

An open-access simulated earthquake ground-motion database for an M7 Hayward Fault earthquake in the San Francisco Bay Region

Earthquake Spectra

1–38

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/87552930251340960

journals.sagepub.com/home/eqs

David McCallen, M.EERI¹ , Arben Pitarka² ,
Houjun Tang¹, Rie Nakata¹,
Khalid M Mosalam, M.EERI³,
Floriana Petrone, M.EERI⁴ ,
Selim Günay³, and Claudio Perez³ 

Abstract

Comprehensive understanding of earthquake ground motions, particularly in the near-fault region of large-magnitude events, is limited by gaps in strong-motion data. This challenge is prominent in areas with high seismic hazard but infrequent large earthquakes where data is sparse and difficult to interpret. These data limitations lead to uncertainties in the development of site-specific ground motions, which are crucial for engineering risk assessments. To address these challenges, physics-based regional-scale ground-motion simulations have been developed. With the emergence of exaflop-scale computing ecosystems, it is now possible to simulate regional earthquake processes at unprecedented fidelity and generate the large number of fault rupture realizations necessary to characterize both intra- and inter-event ground-motion variability. This article introduces a new database of simulated earthquake ground motions, created for applications in earthquake engineering, earthquake planning, and emergency response. The inaugural version of the database features simulated ground motions for a magnitude 7 Hayward Fault earthquake in the San Francisco Bay Region (SFBR), using the Earthquake SIMulation (EQSIM) simulation framework and the Graves–Pitarka kinematic rupture model. The aim is to provide high-fidelity, spatially dense, three-component motions generated on the Department of Energy’s (DOE) newest generation of graphics processing unit (GPU)-accelerated supercomputers. These motions are

¹Lawrence Berkeley National Laboratory, Berkeley, CA, USA

²Lawrence Livermore National Laboratory, Livermore, CA, USA

³Pacific Earthquake Engineering Research Center, University of California, Berkeley, Berkeley, CA, USA

⁴University of Nevada, Reno, Reno, NV, USA

Corresponding author:

David McCallen, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA.

Email: dbmccallen@lbl.gov

being made openly available to the engineering, scientific, and disaster planning communities. In addition, this work develops protocols for the efficient dissemination of these large data sets and emphasizes community engagement to build confidence in their application. This article discusses the methodology behind the data, underlying software verification and validation, scalable data management, and a user interface for data access. The goal is to facilitate widespread use and elicit expert feedback to maximize the utility and exploitation of simulated motions. While the initial focus is on the San Francisco Region, simulations for additional regions will be added as the DOE program progresses.

Keywords

Simulated ground-motion database, regional earthquake ground motions, regional-scale earthquake simulations, open-access ground-motion data, physics-based simulations

Date received: 19 December 2024; accepted: 19 April 2025

Introduction

There has been growing community interest in the application of high-performance computing to simulating earthquake processes at a regional scale (Akinci et al., 2024; Graves, 2022; Graves et al., 2011; Graves and Pitarka, 2010; McCallen et al., 2024b; Maechling et al., 2015; Paolucci et al., 2018, 2021; Petrone et al., 2024; Pitarka et al., 2021a; Rodgers et al., 2020; Smerzini et al., 2024; Taborda and Bielak, 2011; Wang et al., 2024; Yeh and Olsen, 2024; Zhang et al., 2021b). There are two principal approaches to regional-scale earthquake ground-motion simulations. The *hybrid* approach performs deterministic physics-based simulations of low-frequency waveforms (typically to 1 Hz). The high-frequency portion of the seismogram is computed with a stochastic representation and a simplified kinematic rupture model. The final broadband time series is obtained by combining the low- and high-frequency parts of the seismogram. The SCEC Broad Band Platform (e.g. Goulet et al., 2015; Graves and Pitarka, 2010) provides a mature example of the hybrid approach. Alternatively, the *physics-based deterministic* approach utilizes physics-based simulations to directly create full broadband time series, preserving correct spatial time phasing of motions throughout the domain of interest. Fully deterministic simulations have significant computational demands and are progressing based on the major advancements in high-performance computing ecosystems and massively parallel computational software frameworks (McCallen et al., 2024b). The simulated ground-motion database (SGMD) described herein was generated with EarthQuake SIMulation's (EQSIM) physics-based deterministic broadband simulations.

Regional-scale simulations can provide new physical insight into the complex spatial variability of ground motions and infrastructure damage, and the factors that most influence the ground-motion intensity and its spatial distribution. With the available memory and computational speeds realized on today's graphics processing unit (GPU)-accelerated computers, numerical exploration of the modeling parameter space can be achieved through large numbers of simulation realizations to generate ground-motion data that accounts for the uncertainty in model parameters. From the standpoint of characterization of hazard and risk, simulated ground-motion data can also augment the existing sparse observational database of recorded earthquake motions by supplying broadband, site-dependent ground motions. The increasing interest of the engineering community in

the use of simulated ground motions is supported by current seismic design standards, which allow supplementing available records with simulated motions to perform nonlinear response history analyses (e.g. American Society of Civil Engineers (ASCE)/SEI 7-22, 2021; CEN Eurocode 8 (European Standard EN 1998-1:2004, 2004); Standards New Zealand (NZS) 1170.5:2004, 2004). Specifically, the use of simulated motions is contemplated when there are insufficient records from earthquakes with tectonic regimes, magnitudes, and distances similar to those controlling the target spectrum at the site of interest. In addition, at near-field sites, simulated motions can be employed to ensure the incorporation of an adequate proportion of motions exhibiting directionality and impulsive characteristics.

Given the extreme computational demands associated with regional earthquake simulations and the specialized computational expertise required to achieve the most effective use of high-end massively parallel computations, there are a relatively small number of groups that perform regional simulations at extreme scale. In 2021, an International Pacific Rim Forum was organized by the Pacific Earthquake Engineering Research (PEER) Center to discuss the status of regional simulations and to obtain community feedback on mechanisms by which regional-scale simulated motions could be made accessible to the broad research and engineering practitioner communities (McCallen et al., 2022a). The consensus feedback suggested that the creation of an open-access SGMD would provide the opportunity to achieve the broadest access and distribution of simulated ground motions. Subsequently, the US Department of Energy (DOE) Office of Cyber Security, Energy Security and Emergency Response (CESER) initiated a program to develop a regional earthquake ground-motion database that would be available to energy system stakeholders and emergency response planners. Ground motions for the SGMD are being created with the EQSIM fault-to-structure simulation framework, which was developed for emerging GPU-accelerated massively parallel computers under the DOE Exascale Computing Project (McCallen et al., 2021a, 2021b, 2022b, 2024b). In addition, the experience accrued at the PEER Center in providing ground motions suitable for performance-based earthquake engineering (PBEE) evaluations is being leveraged to design an open-access database (Günay and Mosalam, 2013; Stewart et al., 2001). As described in this article, the simulated ground motions that have been generated for the San Francisco Bay Region (SFBR) constitute the first large regional data set to populate the SGMD.

