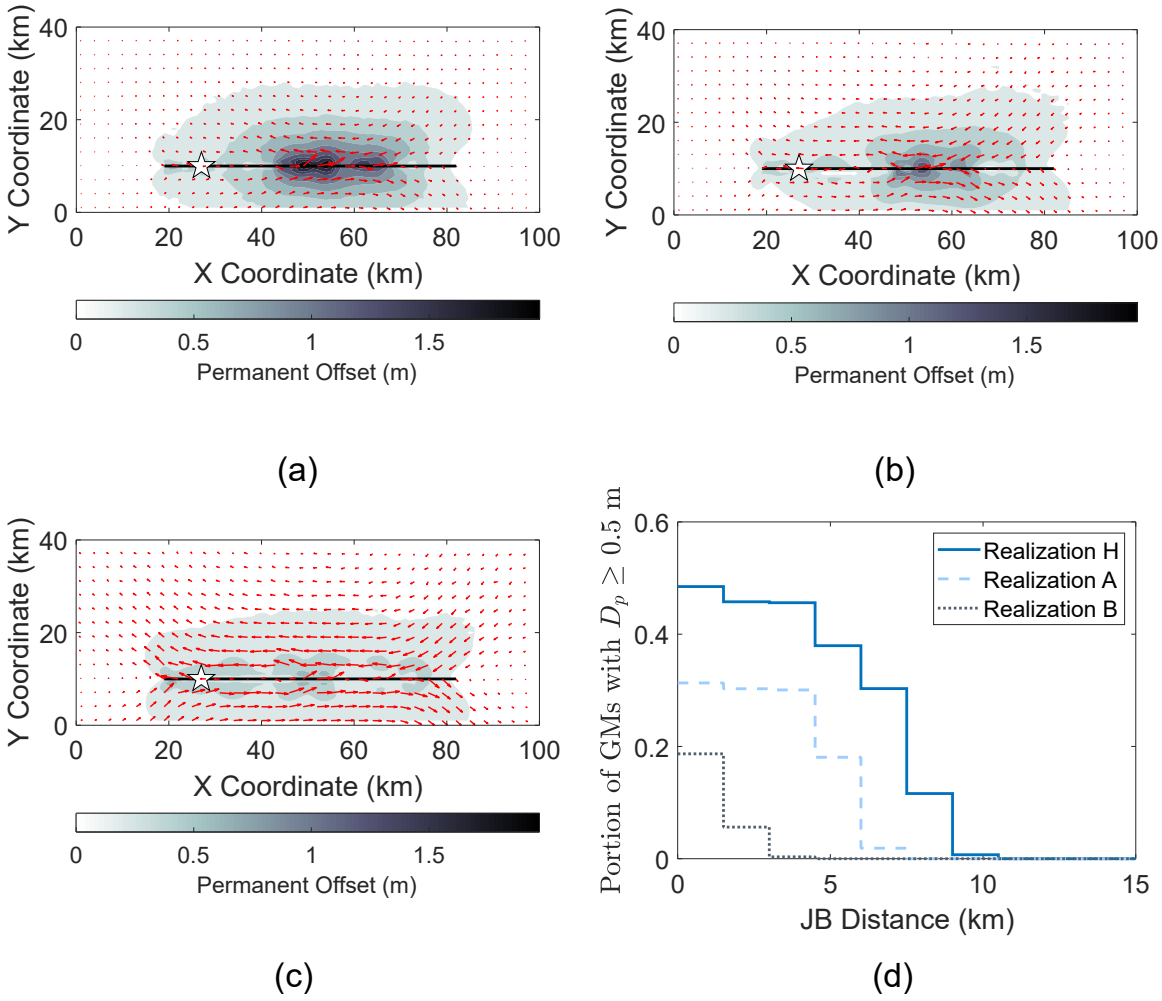

Figure 12: Magnitude and direction of maximum static offset in representative fault rupture simulations: (a) Realization H; (b) Realization A; (c) Realization B; (d) percentage of ground motion records with D_p larger than 0.5 m in each of the three realizations.

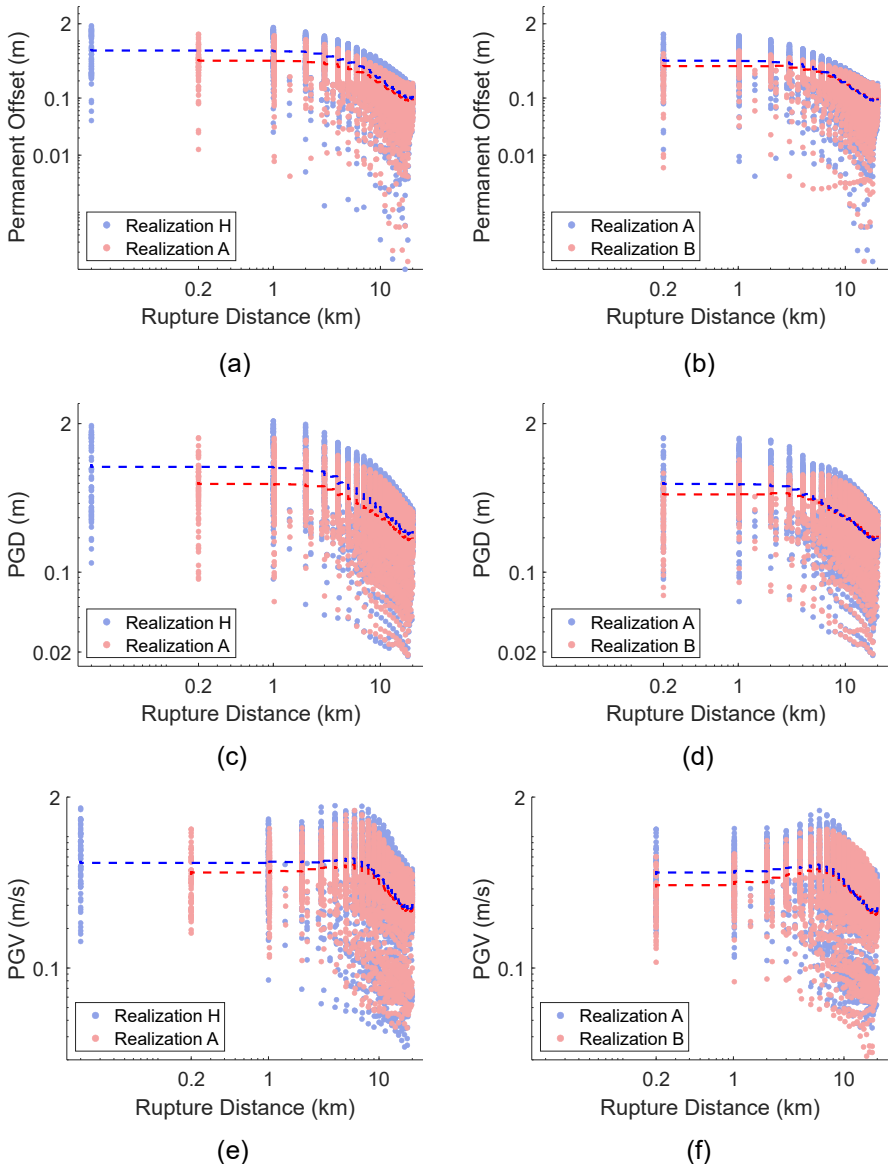

Figure 13: Distance scaling of the intensity of SP ground motions for distances less than 20 km and different rupture realizations as indicated in the legend: (a) and (b) magnitude of the permanent offset at the end of the record; (c) and (d) peak ground displacement; (e) and (f) peak ground velocity.

4.3 Variability of fling characteristics

An important goal of this study is to examine the influence of the rupture simulation parameters on the fling characteristics of simulated near-fault ground motions, and provide guidance for generating future simulations in engineering analysis cases where fling effects are relevant. Despite having a relatively small number of simulated realizations, we qualitatively study the effects of the controlled rupture characteristics on the magnitude of the permanent offset in the near-fault ground motions. The magnitudes and directions of the maximum horizontal permanent offset in representative rupture realizations are shown in Figure 12a, b and c, where the displacement direction represents the orientation which maximizes the horizontal offset at the end of the ground shaking history. The directions of the displacements in all realizations are qualitatively very similar to those of the 1999 M7.1 Hector Mine earthquake mapped by Fialko et al. (2001). The displacements exhibit the following features: (1) along the fault length, the displacement direction is parallel (or close to parallel) to the fault, (2) near the ends of the fault the displacement direction diverges from the fault-parallel direction, and (3) the displacements are primarily anti-symmetric with respect to the fault plane. The maximum displacement magnitude in our simulations is approximately 2.0 m for ruptures at a very shallow depth of 20 m, such as realization H in Figure 12a. This magnitude is, however, substantially lower than that observed in the Hector Mine earthquake, which was reported by Behr et al. (2000) and Fialko et al. (2001) to be approximately 5.5-6 meters. Based on the analysis in the previous section, this difference may be partially attributed to the fact that our rupture simulations are buried (albeit at relatively shallow depths). We note that the magnitudes of fling permanent displacements are nearly the same on the rock and basin soils, and are found to be not sensitive to changes in the velocity structure, which agrees with the theoretical description of fling using static terms in Green's function (Hisada and Bielak, 2003), as opposed to the directivity pulses which are represented by dynamic terms. We also note that changing the depth to the top of the rupture in our simulations from 20 m to 200 m, with everything else being equal, results in a 31% reduction in the magnitude of the maximum permanent offset (Figure 12b), and a substantial reduction of other ground motion intensity measures as well in the near-fault region. For instance, Figure 13a, c and e show the distance scaling of the permanent offset, peak ground displacement, and peak ground velocity of the SP component of both realizations H and A across distances up to 20 km, along with the kernel moving mean for each realization. The figure reveals that all three intensity measures are notably higher for realization H particularly for distances less than 10 km. The shallower rupture also results in a larger number of ground motions with substantial permanent offset (defined to be equal to 0.5 m or larger). The portion of ground motions with $D_p \geq 0.5$ meters is visualized in Figure 12d for each of the three realizations for JB distances up to 15 km. Increasing the depth of the rupture from 20 m to 200 m results in reduction between 33% to 100% of the number of ground motions with $D_p \geq 0.5$ m, within a distance of 9 km from the fault.

In addition, other characteristics of the rupture, specifically the presence and location of the rupture asperities, can significantly influence the magnitude of permanent displacements in the ground motion time history. This influence is visualized via subplots b and c of Figure 12, which contain the permanent displacements corresponding to two identical rupture realizations with a single exception: realization A contains a patch of large slip near the surface, and realization B

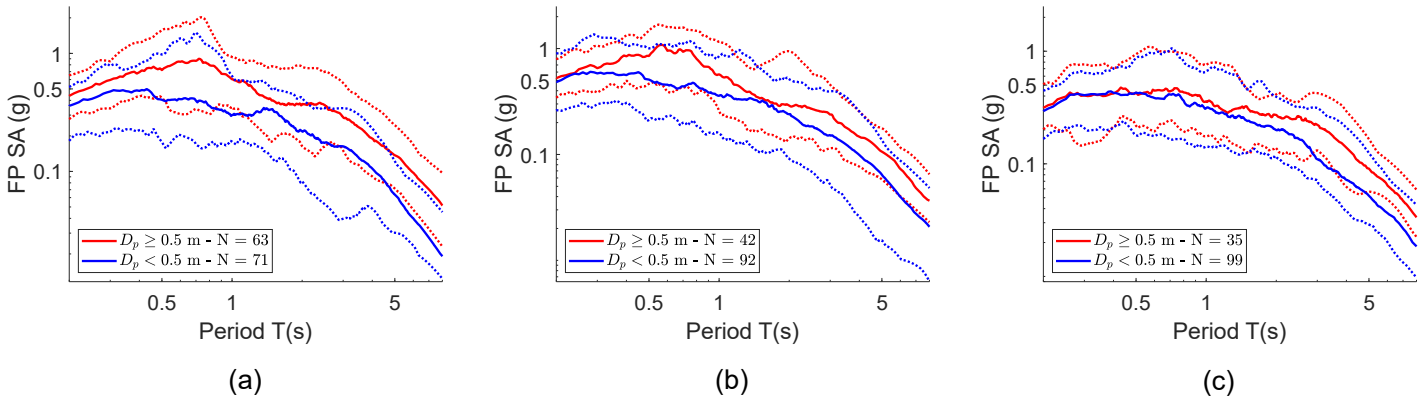

Figure 14: Median (solid line), and 5th and 95th percentile (dashed lines) pseudo-acceleration response spectra of the FP (or equivalently the SP) simulated ground motions at a JB distance of approximately 1 km with static offset exceeding or below 0.5 meters in representative simulations: (a) Realization H; (b) Realization A; (c) Realization B.

does not. The absence of the slip patch results in a reduction of 45% in the magnitude of the fling displacements in realization B, and a reduction between 40 to 100% in the number of records with substantial fling displacement within 6 km of the fault (visualized in Figure 12d). We also found that the D_p, peak ground displacement (PGD) and PGV associated with rupture realizations without slip patches may be substantially lower than those with a shallow slip patch in the near-fault region. In this case, the shallow slip patch in our simulations increases the D_p, PGD and PGV for distances less than 5 km, as featured in Figure 13b, d and f for realizations A and B.

Our analysis also reveals that ground motions with larger permanent offset generally tend to have higher intensity than those with smaller offsets, given the same magnitude and distance parameters. We examined the spectral shapes of the ground motion records with fling step longer than or shorter than 0.5 m from different realizations, particularly in the SP component of the horizontal ground motions, where we primarily observed fling step in our strike-slip simulations with a vertical fault. Ground motion records with substantial fling displacements were found to have higher spectral energy, possibly across a broad range of periods. Figure 14 shows the median and the 5th and 95th percentile spectra for ground motions at a JB distance of 1 km from the three representative realizations H, A, and B, divided into two groups based on whether the associated permanent offset is shorter than 0.5 meters. The plots clearly demonstrate that ground motions with larger fling step have higher spectral intensity than those with shorter fling step at the same near-fault distance. The higher spectral intensity is observed across all periods between 0.2 s and 8 s, especially for realization H which represents the shallowest rupture, and is substantial for periods longer than 1 s for realization B, which is a deeper rupture without distinct slip patches. Overall, a relatively strong correlation between the natural logarithm of the magnitude of the fling displacement and of spectral accelerations is detected in our simulations. Representative data points are displayed in Figure 15a, b and c for the spectral accelerations at the first-mode periods of the three buildings used in this study. The data points represent the SP component of ground motion records

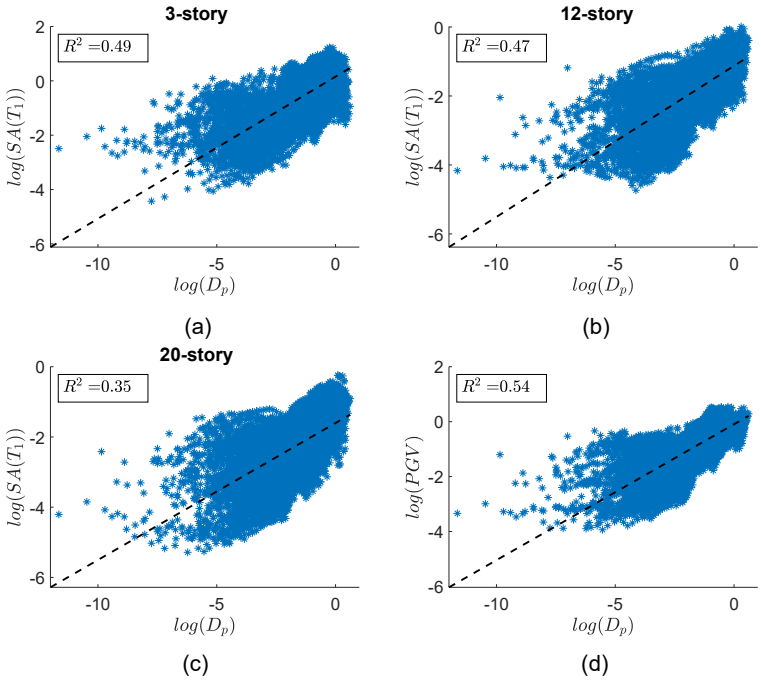

Figure 15: Relationship between the natural logarithm of the magnitude of permanent offset and that of other ground motion intensity measures: (a) spectral acceleration at the first-mode period of the 3-story building; (b) spectral acceleration at the first-mode period of the 12-story building; (c) spectral acceleration at the first-mode period of the 20-story building; (d) peak ground velocity. A linear regression fit and the associated coefficient of determination are also shown in each subplot.

from all rupture realizations within a 20-km JB distance. A linear regression model fit and the corresponding coefficient of determination R^2 value are also displayed, showing that there is a relatively strong linear trend between $ln(D_p)$ and the spectral acceleration at all three periods, and that the dispersion in the linear relationship increases for longer periods. A similar linear correlation is found between the natural logarithm of the fling step and that of the peak ground velocity, shown in Figure 15d.

We note that there is a much larger dispersion in the linear relationships between $ln(Dp)$ and the spectral accelerations of the SN components of the same ground motion records (R^2 values between 0.04 and 0.09, representing a reduction of 73% to 92%, compared to the same D_p-IM correlations for the SP component of the ground motions), because the intensity of the near-fault SN component ground motions in strike-slip vertical-fault ruptures is dominated by rupture directivity effects that result in high spectral energy with negligible permanent offsets. It is, therefore, important to examine whether consideration of ground motions with rupture directivity effects (without fling-dominated ground motions) could be sufficient for representing the expected near-fault ground motion intensity for the analysis of near-fault structures that do not directly cross a fault. In other words, is the influence of fling effects captured implicitly in other ground motion intensity measures, such as spectral shapes and peak ground velocities, or does the fling pulse need to be explicitly considered when selecting ground motions for engineering analysis? In the following sections, we address this question by examining the statistical significance of the fling parameters as predictors of the seismic structural demands, and conducting ground motion selection experiments with and without accounting for the effects of fling.

4.4 Significance of fling parameters as predictors of structural demands

We examined the sensitivity of the structural demands to two parameters which characterize fling pulses: the amplitude of the permanent offset D_p in the ground displacement and the period of the fling pulse T_p, in addition to the peak ground displacement. Initially, a regression model was developed relating a primary IM and an engineering demand parameter (EDP) of the following form: $ln(EDP) = a + b * ln(IM)$, where the spectral acceleration at the first-mode period of the structure was used as the primary IM. We then tested whether the residuals of the EDP-IM relationship exhibit dependence on D_p or T_p. Based on the associated p-values, the residuals of the EDP-$SA(T_1)$ relationship were found to depend on D_p in most cases for the three EDPs we considered: MIDR, RIDR and MFA, but the magnitude and direction of the correlation varied widely depending on the dynamic properties of the building, and on whether the SP near-fault ground motions or all near-fault ground motions (both SP and SN) were considered. For instance, there is little correlation between the residuals of the EDP-$SA(T_1)$ model for the 3-story building and D_p (Spearman correlation coefficient between -0.035 and 0.075), and in some cases, the correlation is negative, meaning that ground motions with larger permanent offset imposed lower structural demands. However, there is a weak positive correlation between the EDP-$SA(T_1)$ residuals and D_p for the 12-story building (Spearman correlation coefficients between 0.095 and 0.208), implying that relatively flexible buildings may be more sensitive to fling effects. When we consider only SP ground motions, the correlations between the EDP-$SA(T_1)$ residuals and D_p are substan-

tially stronger, with the highest correlations observed for the 20-story building EDPs (Spearman coefficients between 0.249 and 0.446) - Representative examples of these correlations are shown in Figure 16. These observations suggest that even though fling effects have a measurable influence on the structural demands, directivity effects–if present–may primarily control the near-fault demands.

Similarly, weak to no correlation was noted between the fling pulse periods and the EDP-$SA(T_1)$ residuals in most cases. The strongest correlation was observed between the EDP-$SA(T_1)$ residuals and T_p for the 20-story building subjected to SP ground motions only (Spearman correlation coefficient between 0.061 and 0.276). Finally, the EDP-$SA(T_1)$ residuals were found to positively correlate with the PGD associated with the ground motion records for all three buildings, and the correlations were relatively stable across both the SN and SP ground motions. The correlation coefficients for all the cases we considered are summarized in Figure 17.

The statistical significance of D_p and PGD, and in some cases T_p for fling pulses, implies that ground motion records with the same $SA(T_1)$ values for a given period T_1 may impose different demands on the structure depending on their associated fling parameters, which is not unexpected because the parameter $SA(T_1)$ alone fails to represent the full intensity characteristics of a ground motion record. Therefore, the preceding analysis does not answer the question of whether the presence of larger fling step correlates with other intensity measures that may already be considered in the process of selecting ground motion records, such as the full spectral shapes or cumulative energy measures. To address this question, we turn to non-parametric regression which can handle highly nonlinear relationships and accommodate several ground motion intensity measures as predictors of the structural demands.

To test whether the structural demands associated with ground motions with similar spectral shapes are sensitive to the fling parameters, we use the non-parametric tree-based ensemble learning RF regression and hypothesis test procedure described in section 3. We conduct the hypothesis testing procedure for RF models in which the predicted variable is one of the scalar EDPs used in this study, and the predictive features consist of a single or a combination of IMs in addition to one of the fling parameters of the underlying records (D_p, T_p or the PGD). The significance of the fling parameter as a predictor variable is tested in each case; all the models we tested are listed in Table 6 along with the model shorthand. The dataset of near-fault ground motions (which is defined in this study to be within 20 km JB distance) is used in the analysis, and consists of 51,066 data points per structure. 70% of the data was used for training, and the remaining 30% for testing (evaluating the MSE). Because of the inherit randomness involved in this testing process, we repeated the RF prediction and testing procedure ten times and obtained a mean p-value, where the training and testing data points were selected randomly from the dataset. While each application of the model and testing procedure yielded slightly different p-values, the statistical significance trends remained generally stable. We used the R package RFtest to conduct the permutation and hypothesis testing with the following parameters which were selected to balance accuracy and computational effort, while adhering to the recommendations in (Coleman et al., 2022): the number of RF trees in each model was selected to be between 100 and 200; the size of the sub-samples used for the training is equal to $n^{0.5}$ where n is the number of observations in the training dataset.

Figure 18 summarizes the models we tested in this study for predicting the MIDR and RIDR,

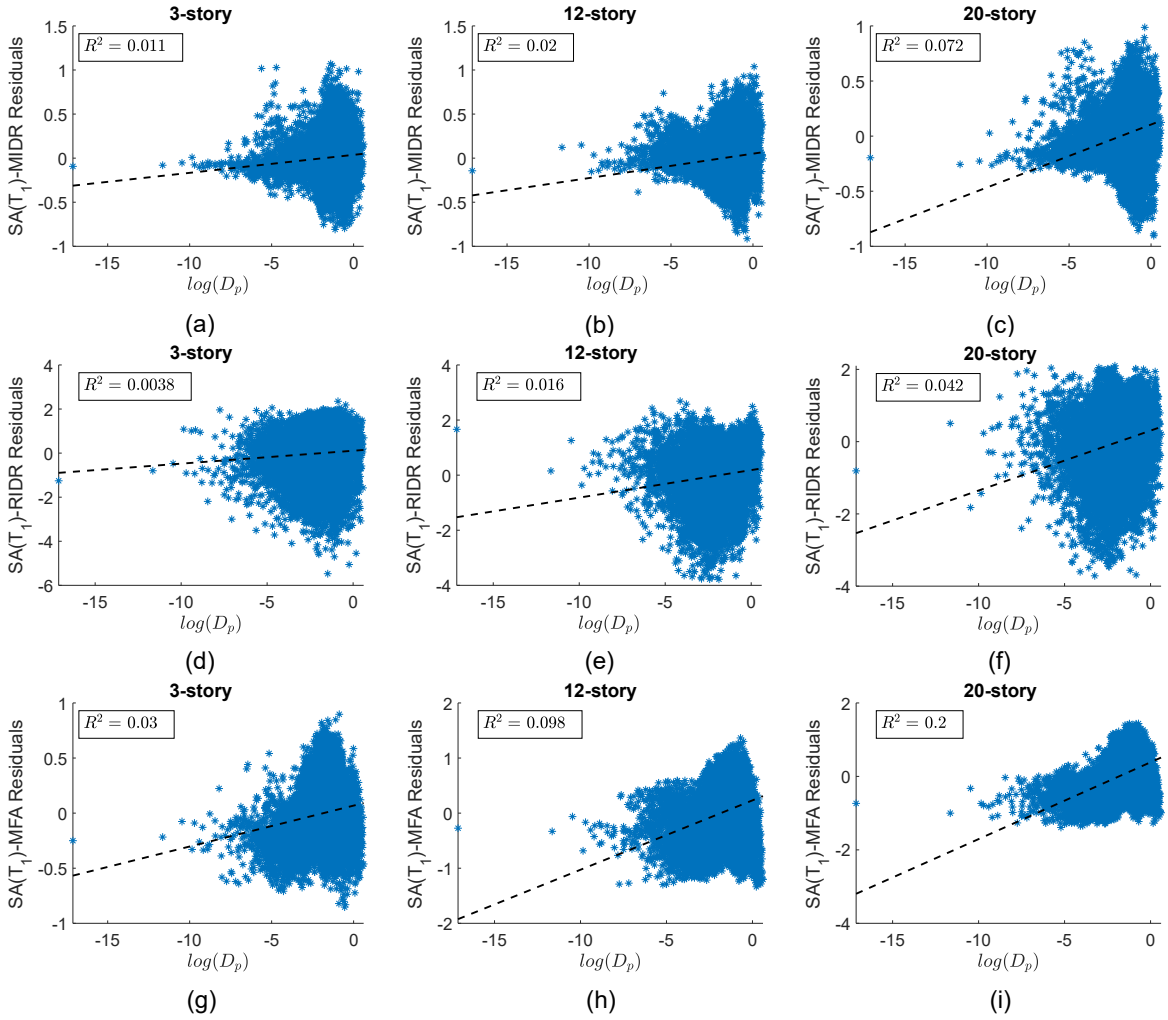

Figure 16: Dependence of the residuals of the $ln(SA(T_1)) - ln(EDP)$ linear regression model on the natural logarithm of the magnitude of the permanent offset associated with near-fault SP ground motions: (a), (b) and (c) for the MIDR demands on the 3-story, 12-story and 20-story buildings, respectively; (d), (e) and (f) for the RIDR demands on the 3-story, 12-story and 20-story buildings, respectively; (g), (h) and (i) for the MFA demands on the 3-story, 12-story and 20-story buildings, respectively.

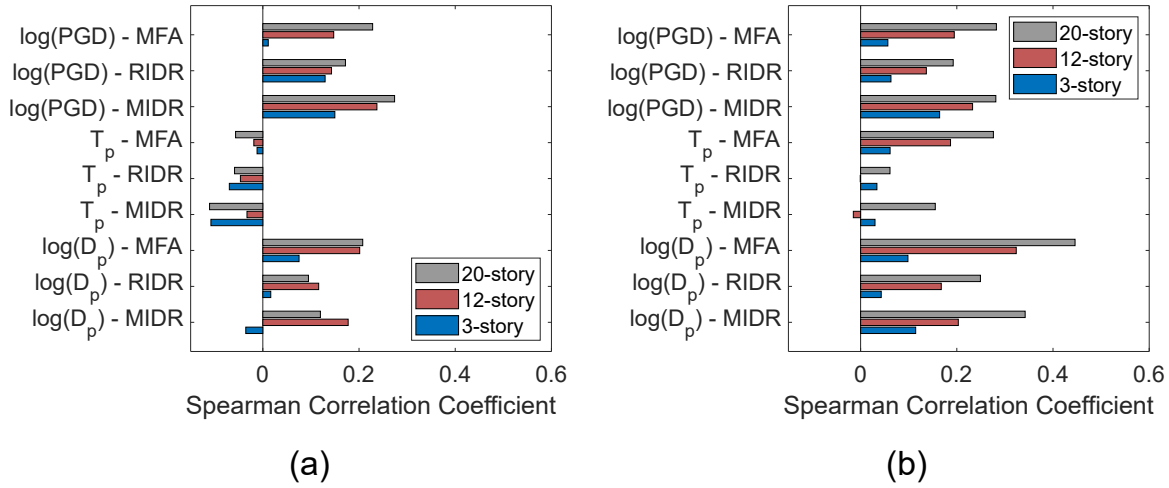

Figure 17: Spearman correlation coefficient between the residuals of the EDP-$SA(T_1)$ model and the fling parameters for different datasets: (a) All near-fault ground motions; (b) near-fault SP ground motions. Both the specific EDP (MIDR, RIDR, or MFA) and fling parameter (D_p, T_p, or PGD) are denoted in the plots.

and the p-values associated with the significance of either D_p, T_p or PGD in each model, and for all three buildings. The p-values represent the probability that adding the fling parameter in question to the predictive model does not significantly reduce the error associated with the predicted EDP. The results summarized in the figure agree with the linear regression results which considered only $SA(T_1)$ as a predictor variable. All subplots of Figure 18 affirm that using a predictive model consisting of only $SA(T_1)$ to predict MIDR or RIDR is not sufficient with respect to any of the three fling parameters. A similar finding is observed when the predictive model consists of both $SA(T_1)$ and PGV, implying that ground motion records with similar $SA(T_1)$ and PGV may impose significantly different demands on structures depending on their fling characteristics.

In contrast, when spectral accelerations at several periods are used as predictive variables, the importance of the fling parameters in predicting MIDR or RIDR decreases substantially, especially for stiffer buildings. For example, subplots a and b demonstrate that neither D_p, T_p nor PGD add any significant predictive power to the model employing the full spectral shape and predicting the MIDR and RIDR demands on the 3-story building. However, for the more flexible 12-story and 20-story buildings (subplots c through f), the dependence of the predicted demands on the fling parameters remained significant in some cases (p-values less than 0.05). Incorporating additional predictive variables in the model representing other aspects of the ground motion intensity tended to further decrease the dependence on the fling parameters. For example, adding either the PGV or the cumulative absolute velocity (CAV) as predictive variables (in addition to the full spectral shape) reduced the dependence of the predicted MIDR on the associated D_p for both the 12-story and 20-story buildings. It is noted, however, that the MIDR demands on the 12-story building remained sensitive to the fling pulse period, regardless of the predictive model. Similarly, the data provides evidence that the RIDR demands on the 20-story building are significantly influenced by

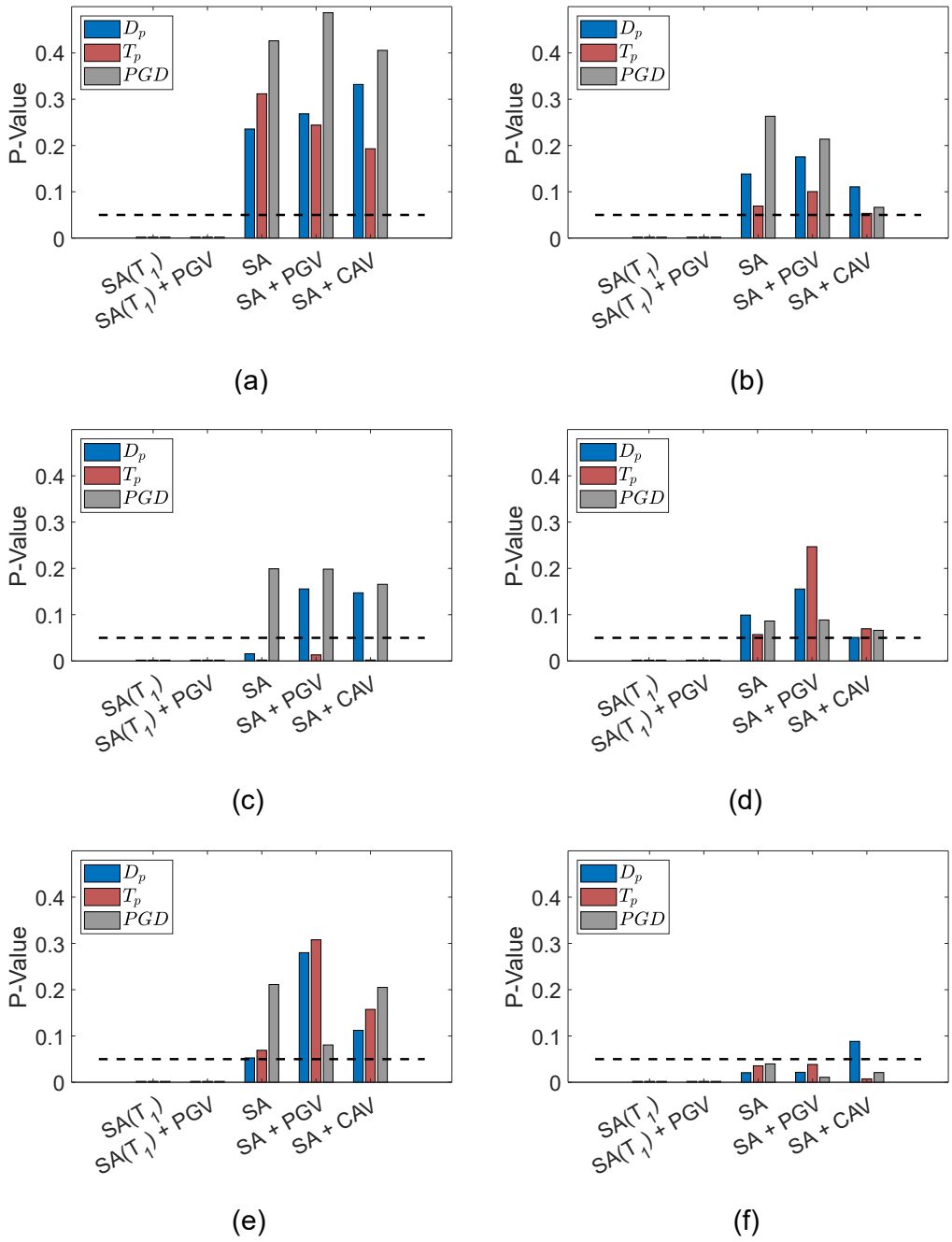

Figure 18: P-values associated with random forest regression models containing the IM combinations indicated on the x-axis, in addition to one of the fling parameters indicated in the legend: (a) and (b) MIDR and RIDR demands, respectively, on the 3-story buildings; (c) and (d) MIDR and RIDR demands, respectively, on the 12-story buildings; (e) and (f) MIDR and RIDR demands, respectively, on the 20-story buildings.