Hayward Fault ground-motion simulations for the SFBR

For the evaluation of regional ground motions for M7 Hayward Fault events, the regional computational model indicated in Figure 1 was employed in the SW4 seismic wave propagation program within the EQSIM framework (Petersson and Sjögreen, 2010, 2012, 2014, 2015; Wang and Petersson, 2019; Zhang et al., 2021a). For the initial data set, the regional simulations were computed at a frequency resolution of 5 Hz with a simulation model minimum shear-wave velocity (V_{smin}) set to 250 m/s and three components of motion were computed (fault-normal, fault parallel, and vertical). The model utilized a linear viscoelastic material (Graves and Day, 2003) and frequency-independent attenuation factors Q_p and Q_s for P and S waves, respectively. Modeling studies comparing simulated motions with observed small SFBR earthquakes indicated that the United States Geological Survey (USGS) velocity model for the SFBR can reasonably be used to simulate ground motions to 5 Hz for many sites (Pinilla-Ramos et al., 2024), and imposing a V_{smin} of 250 m/s allows for efficient run times and the execution of a large number of fault rupture realizations while providing a reasonable estimate of near-surface soft sediment response. The

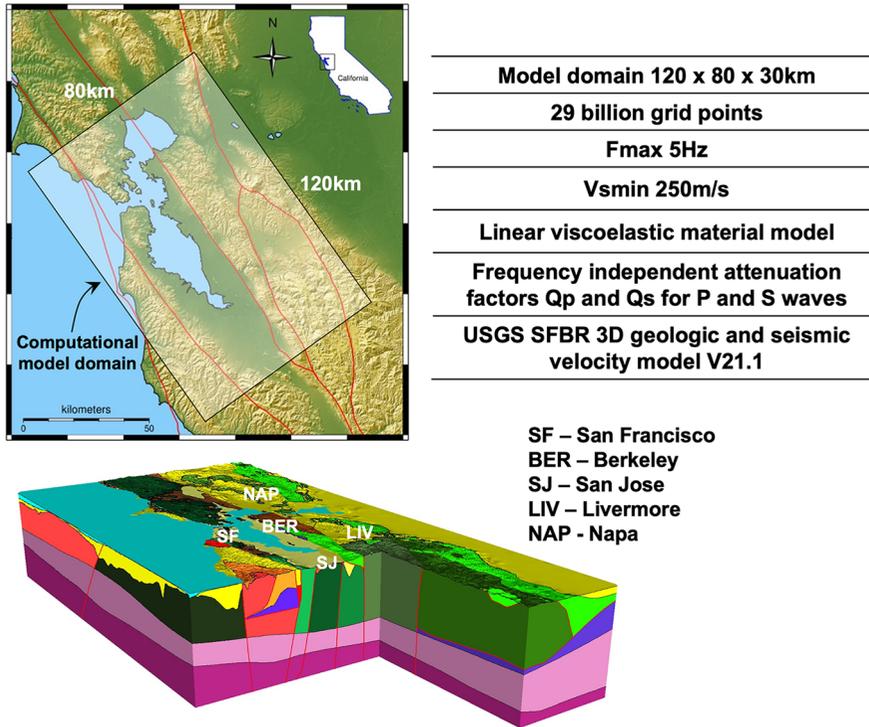
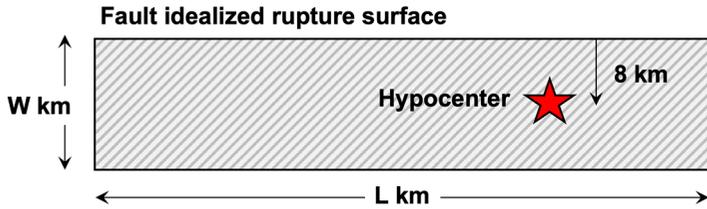


Figure 1. Computational domain considered for the SFBR ($120 \times 80 \times 30$ km), model parameters and the USGS 3D geologic model V21.1 (courtesy USGS).

ground-motion simulations were completed on the Frontier GPU-accelerated exaflop computer at the Oak Ridge National Laboratory (ORNL), the Perlmutter GPU-accelerated computer at Lawrence Berkeley National Laboratory (LBNL), and the Summit GPU-accelerated system at ORNL.

For the initial ground-motion simulations included in the database, the focus was placed on characterization and variability of the earthquake source term generated by a Hayward Fault rupture. Conceptually, uncertainties in geologic material properties could also be included by defining parametric distributions of geologic parameters. For the SFBR, the USGS 3D velocity model was created by assigning geologic properties to identified individual geologic units; this could potentially be extended to assigning baseline geologic properties along with an associated property distribution. However, this was beyond the scope of the current study. For this study, one deterministic geologic structure was considered based on the USGS three-dimensional velocity model for the SFBR as shown in Figure 1 (Aagaard and Hirakawa, 2021).

The precise manner by which a fault rupture initiates and evolves with time during an earthquake has a profound influence on the resulting distribution of ground motions. The rupture process for a given scenario earthquake is not well constrained by existing geophysical data, and it is essential to consider a breadth of rupture possibilities to characterize the potential ground motions. The fault rupture realizations utilized in the SFBR simulations were generated using the Graves and Pitarka (GP) kinematic rupture model (Graves and Pitarka, 2016), with an enhanced slip rate representation (Pitarka et al., 2021b).



- Hypocenter rupture initiation (variable, locations H1 - H5)
- Velocity of rupture propagation = $0.72 V_s$
- Fault rupture width (W) = Depends on event magnitude (15 km for this dataset)
- Fault rupture length (L) = Depends on event magnitude (66 km for this dataset)
- Fault slip (variable)
- Rupture rise time (variable)
- Rupture slip rate (variable)
- Slip rate roughness coefficient = 0.97
- Rise time coefficient = 1.9
- Focal mechanism spatial variation roughness = 2%
- Top of fault rupture below ground surface = 200 m
- Average fault mechanism set to strike slip with rake angle of 150 degrees
- Depth of hypocenter = 8 km
- Fault dipping angle 80 degrees towards the east

Figure 2. Graves–Pitarka kinematic fault rupture parameters.