Table 6: Predictor variables in random forest regression models for the MIDR and RIDR demands

Model Shorthand	Predictor Variables
$SA(T_1)$	$ln(SA(T_1))$[1], fling parameter[2]
$SA(T_1) + PGV$	$ln(SA(T_1))$, $ln(PGV)$, fling parameter
SA	$ln(SA(T_i))$ at T_i = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0 and 6.0 s, fling parameter
$SA + PGV$	$ln(SA(T_i))$ at T_i = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0 and 6.0 s, $ln(PGV)$, fling parameter
$SA + CAV$	$ln(SA(T_i))$ at T_i = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0 and 6.0 s, $ln(CAV)$, fling parameter

[1] T_1 is the first mode period of the building under consideration.
[2] The fling parameter is either $ln(D_p)$, T_p or $ln(PGD)$.

any of the fling parameters, regardless of the predictive model (Figure 18f). Overall, the analysis suggests that selecting ground motion records based on their spectral shape and possibly additional intensity measures may substantially reduce the dependence of the predicted structural demands on the associated fling parameters in most cases, but that the fling parameters could remain important for long-period buildings. In the following section, we further interrogate these observations in the context of ground motion record selection.

4.5 Consideration of fling characteristics in ground motion record selection

4.5.1 Selection methodology

We examine whether neglecting the effects of fling in ground motion record selection introduces additional bias in the predicted distributions of structural demands for locations which may experience strong fling effects. We used the GCIM approach (Bradley, 2010, 2012) to select records from the database of simulated ground motions. The GCIM framework allows for selecting ground motion records such that the mismatch between ensembles of records and target distributions of multiple intensity measures is minimized. We use scenario-based ground motion selection (Tarbali and Bradley, 2015), such that the target intensity is IM distributions obtained directly from multiple fault rupture realizations of a single earthquake scenario and for a specified target rupture distance. We leverage the large database of ground motion simulations 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 distributions, and the final set of selected records. We design hypothetical target IM distributions consisting of only ground motions which has moderate or substantial fling displacements. Ground motion records with moderate and substantial fling are defined as those exceeding a permanent static offset amplitude of 0.5 m and 1.0 m, respectively. The differences between the structural demands imposed by selected ensembles of records to match the target IM distribution consisting of fling-dominated records are examined in three cases:

1. when the selection criteria are based solely on the full spectral shape without consideration

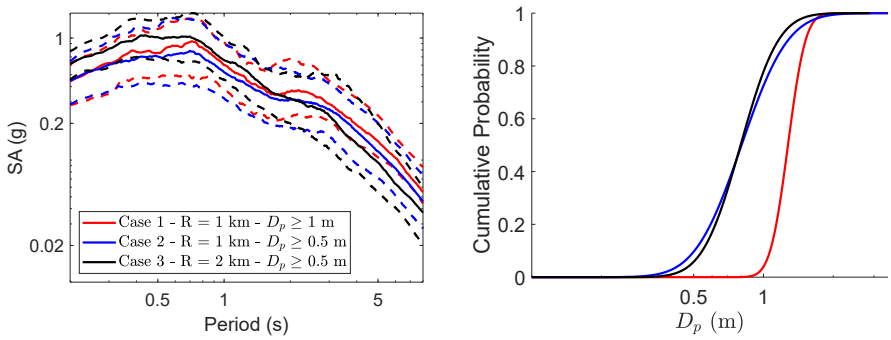

Figure 19: Target ground motion intensity for the three cases indicated in the legend: (a) spectral accelerations; (b) fitted distribution to the magnitude of the permanent offset in the records (intentionally enforced to be larger than either 1.0 m (case 1) or 0.5 m (cases 2 and 3).

of any fling parameters or additional intensity measures; this is denoted as the reference case

2. when the selection criteria explicitly consider one of the fling parameters: the fling amplitude or the fling pulse duration, or the peak ground displacement, or

3. when the selection criteria incorporate intensity measures that represent the cumulative energy or duration of the records, beyond the spectral shape of records. The following additional IMs are specifically considered: PGV, CAV or CAV and the significant duration measure D_{s595}.

Table 7: Characteristics of the target seismic hazard represented in the ground motion selection experiments

Selection Case	Target R (km)	Target D_p (m)
1	1.0	≥ 1.0
2	1.0	≥ 0.5
3	2.0	≥ 0.5

Details of the GCIM are described elsewhere (Bradley, 2010, 2012; Tarbali and Bradley, 2015), and the specific components of ground motion selection in this study were summarized in section 3. A target distribution of an IM at a distance R is obtained by fitting a lognormal distribution to all the simulated records at distance R which follow the particular target conditions ($D_p > 0.5$ m or $D_p > 1.0$ m). The three different target hazards we considered are listed in Table 7, and the acceleration spectra associated with the three target cases are shown in Figure 19a. The figure highlights the different features of the target spectral shape in each case. Specifically, the target spectrum for case1, which is defined by R = 1 km and large fling displacements, has relatively

larger spectral accelerations for periods longer than approximately two seconds, and lower spectral accelerations for periods shorter than two seconds, compared with the target spectrum for case 3, which is defined by R = 2 km and moderate fling displacements. The spectral shape for case 2 is similar to case 1, but with lower accelerations across most periods. Subplot b shows the distribution of D_p corresponding to each case.

At a minimum, we use the acceleration response spectra as the target IMs, and assign equal weights w_i to each spectral ordinate. In cases where additional intensity measures are used, we use a normalized weight of 0.7 for the SA ordinates following the recommendations of Tarbali and Bradley (2015), with an equal marginal weight assigned to each ordinate. The IM combinations and corresponding weights are listed in Table 8.

Table 8: IM combinations tested in the ground motion record selection experiments

Model Shorthand	SA Vector[a]	CAV	D_{s595}	D_p	PGD	PGV	T_p
SA	1.0 [b]	0.0	0.0	0.0	0.0	0.0	0.0
$SA + CAV$	0.7[b]	0.3	0.0	0.0	0.0	0.0	0.0
$SA + CAV + D_{s595}$	0.7	0.1	0.2	0.0	0.0	0.0	0.0
$SA + D_p$	0.7	0.0	0.0	0.3	0.0	0.0	0.0
$SA + PGD$	0.7	0.0	0.0	0.0	0.3	0.0	0.0
$SA + PGV$	0.7	0.0	0.0	0.0	0.0	0.3	0.0
$SA + T_p$	0.7	0.0	0.0	0.0	0.0	0.0	0.3

[a] the SA vector is the spectral acceleration at 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, 5 and 6 seconds

[b] Each spectral acceleration ordinate in the SA vector is assigned an equal normalized weight

4.5.2 Ground motion selection results

Figure 20 summarizes the K-S test statistic of the predicted distributions of MIDR and MFA associated with selection case 1 in Table 7 for a 3-story, 12-story and 20-story building at a distance of 1 km, and a target hazard corresponding to fling displacement amplitude larger than 1.0 m. The analogous results for selection case 3 are shown in Figure 21, which correspond to a target distance of 2 km, and a target fling amplitude larger than 0.5 meters. The K-S test statistic represents the degree of bias associated with the predicted EDP by the selected ensemble of records compared to the target EDP distribution. Each selection experiment is denoted by the target IMs used in the selection process, as indicated on the x-axis. Each box plot summarizes the results of 50 selection experiments using the same selection criteria. To determine whether the selection criteria yield statistically acceptable bias, we compare the bias box plots against the critical value of the K-S test statistic at the 5% significance level, also indicated in the figure with a dashed line. In addition, to assess how the additional IMs influence the bias, we compare all the selection cases against the

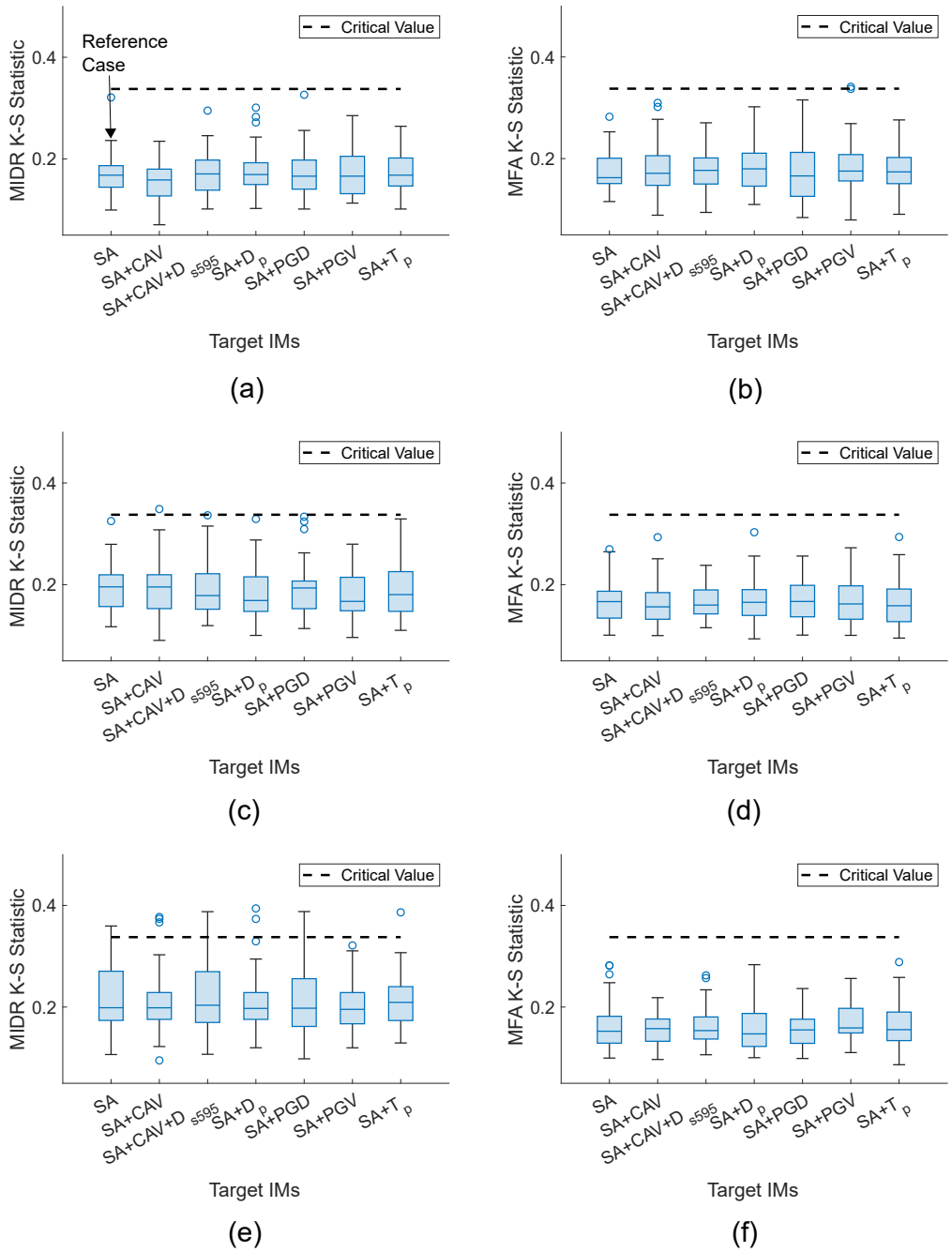

Figure 20: Values of the K-S test statistic of the distributions of EDPs imposed by earthquake records selected to represent the target hazard corresponding to case 1 for the following buildings and EDPs: (a) MIDR demands on the 3-story building; (b) MFA demands on the 3-story building; (c) MIDR demands on the 12-story building; (d) MFA demands on the 12-story building; (e) MIDR demands on the 20-story building; (f) MFA demands on the 20-story building. Each box plot represents the statistics of 50 replicates. The number of selected records in each case is 15.

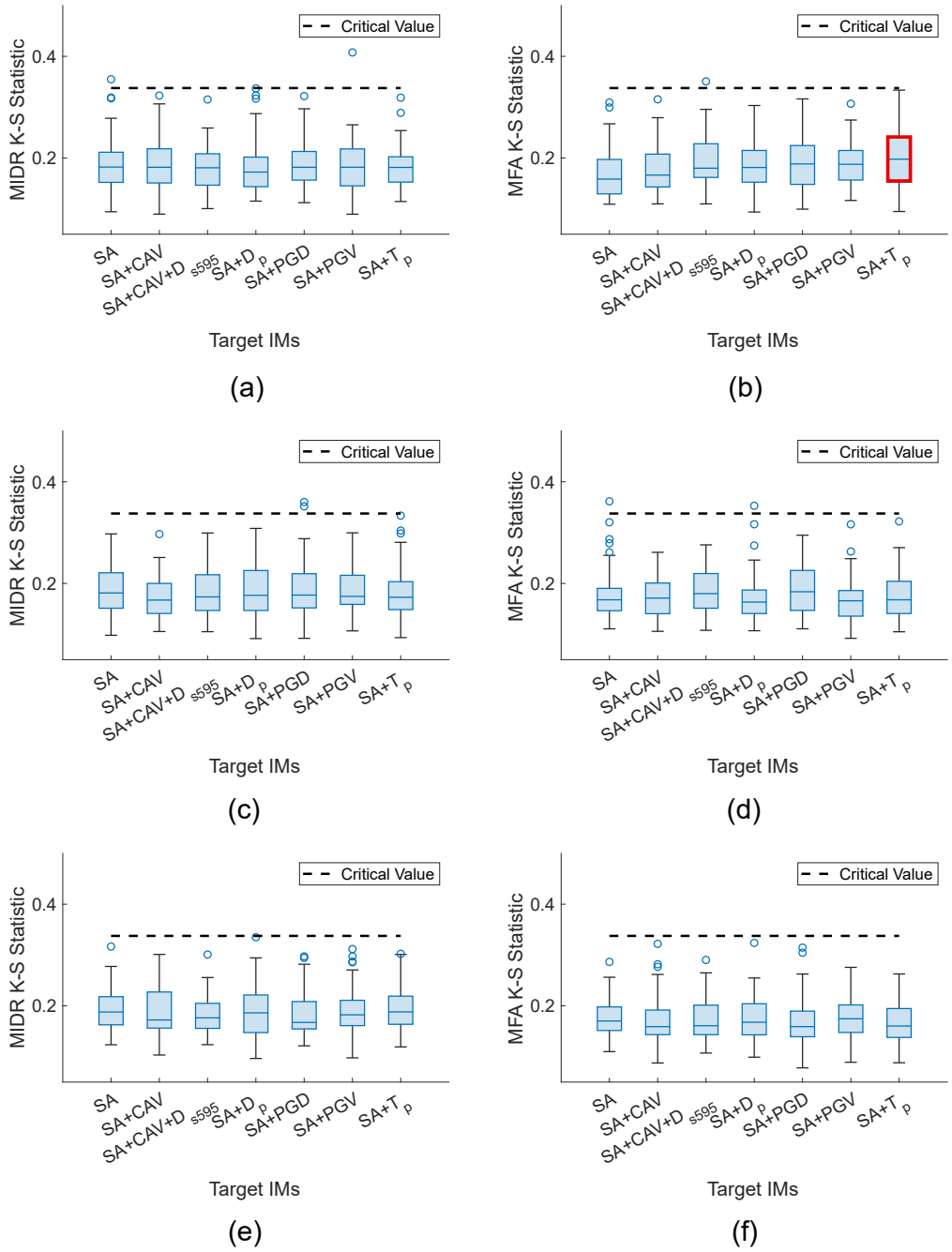

Figure 21: Values of the K-S test statistic of the distributions of EDPs imposed by earthquake records selected to represent the target hazard corresponding to case 3 for the following buildings and EDPs: (a) MIDR demands on the 3-story building; (b) MFA demands on the 3-story building; (c) MIDR demands on the 12-story building; (d) MFA demands on the 12-story building; (e) MIDR demands on the 20-story building; (f) MFA demands on the 20-story building. Each box plot represents the statistics of 50 replicates. The number of selected records in each case is 15. Distributions that are significantly different from the reference case are marked with a red box.

reference case in which only the spectral shape of the records is used as the target IM. This reference case is indicated Figure 20a. We used the non-parametric Kruskal–Wallis one-way analysis of variance with the Dunnett post-hoc multiple comparison procedure to test whether the median bias associated with each IM combination (corresponding to the box plots of Figures 20 and 21) is significantly different from the median bias associated with the reference case at a significant level of 0.05. The single case which was found to have a significantly different median bias was marked in the figure with a red box.

For all selection cases, our results reveal that selecting ensembles of ground motions based on matching only the spectral shape of the target hazard resulted in statistically acceptable bias in all the structural demands, even though the fling characteristics were not explicitly considered in the selection criteria. In addition, no significant reduction in the bias was achieved when we explicitly matched the target distribution of D_p, T_p or PGD for any of the three buildings. Within the critical limits of the adopted K-S test and significance level, this finding implies that the fling characteristics of ground motions may not significantly influence the predicted structural demands, if the ground motion records were selected based on their spectral shape features, and the candidate pool contained records with the appropriate spectral shapes. The results of two selection experiments representative of case 1 for a 20-story building are shown in Figures 22 through 24, and correspond to the following selection criteria: (a) the target hazard consists of the full spectral shape (SA), and (b) the target hazard consists of the full spectral shape in addition to the target distribution of D_p (SA + D_p). Both experiments approximately represent the median bias values for their respective IM combinations in Figure 20e and f.