The key parameters of the GP rupture model are illustrated in Figure 2. Constrained by dynamic rupture modeling and empirically based correlated rupture parameters, the GP model incorporates depth-dependent multi-scale spatial variations of slip, slip rate, slip rake, and rupture velocity that allow for reproducing observed characteristics of near-fault ground motion on a broad frequency range. The slip represents the total permanent fault displacement once the rupture process completes, and the slip rate is the time history of the slip velocity at any location on the fault rupture surface. As such, the slip rate, also called the source time function, defines the spatial and temporal seismic energy release during the fault rupture. The slip rate contributes to the near-fault ground velocities, which is one of the features of near-fault ground motions correlated with structural damage (e.g. Akkar and Özen, 2005). The time derivative of the slip rate represents the slip acceleration that directly contributes to the near-fault ground-motion accelerations. Therefore, the slip rate is one of the most important parameters of the kinematic rupture model. Combined with the fault orientation, given by the strike angle, and the fault dipping orientation with depth, given by the dip angle, the rake angle represents the focal mechanism that controls the seismic wave radiation pattern generated at each point on the fault rupture surface. The rupture velocity characterizes the speed of the rupture front propagation during the earthquake rupture evolution.

For the regional ground-motion data set, the parameters that were varied for generation of fault stochastic rupture realizations, as shown in Figure 3a, included the location of the rupture initiation point, fault slip, rise time, and slip rate. The GP rupture generator also includes an option for creating hybrid ruptures with combined stochastic and deterministic spatial slip features as shown in Figure 3b. The hybrid ruptures are designed to represent

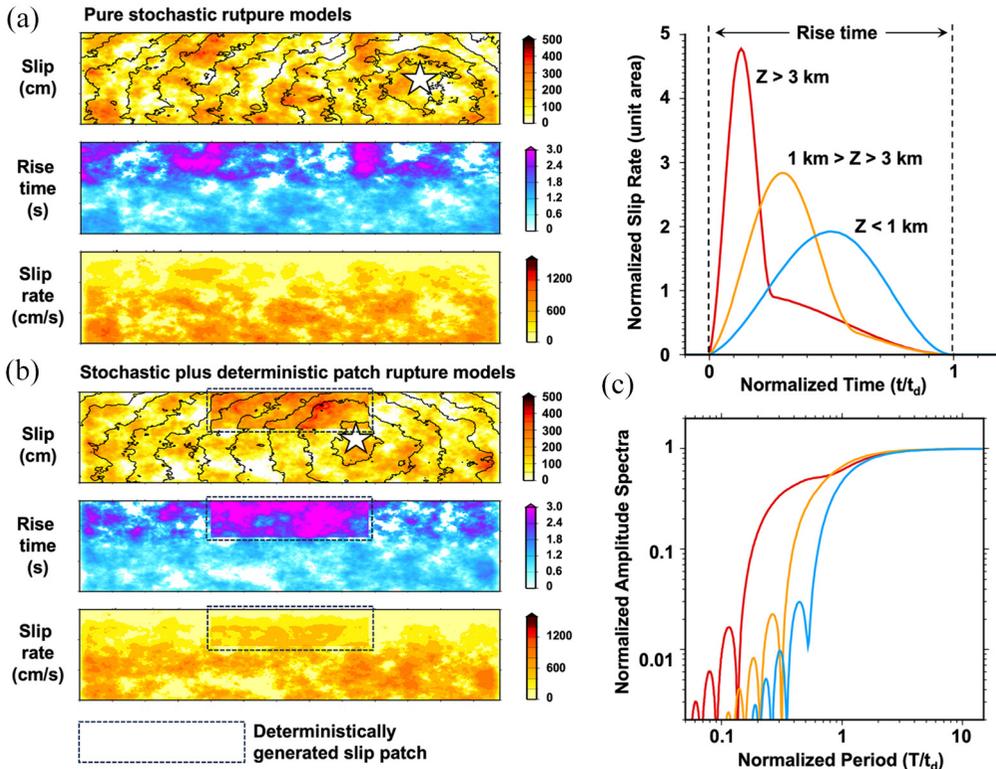


Figure 3. Rupture models considered within the context of the Graves–Pitarka kinematic rupture model for the 50 baseline Hayward Fault rupture realizations: (a) stochastic model of rupture parameters, (b) stochastic rupture model with specified deterministic rupture patches, and (c) Liu slip-rate functions and corresponding Fourier amplitude spectra at different depths adopted in the GP rupture model, Pitarka et al. (2021b).

slip “patches” characterizing areas of enhanced seismic energy release which, depending on their depth, affect different frequency ranges.

The deterministic slip patches are user-defined and are based on the Irikura recipe (Irikura and Miyake, 2011). The resulting rupture process is randomly heterogeneous at different length scales (Pitarka et al., 2017).

As a result of material heterogeneity both in terms of rheology and frictional properties and local stress, the rupture velocity can be highly variable in space. Intrinsically, particularly for large earthquakes, it can reach values that are similar in amplitude or larger than the local shear-wave velocity (near-shear or super-shear rupture velocity). During near-shear and super-shear ruptures, the rupture front propagates with a speed that is close to that of the shear waves propagating toward a site. Therefore, a large-amplitude shear-wave front is formed by the constructive superposition of the shear waves traveling with or ahead of the rupture front. This phenomenon, which creates forward rupture directivity, is most pronounced for strike-slip earthquakes and unilateral ruptures. In the GP rupture model, the rupture velocity mainly correlates with the local slip and is highly variable. Because the average rupture velocity is an input parameter, the rupture generator allows for controlled generation of super-shear rupture scenarios with enhanced rupture

directivity effects. The average rupture velocity was set at 72% of the local shear-wave velocity in all rupture realizations. The general form of the slip-rate function formulation is based on the approach of Liu et al. (2006). The shape of the source time function varies with depth with a linear transition between crustal depths of 1 and 3 km (Pitarka et al., 2021b). This formulation allows the shape of the source time function to transition from “cosine-type,” with a relatively modest high-frequency content in the depth interval 0–1 km, to “Kostrov-type” with a broadband frequency content, at depths greater than 3 km. The spatial correlation between the rise time and slip distributions is controlled by a small stochastic perturbation with a standard deviation and correlation coefficient designed to emulate the roughness of spatial peak slip rate variation and its relative increase with depth including concentrated zones of high slip rate. The rupture model is designed to reproduce observations indicating that the high-frequency part of the earthquake ground motion is primarily controlled by high stress drop areas located in the seismogenic zone starting at a depth of approximately 3 km.

As summarized in the validation discussion below, the GP model has been tested through comparisons of simulated broadband ground motions against empirical ground-motion models as well as validated through direct comparisons with observational data from crustal earthquakes. In the initial database that includes multiple realizations of an M7 earthquake on the Hayward Fault, rupture model variabilities include the rupture initiation and spatial slip distribution, including ruptures with pre-determined large slip patches. The numerical model implementation of the GP kinematic fault rupture is illustrated in Figure 4 which shows time snapshots of the rupture front evolution at selected instants of time during the rupture process.

The Hayward Fault database includes 25 pure stochastic (R1–R25) and 25 hybrid stochastic plus patch (R1patch–R25patch) rupture realizations for a total of 50 M7 Hayward Fault realizations. The rupture models were created to span across five rupture hypocenter locations distributed along the fault and a stochastic distribution of the rupture model parameters indicated in Figure 3. The five hypocenter locations considered are shown in Figure 5 and listed in Table 1, Table 2 specifies the fault rupture nomenclature, the rupture hypocenter, the rupture velocity, and the rupture plane dimensions. In addition to the 50 baseline rupture realizations, the database includes five separate realizations with varying average rupture velocities (R1vr–R5vr) to allow investigation of the influence of fault rupture velocity on the degree of directivity generated and five realizations with variable fault plane dimensions (R1area–R5area) to allow investigation of ground-motion sensitivity to fault rupture area as shown in Table 3.

Figure 6 graphically illustrates the five rupture models associated with the pure stochastic ruptures at hypocenter H1 (ruptures R1–R5). The additional 20 stochastic rupture realizations at hypocenters H2 through H5 and the 25 hybrid ruptures at hypocenters H1 through H5 incorporated in the current SGMD are included in Supplemental attachment 1.

Software verification and validation for establishing confidence in simulated ground motions

For the creation of reliable and usable simulated ground-motion data sets, an important practical consideration is to ensure that utilized simulation software has gone through rigorous verification and validation. Up-to-date software verification on the specific high-performance computer platforms that are being used to generate the simulated motions

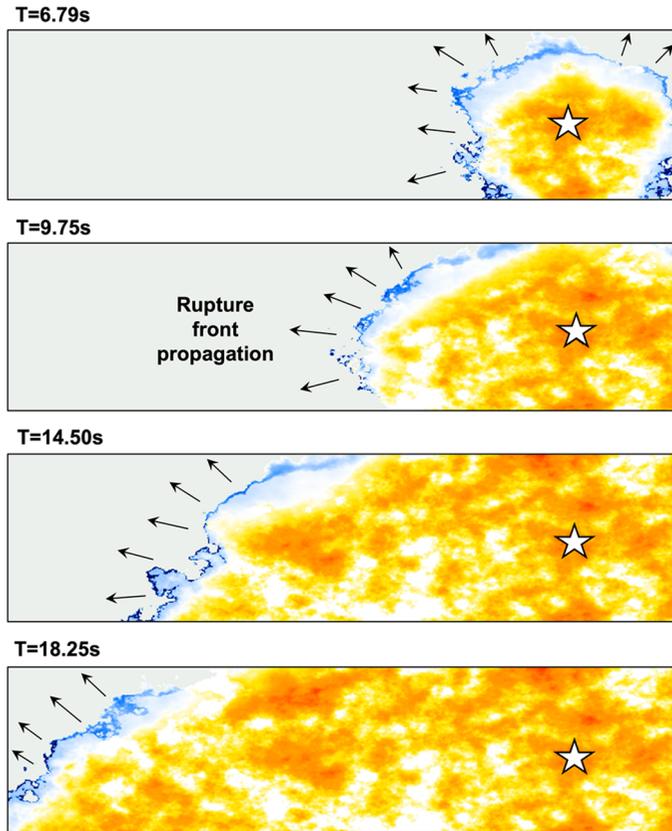


Figure 4. Time snapshots of the numerical model evolution of fault rupture front propagation and fault slip from the GP stochastic rupture model for a Hayward Fault earthquake.

Table I. Hayward Fault rupture hypocenter locations

Hypocenter location	Longitude	Latitude
H1	37.883	-122.247
H2	37.812	-122.185
H3	37.730	-122.109
H4	37.662	-122.052
H5	37.594	-121.991

and validation that demonstrates the physics embodied in the computational models is representative of the actual earthquake processes is essential. If multiple massively parallel computer platforms are being employed in the creation of ground-motion simulated data, the software must be verified across all platforms. For the SW4 seismic wave propagation code utilized in the EQSIM framework, an extensive set of regression test problems has been created that can be efficiently executed in automated batch mode to verify the platform-specific results of SW4. The test suite evaluates the set of key capabilities of the SW4 program through comparison of numerical results for a carefully developed set of

Table 2. Hayward Fault rupture realization parameters for the 50 baseline ruptures included in the simulated ground-motion database^{a,b}

Rupture numbers	Rupture realization designations	Rupture model	Hypocenter
1–5	R1—R5	Stochastic	H1
6–10	R6—R10	Stochastic	H2
11–15	R11—R15	Stochastic	H3
16–20	R16—R20	Stochastic	H4
21–25	R21—R25	Stochastic	H5
26–30	R1patch—R5patch	Hybrid	H1
31–35	R6patch—R10patch	Hybrid	H2
36–40	R11patch—R15patch	Hybrid	H3
41–45	R16patch—R20patch	Hybrid	H4
46–50	R21patch—R25patch	Hybrid	H5

^aThe five rupture realizations at each hypocenter location vary in terms of the automatically generated stochastic rupture parameters for the GP rupture model.

^bFor the 50 baseline models, the velocity of rupture (V_{rupture}) is equal to 72% of the local shear-wave velocity (V_{Slocal}) at each location on the fault and the rupture plane dimension is 64 km in length \times 15 km in width.

Table 3. Hayward Fault rupture realizations for variable average rupture velocity and variable fault rupture area

Number	Rupture realization	Rupture model	Hypocenter	V_{Rupture} (%)	Rupture plane dimensions (km)
1	R1vr	Hybrid	H2	55	64 \times 15
2	R2vr	Hybrid	H2	60	64 \times 15
3	R3vr	Hybrid	H2	65	64 \times 15
4	R4vr	Hybrid	H2	80	64 \times 15
5	R5vr	Hybrid	H2	85	64 \times 15
1	R1area	Hybrid	H3	72	60 \times 15
2	R2area	Hybrid	H3	72	58 \times 15
3	R3area	Hybrid	H3	72	64 \times 14.5
4	R4area	Hybrid	H3	72	66 \times 15
5	R5area	Hybrid	H3	72	62 \times 15

test problems with analytical solutions as well as solutions created by the method of manufactured solutions. The core set of test problems are listed in Supplemental Attachment 2 and can be found in the SW4 GitHub repository.¹

The ability to reliably move the EQSIM workflow across multiple computer platforms with minimum effort has also been facilitated through the utilization of the RAJA C++ software libraries for platform-to-platform portability (Beckingsale et al., 2019). For GPU-accelerated systems, RAJA allows minimization of computer platform-specific inner loop coding changes and substantially simplifies and streamlines a reliable verification process across multiple platforms.