The top subplots of Figure 22 show that the median, 16th percentile and 84th percentile spectral shapes of the selected ensembles of records in both experiments are very similar. Examining the distributions of the fling displacement D_p associated with the selected and target records in the bottom subplots reveals that statistically acceptable D_p values were obtained in both experiments, despite the fact that the D_p distribution was not explicitly matched in experiment a. Importantly, both experiments resulted in acceptable bias in the predicted structural demands, based on the K-S test at a significance level of 0.05, as demonstrated in Figures 23 and 24 which show the target distributions of the MIDR, RIDR and MFA, and the distributions corresponding to the selected ensemble of records in experiments a and b. In addition, the median, 5th percentile and 95th percentile interstory drift (IDR) envelopes predicted along the height of the building are shown in the bottom subplots of Figure 23 and are qualitatively very similar.

Finally, using more comprehensive criteria for ground motion record selection, namely the IM combinations listed in Table 8 based on the SA vector in addition to PGV, CAV, or CAV and D_{s595}, performed comparably to the SA vector alone for all selection cases, without a significant increase or reduction in the bias of the predicted structural demands. These observations suggest that selecting ground motion records based on matching the spectral shape alone may be sufficient with respect to representing the fling characteristics of the target hazard. However, it is important to note that this finding corresponds to selection experiments where the underlying candidate pool of records contained ground motions with the appropriate shapes and fling characteristics, and that no amplitude scaling of the records was performed. Future investigations will examine ground motion selection cases where the underlying database does not contain records with substantial

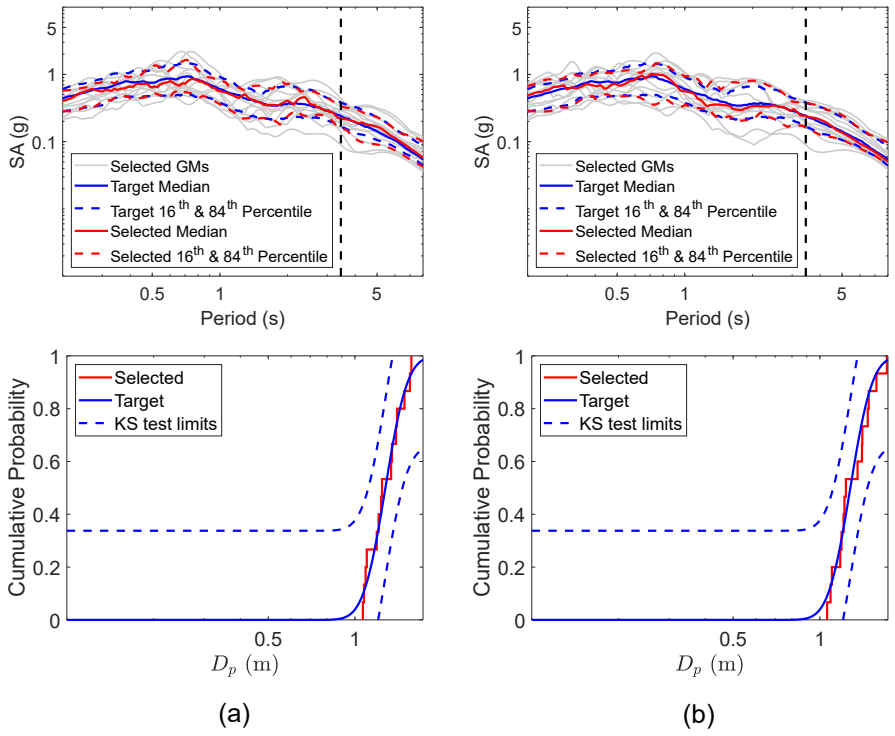

Figure 22: Acceleration response spectra (top) and distributions of the magnitude of fling displacement (bottom) of the target and selected ground motion records corresponding to the following experiments: (a) using the SA vector as the target intensity; (b) using the SA vector and D_p as the target intensity. All cases correspond to a 20-story building at a distance of 1 km with a target fling step larger than 1.0 m. 15 ground motion records were selected in each case.

fling displacements, or where the characteristics of the raw records are altered by scaling.

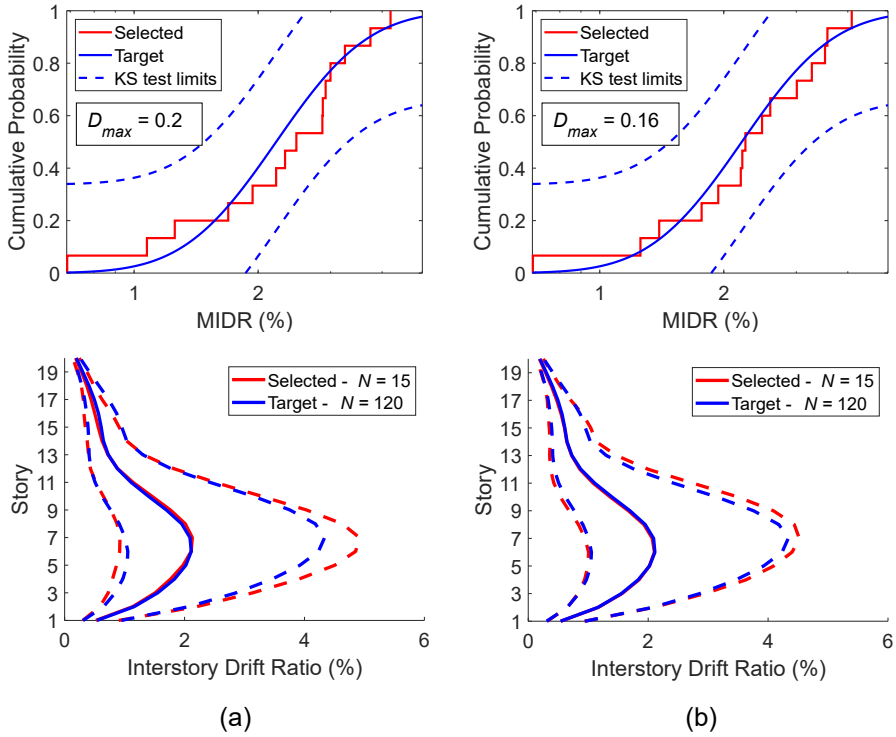

Figure 23: MIDR distributions (top) and the median, and 5th and 95th percentile IDR envelopes (bottom) corresponding to the target and selected ground motion records in the following experiments: (a) using the SA vector as the target intensity; (b) using the SA vector and D_p as the target intensity. All cases correspond to a 20-story building at a distance of 1 km with a target fling step larger than 1.0 m. 15 ground motion records were selected in each case.

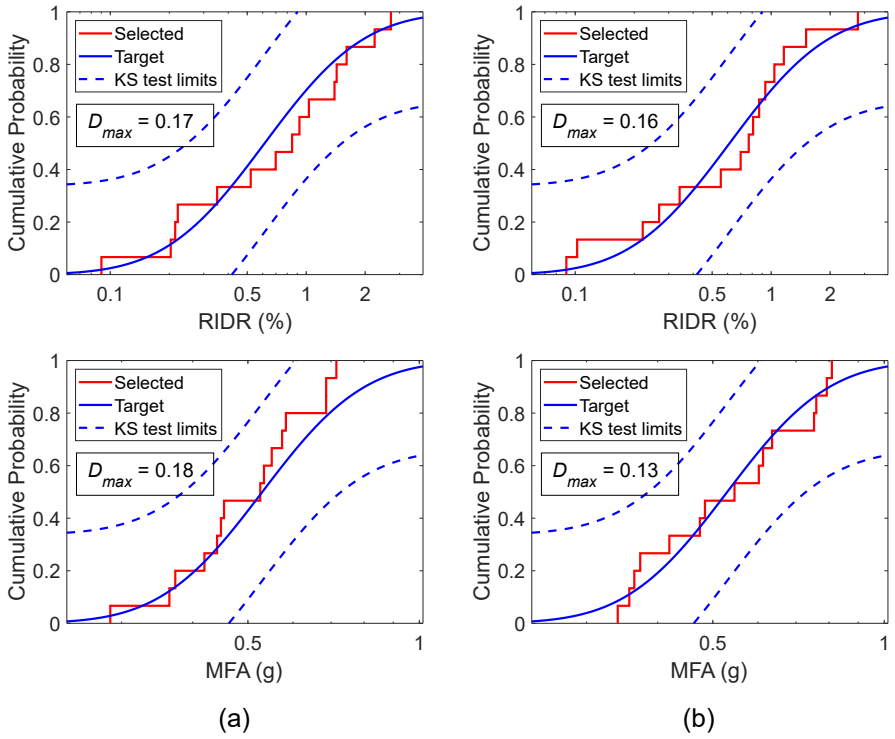

Figure 24: RIDR (top) and MFA (bottom) distributions corresponding to the target and selected ground motion records in the following experiments: (a) using the SA vector as the target intensity; (b) using the SA vector and D_p as the target intensity. All cases correspond to a 20-story building at a distance of 1 km with a target fling step larger than 1.0 m. 15 ground motion records were selected in each case.

5 Use of Simulated Earthquake Ground Motions in Scenario-Based Record Selection

Three-dimensional physics-based simulations have evolved rapidly in recent years and have been used in studying the response of structures to some earthquake scenarios, but they remain relatively underutilized in earthquake engineering applications. Such simulations have the potential to produce synthetic earthquake records that supplement the available database of observational records, especially for relatively large events and short distances. In addition, as such simulations evolve in tandem with scientific understanding of earthquake rupture processes and high-performance computing tools, they may be used to improve the representation of the target seismic hazard for near-fault locations which may be underrepresented in empirical ground motion models. We assess the utility of the database of simulated earthquake ground motions in both aforementioned circumstances. We explore the use of the simulated records in representing the characteristics of near-fault ground shaking for the analysis of building structures, and examine their performance against real earthquake recordings, and the predictions of empirical ground motion models.

5.1 Study design

Suites of ground motion records are selected for the analysis of two near-fault building structures in a scenario-based analysis context. We use a scenario representative of a magnitude 7.0 shallow crustal event, and distance parameters of 1 km and 5 km. The V_{s30} is selected as 380 m/s (which corresponds to the lower limit V_{s30} in the simulated earthquake scenario). The GCIM framework (Bradley, 2010) is used to conduct the ground motion selection experiments. Each selection experiment is defined by the following: (1) the target distance (either 1 or 5 km), (2) the source of the target hazard spectra (either based on available GMPEs, or based on the simulated earthquake scenario for the specified distance and soil properties), (3) the horizontal ground motion component represented by the target spectra (either a SN or SP component), and (4) the database of records from which the suite of records are selected from the analysis (either simulated or real ground motion records). The full study matrix is described in Table 9, where every row corresponds to two different record selection experiments, for a total of 16 experiments. The selection experiments were conducted on the basis of individual ground motion components because we are interested in comparing the targets associated with and the response of structures to either the SN or SP components. We use the selection experiments to assess the performance of the earthquake simulations in the following ways:

- The target hazard spectra for the scenario predicted by the earthquake simulations is compared against the target hazard indicated by the GMPE for the same causal parameters. The median, 16th and 84th percentile spectra corresponding to RotD50 spectra are examined, in addition to two individual perpendicular horizontal components (SN and SP). In addition, the structural demands imposed by selected suites of simulated ground motions based on each target hazard are compared.

- The characteristics and impacts of the near-fault simulated ground motions are holistically

evaluated against those of real near-fault ground motions by comparing the structural demands imposed by suites of simulated and real ground motion records selected to match the same target hazard spectra.

The characteristics of the simulated ground motions and the dataset of recorded near-fault ground motions were described in section 3.

Table 9: Parameters associated with the ground motion selection experiments

Index	Distance (km)	Component	Hazard Target	Selected Records
1, 2	1.0	SN, SP	GMPE	Real
3, 4	1.0	SN, SP	GMPE	Simulated
5, 6	1.0	SN, SP	Simulated	Real
7, 8	1.0	SN, SP	Simulated	Simulated
9, 10	5.0	SN, SP	GMPE	Real
11, 12	5.0	SN, SP	GMPE	Simulated
13, 14	5.0	SN, SP	Simulated	Real
15, 16	5.0	SN, SP	Simulated	Simulated

5.1.1 GMPE target hazard spectra

The target GMPE hazard spectra are based on a weighting of the following empirical models: ASK14 (Abrahamson et al., 2014), BSSA14 (Boore et al., 2014), CB14 (Campbell and Bozorgnia, 2014) and CY14 (Chiou and Youngs, 2014), each receiving a normalized weight of 0.25 in the calculation of the RotD50 spectra. To obtain the SN and SP target spectra from the RotD50 spectra, the approach described in Shahi and Baker (2014b) was implemented, where an empirical model was presented to obtain median horizontal ground motion components in arbitrary orientations from the fault based on the median RotD50 spectrum.

5.1.2 Simulated target spectra

The statistical distributions of the spectral accelerations associated with the simulated target hazard spectra were estimated directly based on the simulated ground motions for the specified magnitude, distance and V_{s30} values. The target distributions should incorporate the anticipated variability due to source, path and site effects. The database of rupture realizations represents some of the aleatory variability associated with the hypocenter location and rupture slip distribution, but is not expected to be comprehensive of all possible conditions. The RotD50 component of the simulated target spectrum was calculated based on the RotD50 definition in Boore (2010).

5.2 Ground motion selection procedure

The selection procedure based on the GCIM framework is followed, as described in section 3. The target hazard consists of 20 spectral acceleration corresponding to the following periods: 0.01,

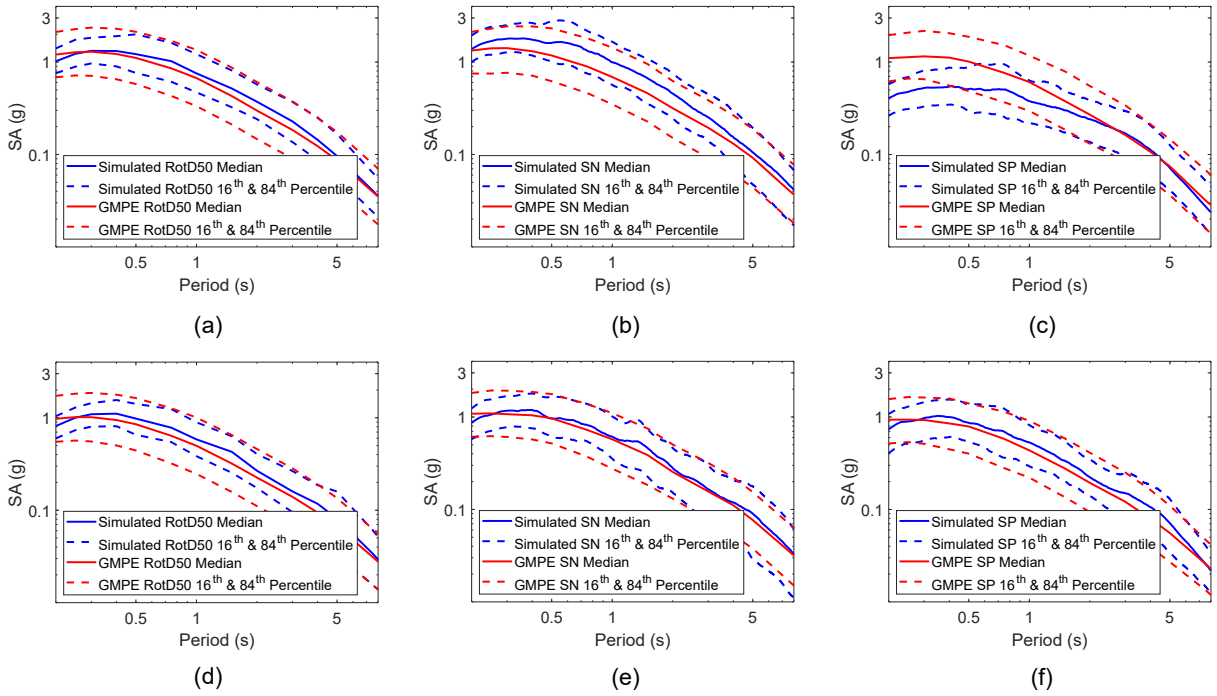

Figure 25: Target hazard spectra based on simulated ground motions and GMPEs: (a) RotD50 component at 1 km; (b) SN component at 1 km; (c) SP component at 1 km; (d) RotD50 component at 5 km; (e) SN component at 5 km; (f) SP component at 5 km

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, 5 and 6 seconds. For the simulated ground motion target, the target IM sample consists of simulated ground motions drawn randomly and independently from the simulated records which constitute the target. Therefore, the correlations between the IMs (SAs at different periods) are inherently incorporated in the selection process). For the GMPE target, random realizations representing the multivariate distribution of the IMs are drawn using the methodology described in Wang (2011), and using the spectral acceleration correlations proposed in Baker and Jayaram (2008). In this study, no scaling of the ground motion records was performed to obtain an appropriate match to the target hazard, because the causal parameters of the database of records were broadly similar to those of the target. However, a more comprehensive study may consider scaling of ground motion records from a broader database, which is a typical practice in earthquake engineering design procedures.