Creation of a high-resolution SGMD typically demands a significant dedicated bank of awarded computer time and requires the utilization of multiple platforms located at multiple supercomputer sites. For the SGMD creation described herein, the three different DOE GPU-accelerated platforms shown in Figure 7 were all utilized for the completion of the

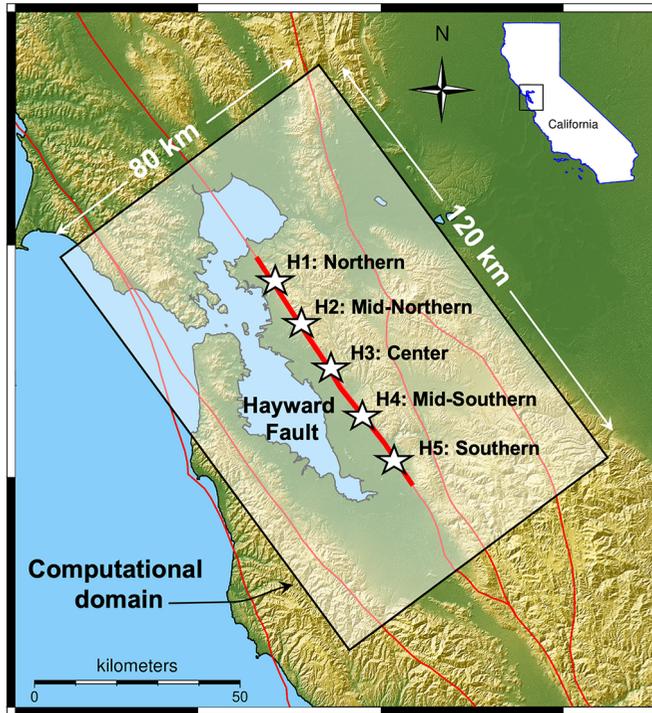


Figure 5. Five hypocenter locations for the 50 rupture realizations.

Hayward Fault rupture realizations. It is essential to maintain rigorous verification testing over the life cycle of a major computational initiative, as the individual machine ecosystems can go through modifications in terms of operating systems and software. For EQSIM simulations, as a best practice, the automated suite of SW4 regression tests is periodically executed across all platforms with particular attention paid to regression testing when major system or software changes are announced. This guards against a surprise discovery that the results from a particular platform could be in error after a large number of realizations have been completed and a large block of high-value supercomputer time, potentially up to tens of thousands of node hours, are expended. The critically important need of ensuring verified software essential to reliable massively parallel simulations has often not received the emphasis required and must be a rigorous element of SGMD creation.

The physics representation validation of the software would ideally be performed as a formal software validation whereby the software is used to model an observed, well-instrumented large earthquake, or ideally multiple large earthquakes, in a rigorous simulation-observational data comparison. However, in modeling earthquake phenomenon, traditional rigorous software validation is a challenge. For many regions, including the SFBR, observational data for even a single large-magnitude Hayward earthquake does not exist. Therefore, there is no existing well-constrained, well-instrumented “experiment” at the response amplitudes of primary interest, and alternative graded approaches to confidence building in the simulation results are necessary based on the data available. Unlike other fields where well-constrained parameter data are available for the system being modeled, earthquake simulations necessarily depend on input data with parametric uncertainty

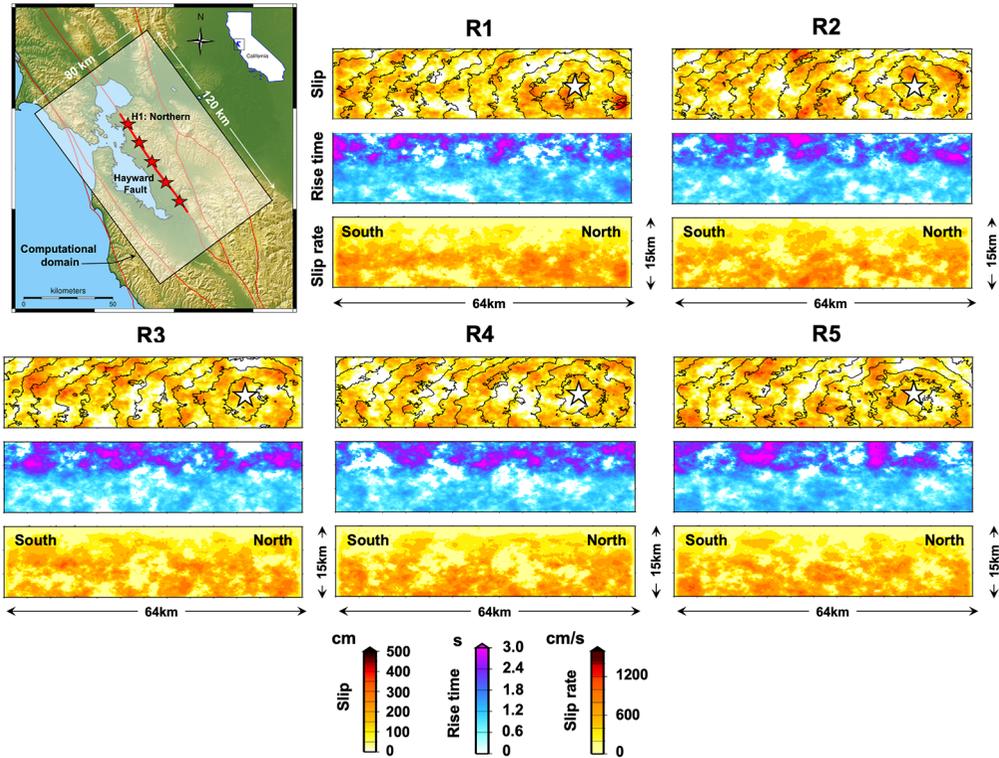


Figure 6. Stochastic rupture models for hypocenter location H1.

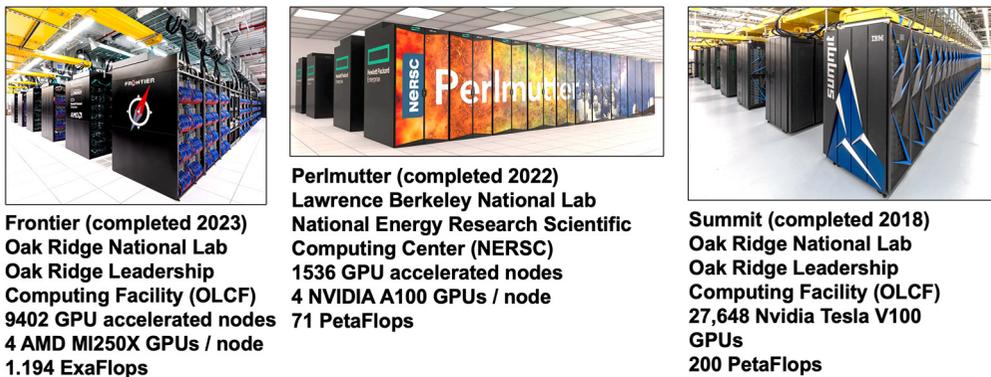


Figure 7. Three GPU-accelerated DOE massively parallel computers utilized in the SFBR Hayward Fault simulations for the SGMD.

including the regional geologic velocity models which have different degrees of refinement and definition depending on the region.

For the EQSIM framework, simulation model results have been subjected to validation and confidence building exercises in multiple ways. First, basic evaluations of EQSIM

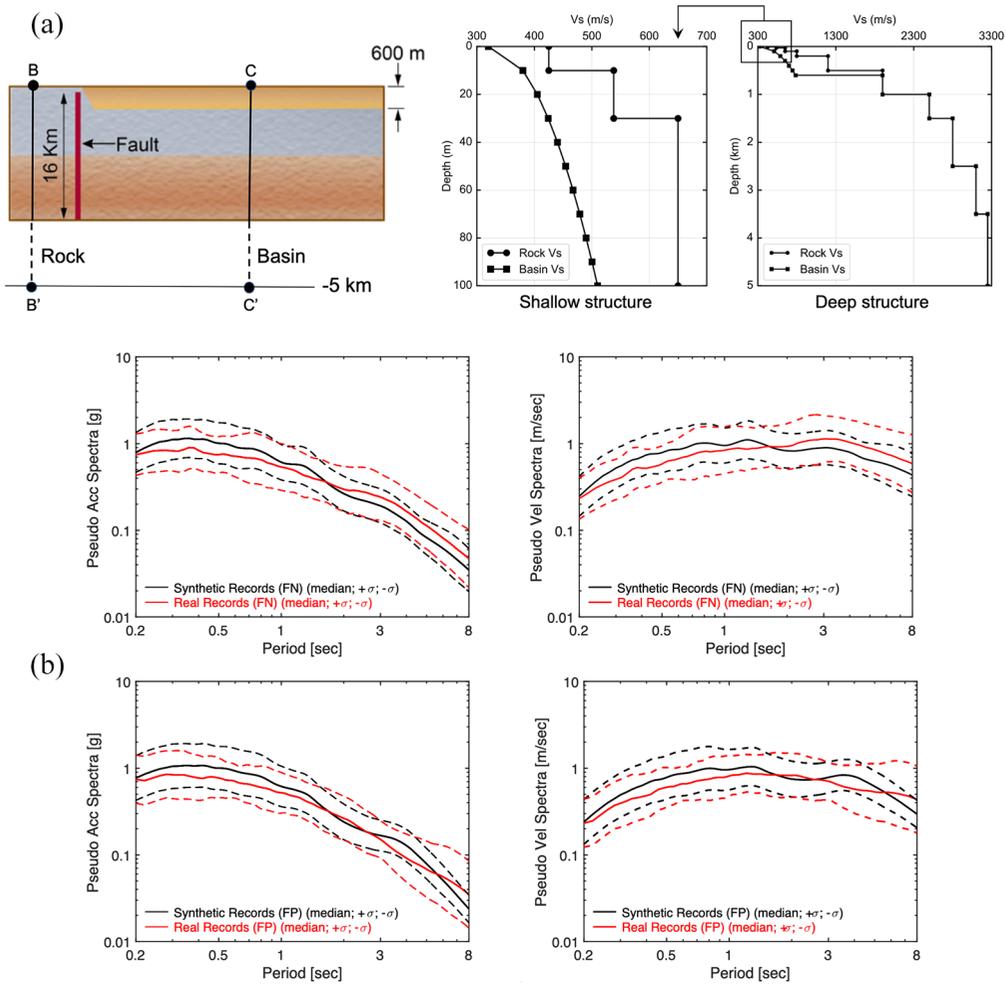


Figure 8. Canonical sedimentary basin with adjacent strike-slip fault used in M7 event fault-to-structure simulations: (a) geologic system and fault location and (b) spectra comparison of 2490 simulated ground motions from three GP rupture realizations with a selected set of 38 measured near-field records for sites within 10 km of the fault.

Source: McCallen et al. (2021b).

fault-to-structure results for a canonical geologic system and a suite of representative buildings have been performed for large-magnitude strike-slip earthquakes as shown in Figure 8 (McCallen et al., 2021b). For evaluation of ground-motion characteristics (Rezaeian et al., 2024), shallow basin response for an M7 strike-slip earthquake was simulated and compared with both empirical ground-motion models and observed historical earthquake waveforms from 38 actual earthquake records with fault and geologic parameters similar to the canonical system. In addition, to evaluate the EQSIM full fault-to-structure workflow structural responses, nonlinear building responses were compared for both simulated and observed earthquake motions for all available sites within 10 km of the fault.

Favorable comparisons between simulated and observed ground motions (Figure 8) and building response (Figure 9) were obtained (McCallen et al., 2021b). In terms of building response, both real records and simulated motions exhibited a log-normal distribution of building drift demand, and median drifts from the set of simulated and observed ground motions were also in good agreement.

Evaluations of simulated ground motions and the GP rupture model generator have been completed through comparison to empirical ground-motion models (Goulet et al., 2015), as well as comparisons with a significant number of observed large earthquakes (Akinci et al., 2024; Bradley et al., 2017; Graves, 2022; Graves and Pitarka, 2010, 2016; Lee et al., 2020, 2022; Pitarka et al., 2017, 2020, 2021a). These thorough evaluations, performed at the amplitude and energy levels of large and intermediate magnitude earthquakes with realistic 3D geologic conditions, relied on direct comparisons between recorded and simulated ground motions on a range of 0–5 Hz. Quantitative evaluations have demonstrated that the GP model is well suited to deterministic broadband simulation techniques. Moreover, being driven by physics-based rupture modeling, in addition to calibration by empirical correlations between rupture parameters as a function of magnitude, the GP model is capable of representation of multi-scale source effects and related near-fault ground-motion variability and amplitude. As more earthquake data have become available, the GP model has gone through refinements, including enhanced representations of large-scale and small-scale slip and slip rate spatial variations that better reproduce the observed frequency content of the generated seismic energy (e.g. Graves and Pitarka, 2015; Pitarka et al., 2021b). These refinements have improved the quality of the simulated near-fault ground-motion variability and permanent displacement (Graves, 2022).