5.3 Results and discussion

5.3.1 Comparison between the simulated and GMPE target hazard spectra

Figure 25a, b and c show the median, 16th percentile and 84th percentile target hazard spectra for a scenario with M = 7.0 and R = 1 km based on the simulated records for this scenario, plotted against the analogous target spectra based the GMPE for the same causal parameters. Subplot a

demonstrates the reasonable agreement between the simulated and GMPE RotD50 target spectra across a broad range of periods. Specifically, we compare the spectra across periods between 0.2 seconds (which is the lower limit represented by the earthquake simulations) and 8.0 seconds. As expected, because the simulated records represent a small number of realizations, they have lower variability than the GMPE target; therefore, the 16th percentile of the simulated target tends to be notably higher than the 16th percentile of the GMPE target across most periods.

The individual horizontal component target simulated spectra, however, differ more substantially from the estimated GMPE target spectra for the SN and SP components. Examination of the SN and SP component targets in Figure 25b and c reveals that the simulated SN component median and 16th percentile spectra are higher than the GMPE spectra in the short-period to mid-period range, whereas the 84th percentile spectrum of both targets appears to be similar across all periods. These trends are consistent with the trends observed when comparing the database of SN simulated near-fault ground motions against the real SN records, as discussed previously. On the other hand, the simulated SP target spectral accelerations are substantially smaller than the GMPE target spectral accelerations for periods up to approximately 2 seconds, as the plots reveals. It's important to note that the SP target spectra trends do not appear to be consistent with the trends observed in comparing the simulated SP near-fault records against the real records for distances between 0 and 15 km, shown in Figure 1, and the distance scaling trends noted previously in section 3. There are several potential reasons for this apparent inconsistency. First, because of the very small number of recorded ground motions for distances less than 5 km, the ground motion intensity at such short distances may be poorly represented by the GMPE model. Second, the particular earthquake scenario simulations used in this study, which represent a pure strike-slip earthquake with a vertical fault, may be characterized by higher polarization in the ground motions at very short distances, leading to higher SN and lower SP accelerations than the general trends observed across the various fault geometries and rupture mechanisms employed in constructing the empirical GMPE. Examining the analogous target spectra for a distance of 5 km in Figure 25d, e and f demonstrates the consistency between the RotD50, SN and SP simulated and GMPE target hazard spectra, especially the median and 84th percentile spectra, but the 16th percentile spectrum of the simulated target remains somewhat higher than the 16th percentile GMPE spectrum.

5.3.2 Structural demands imposed by records matching the GMPE and simulated spectra

The differences between the simulated and GMPE targets are further highlighted in the context of their influence on the demands imposed on building structures with different dynamic characteristics. Figure 26 shows the results of record selection experiments from the database of simulated records, where subplot a shows the results of using the SN simulated target hazard spectra, and subplot b represents the analogous results using the SN GMPE target spectra. The global residual measure R_{IM} is shown in each subplot and is almost identical for both experiments. Figures 27 and 28 show the cumulative distributions of the structural demands corresponding to the selected suites of records in both cases on the 3-story and 20-story buildings, respectively. Empirical cumulative distributions for the MIDR, RIDR and MFA are plotted for each building, along with the fitted lognormal distributions. In addition, the median, 5th percentile and 95th percentile IDR envelopes

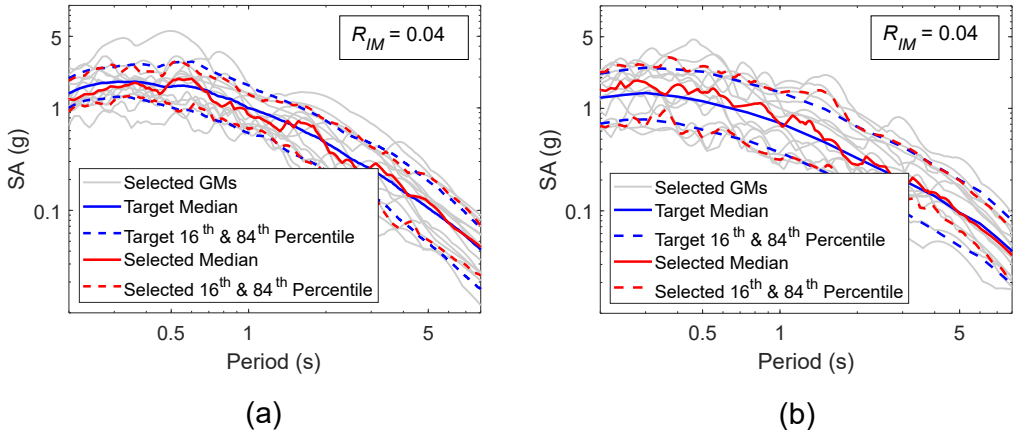

Figure 26: Selected ground motions matching a SN (a) simulated target, and (b) GMPE target for R = 1 km and V_{s30} = 380 m/s. 15 ground motion records were selected in each case.

corresponding to the selected records in each case are shown in subplot d.

As expected based on the trends of the target spectra, the demands imposed by the records selected to match the simulated SN target spectra appear to be higher than those imposed by the records selected to match the target GMPE SN spectra for both buildings. However, the differences between the demands imposed by each suite of records are not significant, based on the Mann-Whitney U (MWU) hypothesis test where we test whether there is enough evidence in the data to reject the null hypothesis that both samples come from distributions with equal medians. The null hypothesis is not rejected in this case at a significance level of 0.05 for MIDR, RIDR and MFA sample pairs for each building (p-values between 0.26 and 0.56 for the 3-story building, and between 0.38 and 0.53 for the 20-story building). The differences observed in Figures 27 and 28 tend to be higher at the tails of the distribution, which correspond to the probabilities of not exceeding smaller demand levels. For example, the probability of not exceeding an MIDR of 1.5% for the 3-story building is predicted by the records matching the simulated target to be only 25%, but is predicted to be over 40% by the records matching the GMPE target. Similarly, for the 20-story building, the probability of not exceeding an MIDR of 1.5% is predicted by the records matching the simulated target to be 40%, but is predicted to be approximately 55% by the records matching the GMPE target.

Substantially larger differences are observed between the structural demands predicted by matching the SP simulated or GMPE target hazard spectra (Figure 29), as expected based on the trends discussed above for the target spectra. This is particularly the case for structural demands controlled by the high-frequency content of the ground motion records, such as the demands on short-period building, and the acceleration-based EDPs (i.e., the MFA). These trends are shown in Figures 30 and 31 for the 3-story and 20-story, respectively. Much lower demands on the 3-story building are predicted by matching a simulated hazard target, compared to a GMPE target. In the case of the 20-story building, the cumulative distributions of the displacement-based MIDR demands predicted in both cases are not significantly different, based on the MWU statistical test

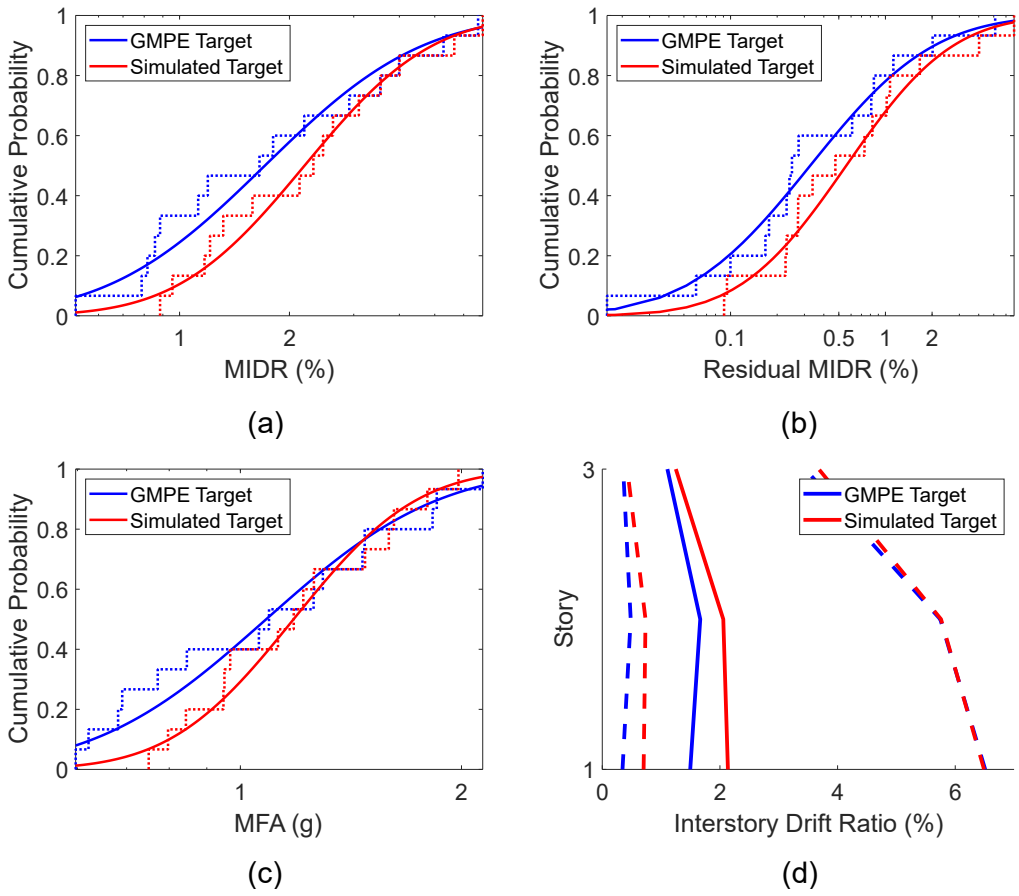

Figure 27: Empirical and fitted cumulative distributions of the demands imposed on the 3-story building by ground motions selected to match GMPE and simulated SN target spectra at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records corresponding to each target.

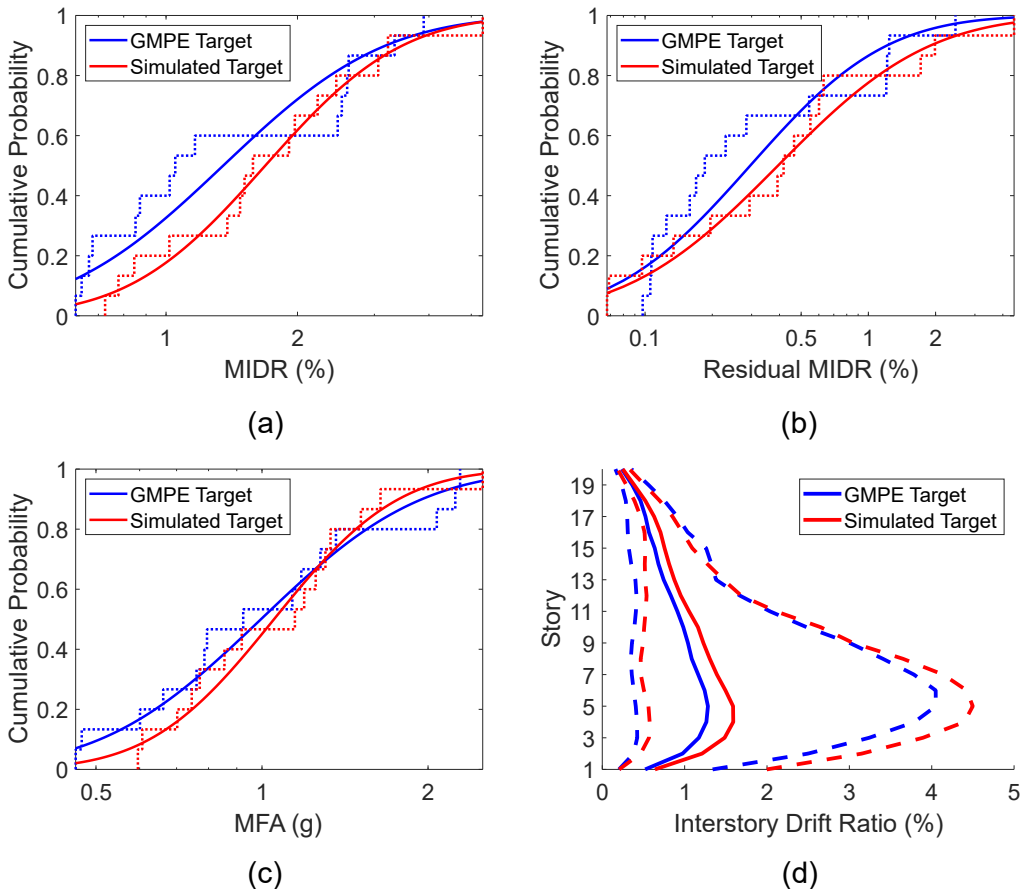

Figure 28: Empirical and fitted cumulative distributions of the demands imposed on the 20-story building by ground motions selected to match GMPE and simulated SN target spectra at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records corresponding to each target.

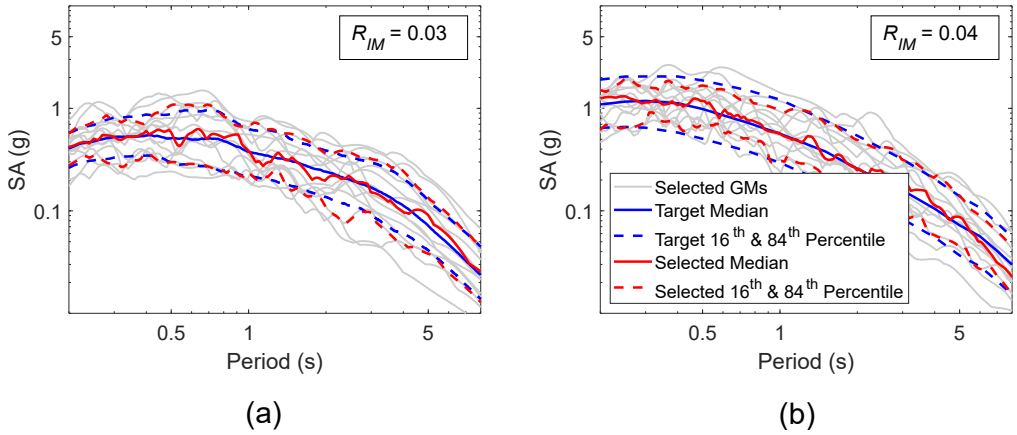

Figure 29: Selected ground motions matching a SP (a) simulated target and (b) GMPE target for R = 1 km and Vs30 = 380 m/s. 15 ground motion records were selected in each case.

(p-value = 0.87). However, the cumulative distributions of the MFA demands are significantly different based on the same test (p-value = 0.0001). For example, using a simulated target leads to a probability of not exceeding an MFA of 0.5 g of approximately 80%, whereas using a GMPE target leads to a probability of only 10%. On the other hand, the probability of not exceeding a MIDR of 1.5% is predicted by matching the GMPE and simulated hazard targets to be 60% and 67%, respectively.

In agreement with the trends observed for the simulated and GMPE target spectra at a distance of 5 km, most of the differences between the demands imposed by suites of records which were selected to match a simulated or GMPE hazard target diminish for both structures at a distance of 5 km. As observed in Figure 32, the cumulative distributions of the MIDR, RIDR and MFA demands on the 3-story building imposed in both cases are not significantly different based on the MWU test (p-values between 0.38 and 0.93). The null hypothesis is also not rejected for the analogous cumulative distributions of the MIDR, RIDR and MFA demands on the 20-story building (shown in Figure 33), indicating no statistically significant difference between using a GMPE and simulated scenario hazard target for this case (MWU test p-values between 0.25 and 0.62).

5.3.3 Characteristics of simulated and real ground motions selected to match the same hazard target

In addition to assessing the potential for using simulated earthquake scenarios to represent the target hazard in seismic risk analysis, we also examine the appropriateness of using databases of simulated earthquake ground motions instead of, or in supplement to, real ground motions in the analysis of structures, after the target hazard parameters have been established. The use of simulated ground motion records is useful in cases where there are no real records in the available observational databases with the desired characteristics (for example, records which correspond to distances less than 10 km and relatively large magnitude events).