Specific to the SFBR, a number of validation exercises have been completed. While no large Hayward Fault earthquakes have occurred since the last event in 1868, there have been multiple small events that have been observed with regional strong-motion instrumentation. Figure 10, for example, illustrates seven moderate magnitude events ranging from M 3.8 to 4.4 that have occurred along the Hayward Fault. As part of the EQSIM ground-motion evaluations, all seven of these events have been simulated and comparisons between simulated and measured ground motions have been completed at a large number of instrumented sites (Pinilla-Ramos et al., 2024). These comparisons have demonstrated good agreement between simulated and observed motions at many sites, with small Fourier amplitude ratio residuals between 0 and 5 Hz, as summarized in Figure 10. These comparisons, along with the work of others (Thompson, 2018; Wills et al., 2015), also point to specific areas where the SFBR geologic velocity model could benefit from additional refinement, particularly for the soft near-surface layers.

In addition to small event simulations, extensive comparisons have been completed between M7 Hayward Fault event simulated motions and existing empirical ground-motion models (Abrahamson et al., 2014; Boore et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014) and suites of observed earthquake motions with characteristics similar to the simulated earthquakes. The objective of these evaluations was two-fold: comparing ground-motion intensities relevant to the response of built systems (e.g. spectral amplitudes, significant duration, polarization, inter-period correlation) and assessing structural responses as obtained from suites of simulated and recorded time series to identify systematic dependencies of demand parameters on simulation model variables. These comparisons have demonstrated reasonable agreement between simulated and empirical intensity measures (Petrone et al., 2021a, 2021b). Notably, exceptions where the

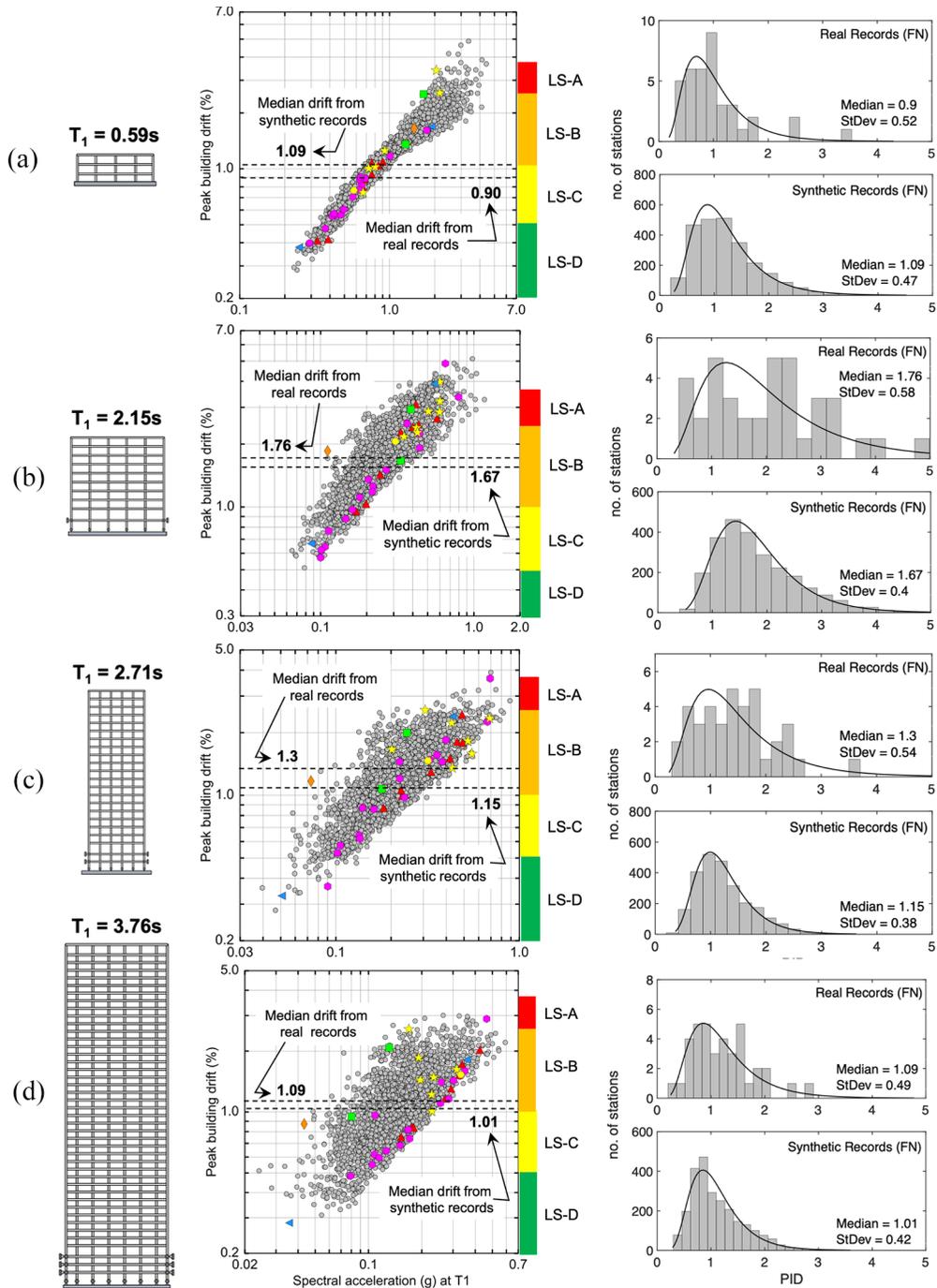


Figure 9. Steel moment frame building Peak Interstory Drift Ratio (PIDR) demand comparisons for 2490 simulated ground motions (gray dots) and 38 real records (color dots) at sites within 10 km of a strike-slip fault for: (a) three-story building, (b) nine-story building, (c) twenty-story building, and (d) forty-story building vertical color bar shows building drift defined limit states from ASCE 43-19. Source: McCallen et al. (2021b).