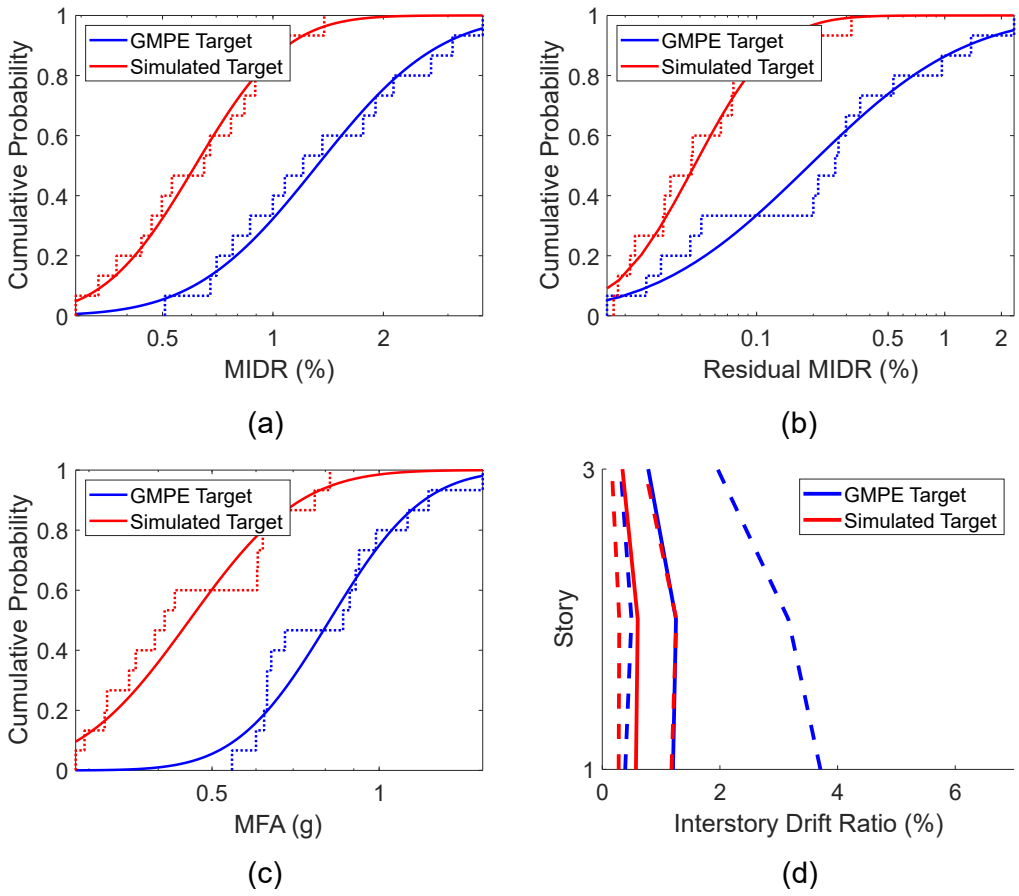

Figure 30: Empirical and fitted cumulative distributions of the demands imposed on the 3-story building by ground motions selected to match GMPE and simulated SP target spectra at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records corresponding to each target.

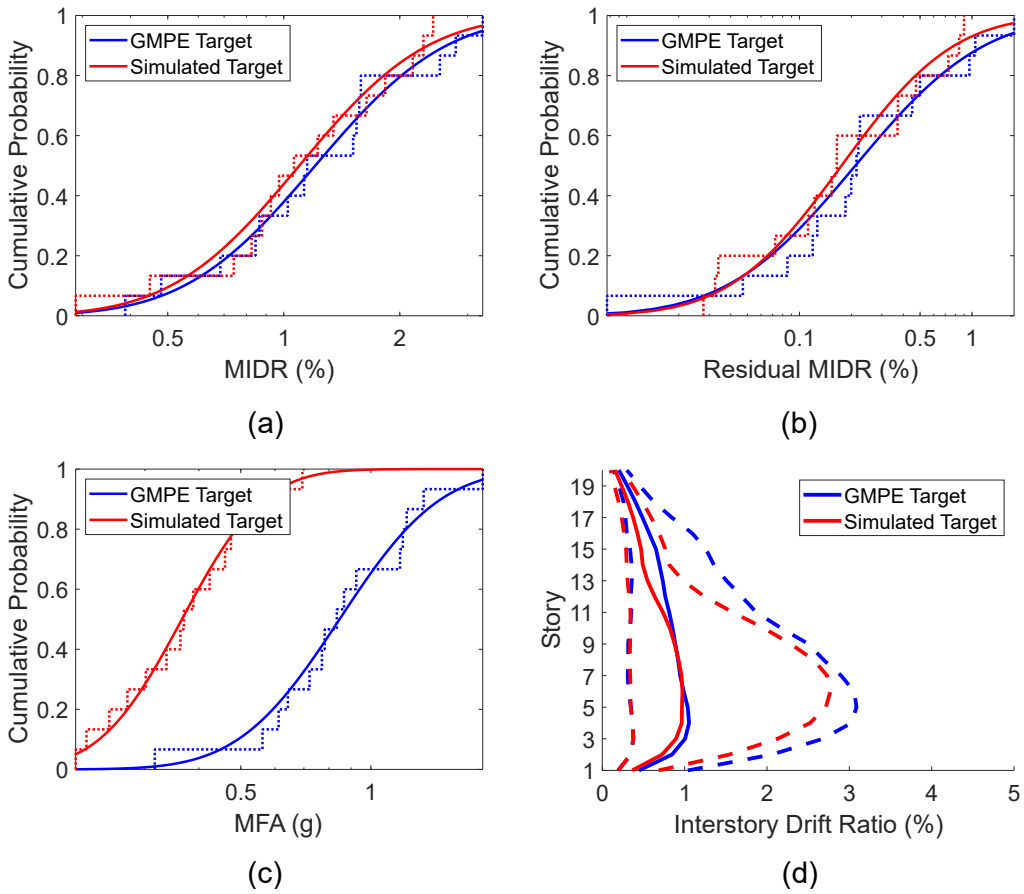

Figure 31: Empirical and fitted cumulative distributions of the demands imposed on the 20-story building by ground motions selected to match GMPE and simulated SP target spectra at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records corresponding to each target.

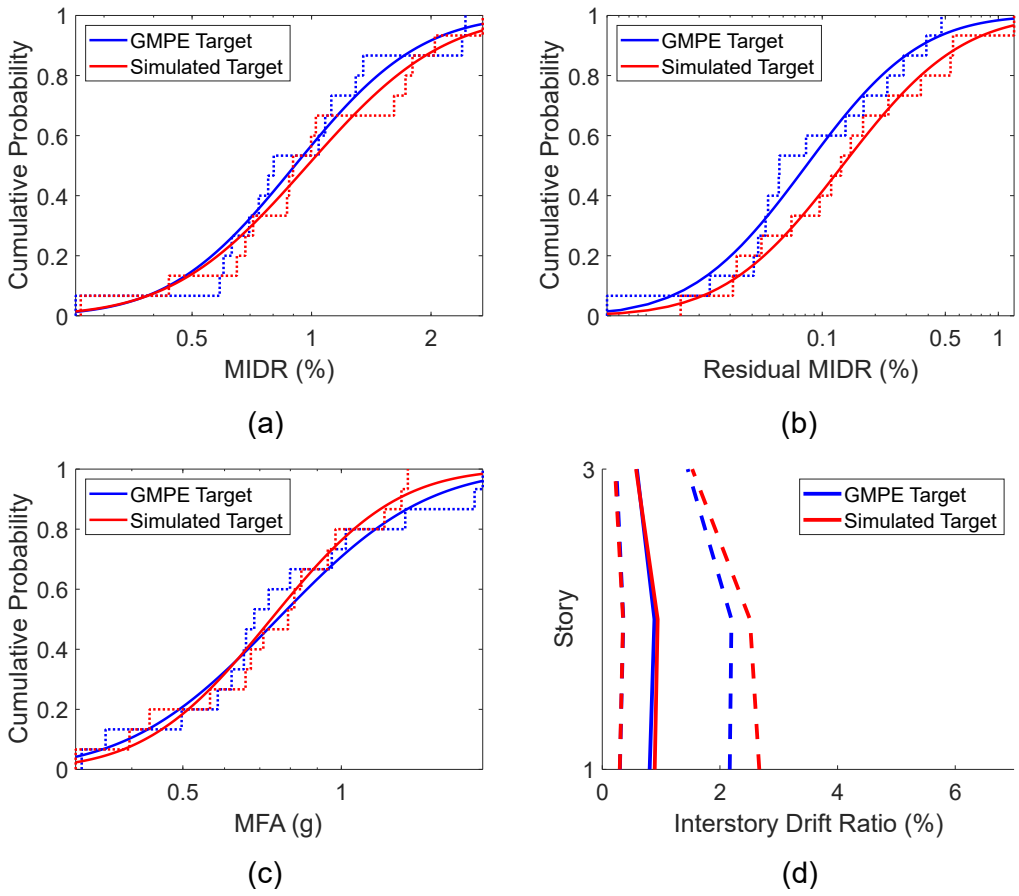

Figure 32: Empirical and fitted cumulative distributions of the demands imposed on the 3-story building by ground motions selected to match GMPE and simulated SP target spectra at R = 5 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records corresponding to each target.

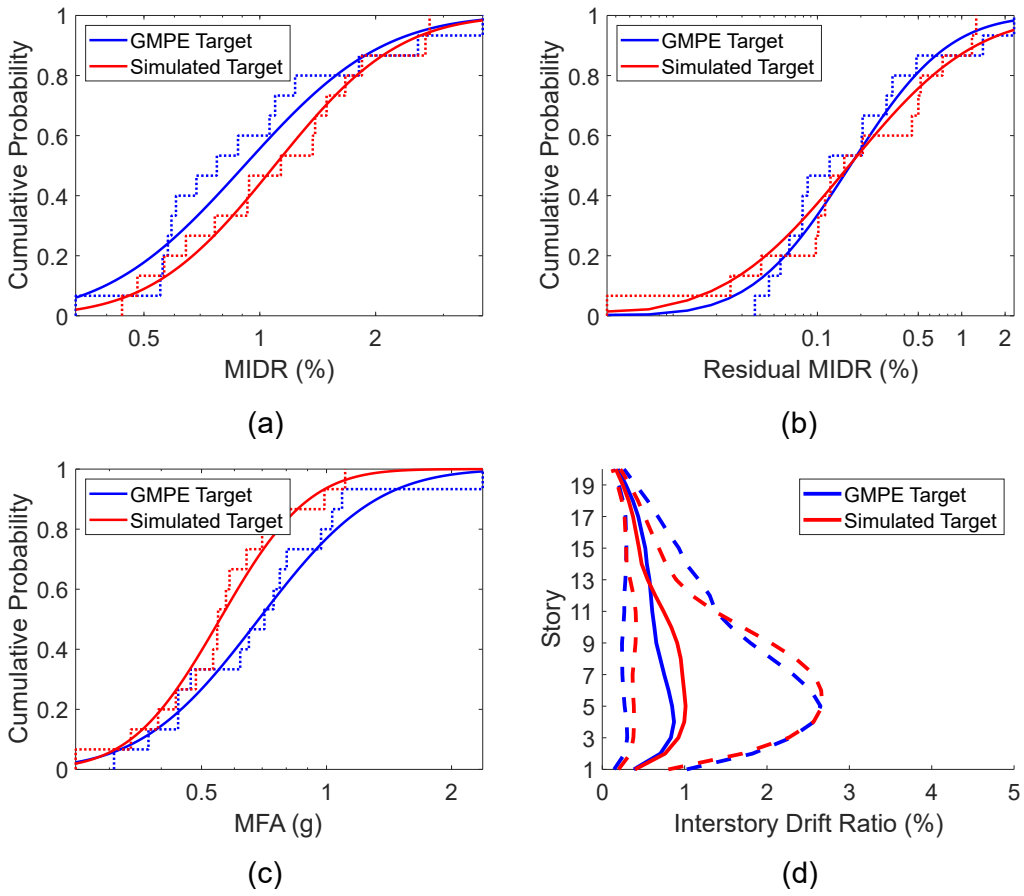

Figure 33: Empirical and fitted cumulative distributions of the demands imposed on the 20-story building by ground motions selected to match GMPE and simulated SP target spectra at R = 5 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records corresponding to each target.

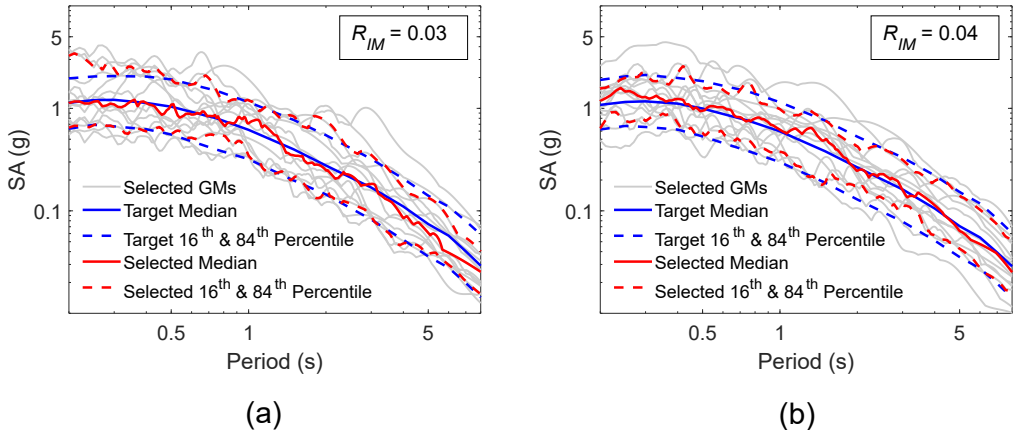

Figure 34: (a) Recorded ground motions and (b) simulated ground motions selected to match a SP GMPE target spectrum for R = 1 km and V_{s30} = 380 m/s.

Figure 34 shows the results of ground motion selection experiments conducted to match a target hazard based on the SP GMPE spectra at a distance of 1 km. Subplot a shows the median, 16th percentile and 84th percentile of the selected records from the database of real near-fault records, whereas subplot b represents records selected from the database of simulated ground motions. It is noted that both selection experiments have a very similar global residual (noted in each subplot); i.e., the distributions associated with both suites of selected ground motions represent those associated with the target hazard spectra to a statistically similar extent. The same observations can be made for the distributions of the selected real and simulated ground motions to match a SN GMPE target, which are shown in Figure 37.

The structural demands corresponding to both suites of selected simulated and real records imposed on the 3-story and 20-story buildings are displayed in Figures 35 and 36, respectively, and demonstrate the consistency between the cumulative demand distributions imposed by simulated and recorded ground motions for the MIDR, RIDR and MFA, with no evidence of a statistically significant difference between each pair of samples (two-sample MWU test, p-values between 0.28 and 1.0). More notable differences are observed between the demands imposed by the suites of selected simulated and real records matching target SN GMPE spectra (compared to the SP target spectra), especially in the cumulative distribution of the displacement-based demands on the 20-story building (Figures 38 and 39). However, the results do not suggest that the distributions are significantly different based on the statistical testing procedure at the 0.05 significance level (p-values between 0.08 and 0.97). Nonetheless, we note that the structural demand trends suggest that the simulated SN ground motions may tend to overestimate the structural demands in the short-period range (e.g., the demands on the 3-story building) and underestimate the demands in the long-period range (e.g., the MIDR demands on the 20-story building) compared to real ground motion records selected to match the same target hazard representing the SN component.

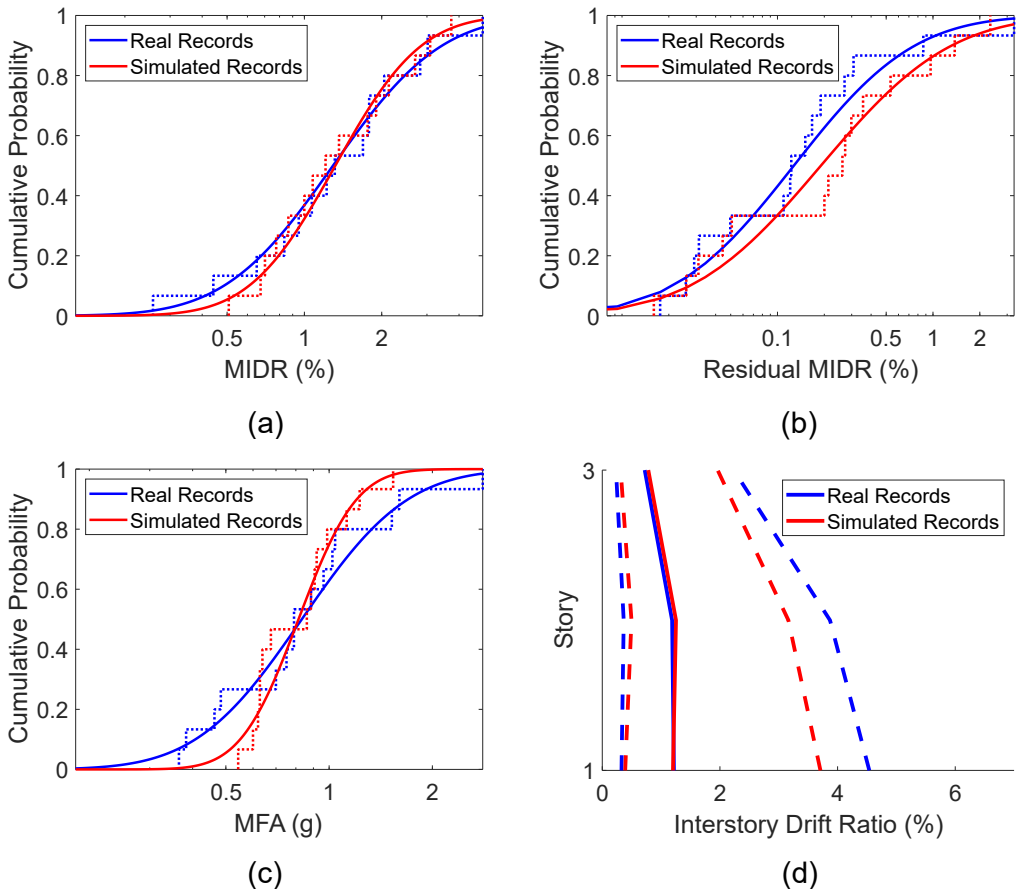

Figure 35: Empirical and fitted cumulative distributions of the demands imposed on the 3-story building by real and simulated ground motions selected to match a GMPE SP target spectrum at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records in each case.

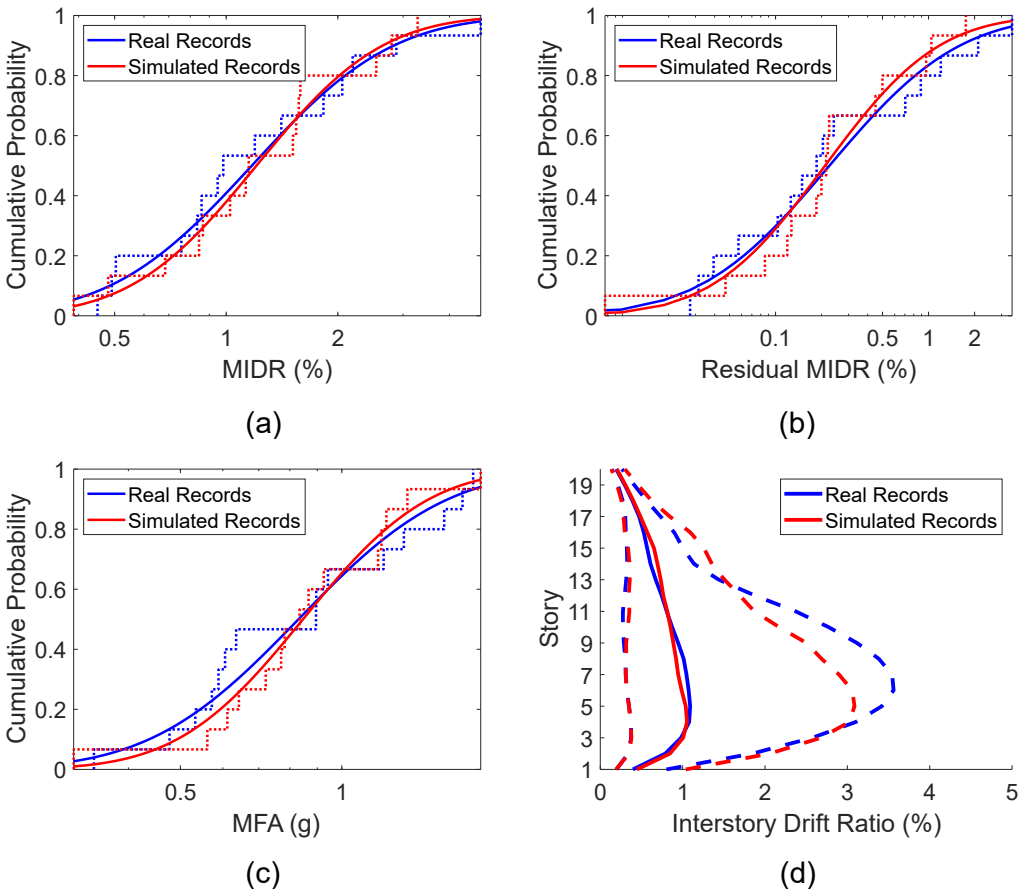

Figure 36: Empirical and fitted cumulative distributions of the demands imposed on the 20-story building by real and simulated ground motions selected to match a GMPE SP target spectrum at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records in each case.

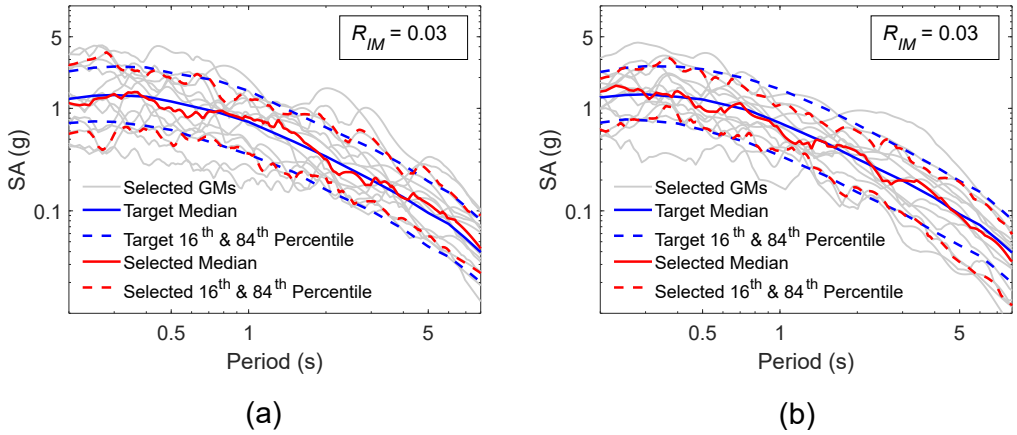

Figure 37: (a) Recorded ground motions and (b) simulated ground motions selected to match a SN GMPE target spectrum for R = 1 km and V_{s30} = 380 m/s.

6 Conclusions

The selection of earthquake records for performing nonlinear analysis of structures located near active faults is a challenging task due to the sparsity of available near-fault field observations, and limited understanding of the unique characteristics of near-fault shaking that impact the response of engineered structures. In this work, a large, spatially uniform, and dense dataset of simulated earthquake records was used to examine the important features of near-fault ground shaking for the analysis of low-, mid-, and high-rise buildings representative of modern seismic code standards. This study systematically examined the dependence of several engineering demand parameters on the directivity and fling pulse features of near-fault simulated earthquake ground motions. Thousands of hypothetical earthquake record selection experiments were performed to evaluate the bias in the predicted structural demands associated with certain selection criteria of interest, such as including and excluding the directivity pulse parameters, the fling displacement parameters, and other relevant IMs. Finally, the intensity characteristics of the simulated earthquake ground motions are holistically compared against those of near-fault real records, and GMPEs representing near-fault hazard spectra, with focus on the bias that could propagate to the predicted structural demands in engineering analysis applications. Based on the findings of this study, the following conclusions and recommendations are made:

- Classification of near-fault records into pulse/non-pulse based on the presence of dominating directivity pulse features was not found to be necessary for selecting near-fault records for engineering analysis, as long as the records were selected based on their full spectral shapes, and the target hazard is representative of near-fault locations. In addition to the shortcomings of this binary approach discussed in Kenawy et al. (2023), the selection experiments performed in this study reveal that there is no statistically significant difference between the structural demands imposed by suites of records selected from pools of candidate records that are exclusively pulse-type, or mixed (contain both pulse and non-pulse

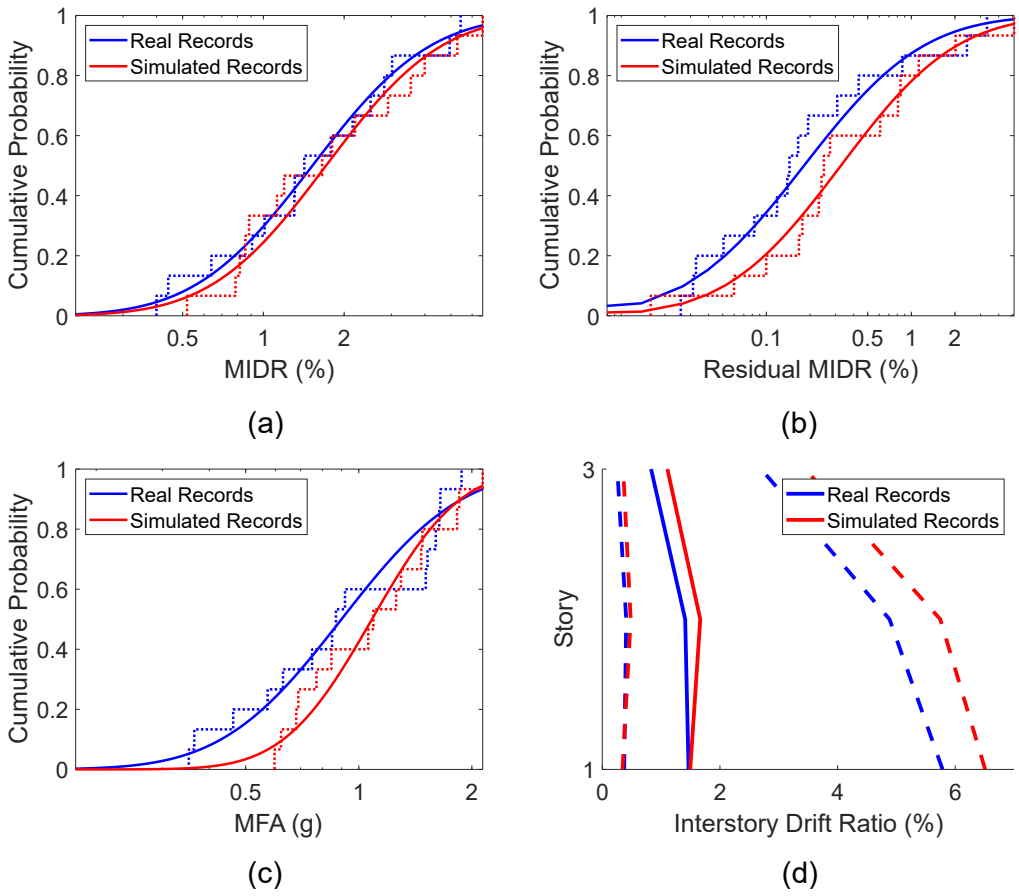

Figure 38: Empirical and fitted cumulative distributions of the demands imposed on the 3-story building by real and simulated ground motions selected to match a GMPE SN target spectrum at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records in each case.

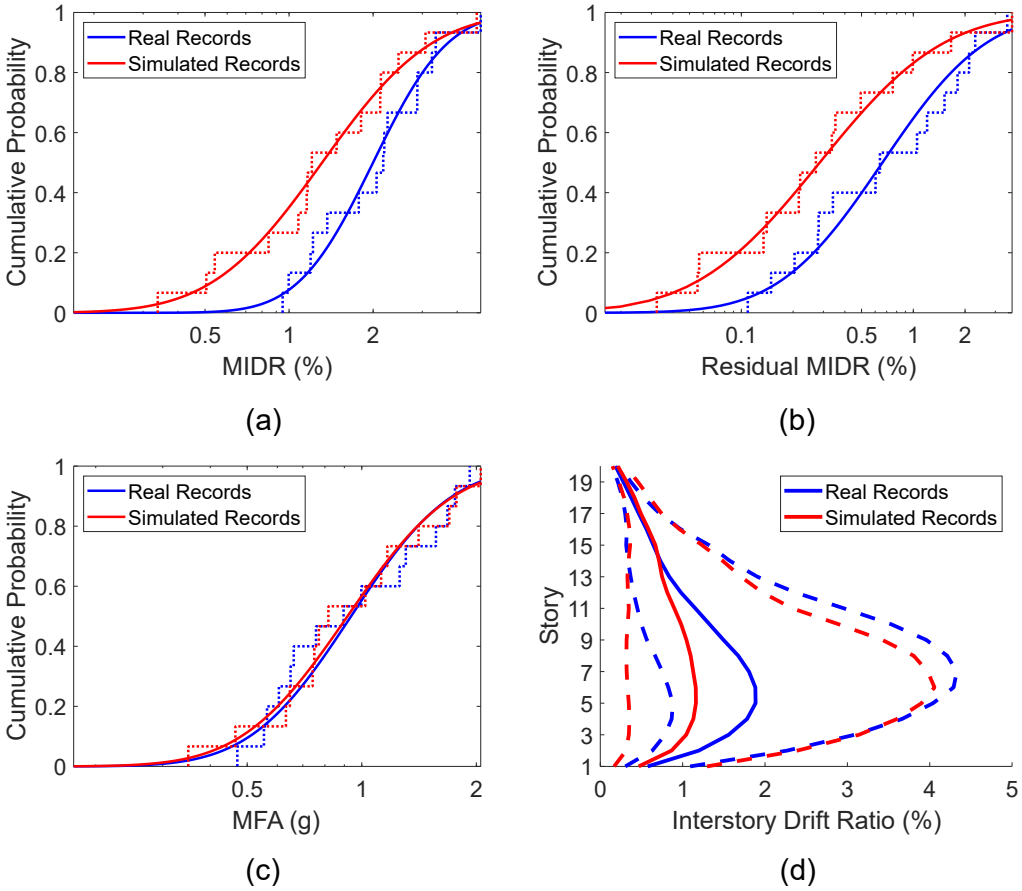

Figure 39: Empirical and fitted cumulative distributions of the demands imposed on the 20-story building by real and simulated ground motions selected to match a GMPE SN target spectrum at R = 1 km: (a) MIDR; (b) RIDR; (c) MFA. Subplot (d) shows the median, and 5th and 95th percentiles of the IDR envelopes imposed by the selected suite of records in each case.

records), even when the target hazard is highly dominated by directivity effects. The findings of this study, therefore, support the view that the intensity of near-fault records is a continuous phenomenon (Tarbali et al., 2019), and is reflected in the spectral shape of the records and other relevant intensity measures. Improving the seismic analysis of near-fault structures should, therefore, focus on improving the representation of the target hazard spectra for near-fault locations using both real and physics-based simulated ground motions.

- Near-fault ground motion records associated with large fling pulses may have different spectral shapes than those without fling pulses and have the same causal parameters (magnitude and distance). Fling-containing records in this study tended to have higher spectral energy, especially in the long period range. Statistical analysis of the seismic demands on buildings with different dynamic characteristics revealed that the parameters representing fling were important predictors of the seismic demands when the spectral shape of the underlying records was not considered, particularly for flexible buildings, but their importance decreased significantly when the records were defined by their full spectral shapes. Ground motion selection experiments which match a target hazard at locations with substantial fling effects revealed that selecting records based on the full spectral shape alone was sufficient for representing the impacts of fling pulses on structures. It is unclear, however, whether target hazard spectra based on GMPEs would sufficiently represent the spectral shapes associated with large-fling-containing ground motions, and further work is needed in that regard.

- For the magnitude 7.0 shallow crustal event considered in this study, the structural demand trends associated with selected suites of simulated near-fault ground motions are strongly consistent with those associated with real near-fault ground motions, when the records are selected based on spectral shape features. Therefore, for near-fault locations where appropriate real ground motions are not available, physics-based simulated ground motions such as those used in this study present a viable substitute for the analysis of near-fault structures.

- Based on the relatively small number of realizations of the magnitude 7.0 event considered in this study with a pure strike-slip mechanism and a vertical fault, the simulated earthquake scenarios may not yet be appropriate for representing a broader target hazard at very small distances, on the order of 1 km, until more variability is added to the database of simulations. However, for relatively larger near-fault distances (5 km in this study), they are highly consistent with GMPE-based target hazard spectra, even with a relatively small number of rupture realizations and a simplistic velocity structure. The inconsistency of the seismic hazard spectra based on earthquake simulations and GMPE models at very short distances should be the focus of future research, and will likely diminish as more real near-fault records become available, and more earthquake simulations are generated for different rupture conditions.

Although the characteristics of the simulated ground motions are shown to be consistent with real near-fault records, it is important to note that the findings of this study are based on the analysis of site-specific simulated earthquake records only, and that the computational domain in this study contains relatively simplistic geology corresponding to either generic rock or soft soil conditions. The findings of this study are anticipated to be applicable to the selection of real ground

motions, as long as the target response spectra are appropriately representative. However, these findings will continue to be tested as more simulations and field observations of near-fault ground motions become available. Future research should also consider the effects of differing site conditions between the target and selected records. It is also important to note that raw records were used in this study, without scaling or alteration. Consequently, the effects of record scaling on the representation of directivity or fling pulse effects were not considered and are left for future research. Finally, the database of simulations consists of rupture realizations of a single magnitude 7.0 strike-slip event with a vertical fault, and the study generally focuses only on the variability induced by the rupture slip characteristics and source-to-site distance. Therefore, the study does not consider the variability associated with records from events with different moment magnitudes. In addition, the findings of this study should be considered applicable to strike-slip events, but may not capture the complexity associated with other rupture mechanisms.

7 Project Data

The results of the structural simulations conducted in this project will be made available on the DesignSafe data depot. The structural simulation models used in this study are available on the author's Github repository at https://github.com/mmkenawy/Regional_RC_Simulations.

8 Publications and Dissemination

A publication containing some of the findings of this study has been recently published in the journal Earthquake Spectra (Kenawy et al., 2023). A conference paper has been prepared for presentation at the 2024 18th World Conference on Earthquake Engineering in Milan, Italy. A second journal article summarizing additional findings is currently under preparation.

9 Acknowledgment of Support and Disclaimer

This material is based upon work supported by the U.S. Geological Survey under Grant No. G22AP00380. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey. The rupture simulations in this study were contributed by Dr. Arben Pitarka, whose work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract Number DE-AC52-07NA27344. A portion of the ground motion simulations was performed on the LLNL's High-Performance Computing (HPC) Systems and were funded by the LLNL Computing Grand Challenge 16 'Broadband Earthquake Ground Motion Simulations on LLNL's Next Generation HPC Systems'. The study also uses fault rupture simulations which were produced under 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, whose principal investigator is Dr. David McCallen at the University of Nevada, Reno, and Lawrence Berkeley National Laboratory. The author is grateful for Dr. Jeff Bayless's input toward completing the ground motion selection study, and the support of Dr. John Osteraas throughout the duration of the project.

References

Abrahamson, N. A. (2000). Effects of rupture directivity on probabilistic seismic hazard analysis. Proceedings of the 6th international conference on seismic zonation, Palm Springs, CA, pages 151–156.

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.

Akkar, S. and Boore, D. M. (2009). On baseline corrections and uncertainty in response spectrafor baseline variations commonly encountered in digital accelerograph records. *Bulletin of the Seismological Society of America*, 99(3):1671–1690.

Alavi, B. and Krawinkler, H. (2004). Behavior of moment-resisting frame structures subjected to near-fault ground motions. *Earthquake engineering & structural dynamics*, 33(6):687–706.

Almufti, I., Motamed, R., Grant, D. N., and Willford, M. (2015). Incorporation of velocity pulses in design ground motions for response history analysis using a probabilistic framework. *Earthquake Spectra*, 31(3):1647–1666.

American Concrete Institute (2014). *Building Code Requirements for Structural Concrete (ACI 318-14)*.

American Society of Civil Engineers (2017). *Minimum Design Loads and Associated Criteria for Buildings and Other Structures (ASCE/SEI 7-16)*.

American Society of Civil Engineers (2022). *Minimum Design Loads and Associated Criteria for Buildings and Other Structures (ASCE/SEI 7-22)*.

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.

Archuleta, R., Steidl, J., and Squibb, M. (2005). The Cosmos Virtual Data Center: Open dissemination of worldwide strong motion data. In *Directions in Strong Motion Instrumentation*, pages 209–222. Springer.

Baker, J. W. (2007). Quantitative classification of near-fault ground motions using wavelet analysis. *Bulletin of the seismological society of America*, 97(5):1486–1501.

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. (2004). Choice of a vector of ground motion intensity measures for seismic demand hazard analysis. In *13th World conference on earthquake engineering*, number 3384. Citeseer.

Baker, J. W. and Jayaram, N. (2008). Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra*, 24(1):299–317.

Bayless, J. and Abrahamson, N. (2023). A method for generating spectrum-compatible earthquake ground motion time histories with permanent displacement. In *Proc. of the 2023 United States Society on Dams Annual Conference, Charleston, South Carolina*.

Behr, J., Bryant, B., Given, D., Gross, K., Hafner, K., Hardebeck, J., Hauksson, E., Heaton, T., Hough, S., Hudnut, K., et al. (2000). Preliminary report on the 16 October 1999 m 7.1 Hector Mine, California, earthquake. *Seismological Research Letters*, 71(1):11–23.

Bertero, V. V., Mahin, S. A., and Herrera, R. A. (1978). Aseismic design implications of near-fault san fernando earthquake records. *Earthquake engineering & structural dynamics*, 6(1):31–42.

Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1):1063–1095.

Bijelić, N., Lin, T., and Deierlein, G. G. (2019a). Evaluation of building collapse risk and drift demands by nonlinear structural analyses using conventional hazard analysis versus direct simulation with cybershake seismograms. *Bulletin of the Seismological Society of America*, 109(5):1812–1828.

Bijelić, N., Lin, T., and Deierlein, G. G. (2019b). Quantification of the influence of deep basin effects on structural collapse using scec cybershake earthquake ground motion simulations. *Earthquake Spectra*, 35(4):1845–1864.

Bolt, B. (2002). Estimation of strong seismic ground motions. *International handbook of earthquake and engineering seismology*, pages 983–1001.

Boore, D. M. (2001). Effect of baseline corrections on displacements and response spectra for several recordings of the 1999 Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America*, 91(5):1199–1211.

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. and Bommer, J. J. (2005). Processing of strong-motion accelerograms: needs, options and consequences. *Soil Dynamics and Earthquake Engineering*, 25(2):93–115.

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.

Bradley, B. A. (2019). On-going challenges in physics-based ground motion prediction and insights from the 2010–2011 Canterbury and 2016 Kaikoura, New Zealand earthquakes. *Soil Dynamics and Earthquake Engineering*, 124:354–364.

Bradley, B. A., Burks, L. S., and Baker, J. W. (2015). Ground motion selection for simulation-based seismic hazard and structural reliability assessment. *Earthquake Engineering & Structural Dynamics*, 44(13):2321–2340.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Breiman, L. (2004). Consistency for a simple model of random forests. *University of California at Berkeley. Technical Report*, 670.

Burks, L. and Baker, J. (2015). Ground motion time series to accompany 'a predictive model for fling-step in near-fault ground motions based on recordings and simulations'. Processed data, Stanford University.

Burks, L. S. (2014). *Ground motion simulations: Validation and application for civil engineering problems*. PhD thesis, Stanford University, 450 Serra Mall, Stanford, CA.

Burks, L. S. and Baker, J. W. (2014). Fling in near-fault ground motions and its effect on structural collapse capacity. In *Tenth US National Conference on Earthquake Engineering Frontiers of Earthquake Engineering*, pages 21–25.

Burks, L. S. and Baker, J. W. (2016). A predictive model for fling-step in near-fault ground motions based on recordings and simulations. *Soil Dynamics and Earthquake Engineering*, 80:119–126.

Burks, L. S., Zimmerman, R. B., and Baker, J. W. (2015). Evaluation of hybrid broadband ground motion simulations for response history analysis and design. *Earthquake Spectra*, 31(3):1691–1710.

Calugaru, V. and Panagiotou, M. (2012). Response of tall cantilever wall buildings to strong pulse type seismic excitation. *Earthquake engineering & structural dynamics*, 41(9):1301–1318.

Campbell, K. W. and Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra*, 30(3):1087–1115.

Champion, C. and Liel, A. (2012). The effect of near-fault directivity on building seismic collapse risk. *Earthquake Engineering & Structural Dynamics*, 41(10):1391–1409.

Chiou, B. S.-J. and Youngs, R. R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 30(3):1117–1153.

Coleman, T., Peng, W., and Mentch, L. (2022). Scalable and efficient hypothesis testing with random forests. *The Journal of Machine Learning Research*, 35(1).

Daei, A., Poursha, M., and Zarrin, M. (2023). Sensitivity of structural responses to the processing of near-fault ground motion records containing fling-step. *Advances in Structural Engineering*, 26(6):1114–1129.

Dalguer, L. A., Day, S. M., Atkinson, G. M., and Chen, R. (2021). Introduction to the special section on fault displacement and near-source ground-motion models. *Bulletin of the Seismological Society of America*, 111(5):2271–2274.

de la Torre, C. A., Bradley, B. A., and Lee, R. L. (2020). Modeling nonlinear site effects in physics-based ground motion simulations of the 2010–2011 Canterbury earthquake sequence. *Earthquake Spectra*, 36(2):856–879.

Dreger, D. S., Beroza, G. C., Day, S. M., Goulet, C. A., Jordan, T. H., Spudich, P. A., and Stewart, J. P. (2015). Validation of the SCEC broadband platform v14.3 simulation methods using pseudospectral acceleration data. *Seismological Research Letters*, 86(1):39–47.

D'Amico, M., Felicetta, C., Schiappapietra, E., Pacor, F., Gallovič, F., Paolucci, R., Puglia, R., Lanzano, G., Sgobba, S., and Luzi, L. (2018). Fling effects from near-source strong-motion records: Insights from Mw 6.5, 2016, Norcia earthquake (Central Italy). *Seismol. Res. Lett*.

Eads, L., Miranda, E., and Lignos, D. G. (2015). Average spectral acceleration as an intensity measure for collapse risk assessment. *Earthquake Engineering & Structural Dynamics*, 44(12):2057–2073.

Fialko, Y., Simons, M., and Agnew, D. (2001). The complete (3-D) surface displacement field in the epicentral area of the 1999 Mw7. 1 Hector Mine earthquake, California, from space geodetic observations. *Geophysical research letters*, 28(16):3063–3066.

Garini, E., Gazetas, G., and Anastasopoulos, I. (2011). Asymmetric Newmark sliding caused by motions containing severe directivity and fling pulses. *Géotechnique*, 61(9):733–756.

Gazetas, G., Garini, E., and Georgarakos, T. (2008). Effects of near-fault ground shaking on sliding systems. In *Geotechnical Earthquake Engineering and Soil Dynamics IV*, pages 1–23.

Giberson, M. F. (1967). *The response of nonlinear multi-story structures subjected to earthquake excitation*. PhD thesis, California Institute of Technology, Pasadena, CA.

Graves, R. and Pitarka, A. (2016). Kinematic ground-motion simulations on rough faults including effects of 3D stochastic velocity perturbations. *Bulletin of the Seismological Society of America*, 106(5):2136–2153.

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.

Gregor, N., Abrahamson, N. A., Atkinson, G. M., Boore, D. M., Bozorgnia, Y., Campbell, K. W., Chiou, B. S.-J., Idriss, I., Kamai, R., Seyhan, E., et al. (2014). Comparison of NGA-West2 GMPEs. *Earthquake Spectra*, 30(3):1179–1197.

Grigoriu, M. (2011). To scale or not to scale seismic ground-acceleration records. *Journal of engineering mechanics*, 137(4):284–293.

Hall, J. F., Heaton, T. H., Halling, M. W., and Wald, D. J. (1995). Near-source ground motion and its effects on flexible buildings. *Earthquake spectra*, 11(4):569–605.

Hamidi, H., Karbassi, A., and Lestuzzi, P. (2020). Seismic response of RC buildings subjected to fling-step in the near-fault region. *Structural Concrete*, 21(5):1919–1937.

Haselton, C. B., Liel, A. B., Taylor-Lange, S. C., and Deierlein, G. G. (2016). Calibration of model to simulate response of reinforced concrete beam-columns to collapse. *ACI Structural Journal*, 113(6).

Hayden, C. P., Bray, J. D., and Abrahamson, N. A. (2014). Selection of near-fault pulse motions. *Journal of Geotechnical and Geoenvironmental Engineering*, 140(7):04014030.

Hisada, Y. and Bielak, J. (2003). A theoretical method for computing near-fault ground motions in layered half-spaces considering static offset due to surface faulting, with a physical interpretation of fling step and rupture directivity. *Bulletin of the Seismological Society of America*, 93(3):1154–1168.

Hisada, Y. and Tanaka, S. (2021). What is fling step? Its theory, simulation method, and applications to strong ground motion near surface fault ruptures. *Bulletin of the Seismological Society of America*, 111(5):2486–2506.

Housner, G. and Trifunac, M. (1967). Analysis of accelerograms—parkfield earthquake. *Bulletin of the seismological society of America*, 57(6):1193–1220.

Huang, Y.-N., Whittaker, A. S., Luco, N., and Hamburger, R. O. (2011). Scaling earthquake ground motions for performance-based assessment of buildings. *Journal of Structural Engineering*, 137(3):311–321.

Ibarra, L. F., Medina, R. A., and Krawinkler, H. (2005). Hysteretic models that incorporate strength and stiffness deterioration. *Earthquake engineering & structural dynamics*, 34(12):1489–1511.

Iwan, W., Huang, C.-T., and Guyader, A. C. (2000). Important features of the response of inelastic structures to near-field ground motion. Proc. of the 12th World Conf. on Earthquake Engineering. International Association for Earthquake Engineering.

Jeong, S. and Bradley, B. A. (2017). Amplification of strong ground motions at Heathcote Valley during the 2010–2011 Canterbury earthquakes: The role of 2D nonlinear site response. *Bulletin of the Seismological Society of America*, 107(5):2117–2130.

Kalkan, E. and Kunnath, S. K. (2006). Effects of fling step and forward directivity on seismic response of buildings. *Earthquake spectra*, 22(2):367–390.

Kamai, R., Abrahamson, N., and Graves, R. (2014). Adding fling effects to processed ground-motion time histories. *Bulletin of the Seismological Society of America*, 104(4):1914–1929.

Karabulut, H. and Bouchon, M. (2007). Spatial variability and non-linearity of strong ground motion near a fault. *Geophysical Journal International*, 170(1):262–274.

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. Technical report, 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.

Kenawy, M., McCallen, D., and Pitarka, A. (2023). Characteristics and selection of near-fault simulated earthquake ground motions for nonlinear analysis of buildings. *Earthquake Spectra*.

Kohrangi, M., Vamvatsikos, D., and Bazzurro, P. (2019). Pulse-like versus non-pulse-like ground motion records: spectral shape comparisons and record selection strategies. *Earthquake Engineering & Structural Dynamics*, 48(1):46–64.

Li, C., Kunnath, S., Zuo, Z., Peng, W., and Zhai, C. (2020). Effects of early-arriving pulse-like ground motions on seismic demands in RC frame structures. *Soil Dynamics and Earthquake Engineering*, 130:105997.

Liossatou, E. and Fardis, M. N. (2016). Near-fault effects on residual displacements of RC structures. *Earthquake engineering & structural dynamics*, 45(9):1391–1409.

Luco, N. and Cornell, C. A. (2007). Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthquake Spectra*, 23(2):357–392.

MacRae, G. A., Morrow, D. V., and Roeder, C. W. (2001). Near-fault ground motion effects on simple structures. *Journal of Structural Engineering*, 127(9):996–1004.

Malhotra, P. K. (1999). Response of buildings to near-field pulse-like ground motions. *Earthquake engineering & structural dynamics*, 28(11):1309–1326.

Marafi, N. A., Eberhard, M. O., Berman, J. W., Wirth, E. A., and Frankel, A. D. (2019). Impacts of simulated M9 cascadia subduction zone motions on idealized systems. *Earthquake Spectra*, 35(3):1261–1287.

Mavroeidis, G., Dong, G., and Papageorgiou, A. (2004). Near-fault ground motions, and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems. *Earthquake Engineering & Structural Dynamics*, 33(9):1023–1049.

McCallen, D., Petersson, A., Rodgers, A., Pitarka, A., Miah, M., Petrone, F., Sjogreen, B., Abrahamson, N., and Tang, H. (2021a). 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.

Mccallen, D., Petersson, N., Rodgers, A., Miah, M., Pitarka, A., Petrone, F., and Tang, H. (2020). The earthquake simulation (EQSIM) framework for physics-based fault-to-structure simulations. Proceedings of the 17th World Conference on Earthquake Engineering, Sendai, Japan.

McCallen, D., Petrone, F., Miah, M., Pitarka, A., Rodgers, A., and Abrahamson, N. (2021b). EQSIM—A multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers, part II: Regional simulations of building response. *Earthquake Spectra*, 37(2):736–761.

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

Mentch, L. and Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *The Journal of Machine Learning Research*, 17(1):841–881.

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

Paolucci, R., Smerzini, C., and Vanini, M. (2021). BB-SPEEDset: A validated dataset of broadband near-source earthquake ground motions from 3D physics-based numerical simulations. *Bulletin of the Seismological Society of America*, 111(5):2527–2545.

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*, pages 1–23.

Pitarka, A., Graves, R., Irikura, K., Miyakoshi, K., Wu, C., Kawase, H., Rodgers, A., and McCallen, D. (2022). Refinements to the graves–pitarka kinematic rupture generator, including a dynamically consistent slip-rate function, applied to the 2019 Mw 7.1 Ridgecrest earthquake. *Bulletin of the Seismological Society of America*, 112(1):287–306.

Pitarka, A., Irikura, K., Iwata, T., and Kagawa, T. (1996). Basin structure effects in the Kobe area inferred from the modeling of ground motions from two aftershocks of the january 17, 1995, hyogo-ken nanbu earthquake. *Journal of Physics of the Earth*, 44(5):563–576.

Reddy, K. K., Somala, S. N., and Tsang, H.-H. (2021). Response of inelastic SDOF systems subjected to dynamic rupture simulations involving directivity and fling step. *Soil Dynamics and Earthquake Engineering*, 151:106992.

Rodgers, A. J., Pitarka, A., and McCallen, D. B. (2019). The effect of fault geometry and minimum shear wavespeed on 3D ground-motion simulations for an Mw 6.5 Hayward fault scenario earthquake, San Francisco Bay Area, Northern California. *Bulletin of the Seismological Society of America*, 109(4):1265–1281.

Schiappapietra, E., Lanzano, G., and Sgobba, S. (2022). Empirical predictive models for fling step and displacement response spectra based on the NESS database. *Soil Dynamics and Earthquake Engineering*, 158:107294.

Sgobba, S., Felicetta, C., Lanzano, G., Ramadan, F., D'Amico, M., and Pacor, F. (2021). NESS2.0: An updated version of the worldwide dataset for calibrating and adjusting ground-motion models in near source. *Bulletin of the Seismological Society of America*, 111(5):2358–2378.

Shabestari, K. T. and Yamazaki, F. (2003). Near-fault spatial variation in strong ground motion due to rupture directivity and hanging wall effects from the Chi-Chi, Taiwan earthquake. *Earthquake engineering & structural dynamics*, 32(14):2197–2219.

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. (2014a). An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions. *Bulletin of the Seismological Society of America*, 104(5):2456–2466.

Shahi, S. K. and Baker, J. W. (2014b). NGA-West2 models for ground motion directionality. *Earthquake Spectra*, 30(3):1285–1300.

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.

Somerville, P. G. (2003). Magnitude scaling of the near fault rupture directivity pulse. *Physics of the earth and planetary interiors*, 137(1-4):201–212.

Somerville, P. G., Smith, N. F., Graves, R. W., and Abrahamson, N. A. (1997). Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effects of rupture directivity. *Seismological research letters*, 68(1):199–222.

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.

Tothong, P. and Cornell, C. A. (2008). Structural performance assessment under near-source pulse-like ground motions using advanced ground motion intensity measures. *Earthquake Engineering & Structural Dynamics*, 37(7):1013–1037.

Ventura, C. E., Archila, M., Bebamzadeh, A., and Liam Finn, W. (2011). Large coseismic displacements and tall buildings. *The Structural Design of Tall and Special Buildings*, 20:85–99.

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.

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*, page 87552930211014540.