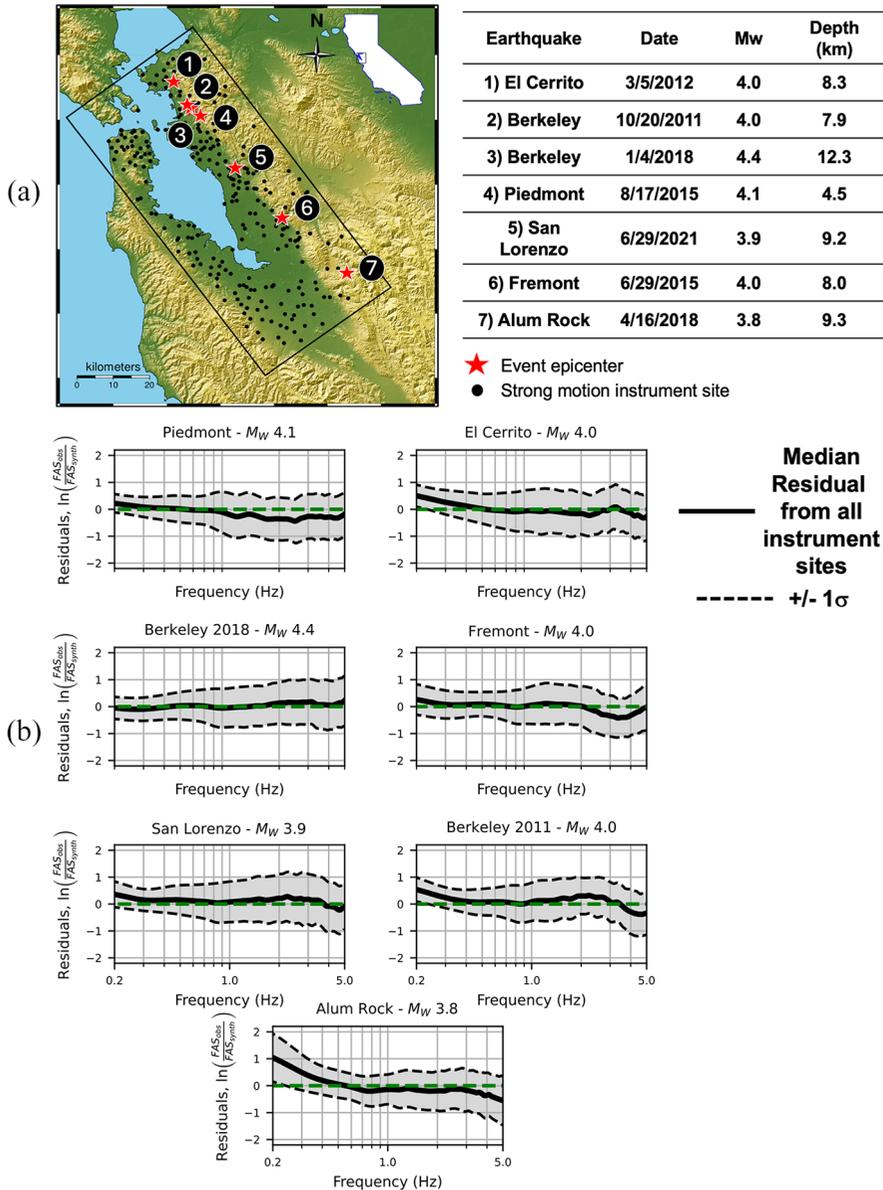


Figure 10. Small events used in simulated/observed ground-motion comparisons: (a) event locations and magnitudes and (b) comparison of within-event Fourier amplitude residuals for horizontal component of ground motion averaged over all recording stations. Source: Pinilla-Ramos et al. (2024).

simulated motions exhibit systematic deviation from the empirical relationships were observed in areas where significant forward and backward rupture directivity effects can result in simulated motions above and below empirical predictions (Petrone et al., 2024). For such cases, further comparisons were carried out with empirical models for forward-

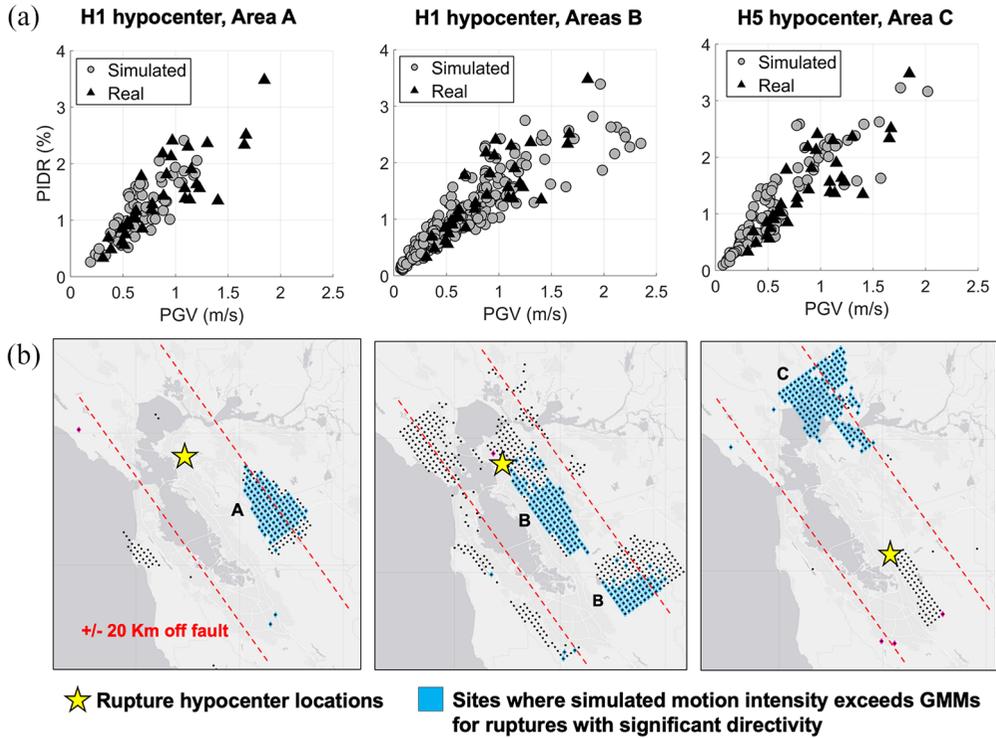


Figure 11. Comparison of the response of a twenty-story building (shown in Figure 9) subjected to simulated and real (observed) motions containing forward directivity: (a) PIDR vs PGV for fault-normal motions and (b) sites within 20 km of the fault where the Hayward Fault simulated motions exhibit directivity effect for far northern and far southern hypocenters.

Source: Petrone et al. (2024).

directivity motions in the near-field (Bray and Rodriguez-Marek, 2004), demonstrating improved agreement of simulated and predicted ground-motion intensities.

In this regard, Figure 11 illustrates the comparison of the response of a twenty-story steel building subject to simulated motions containing near-field forward directivity and observed records from shallow crustal earthquakes identified as containing forward directivity based on the geometric conditions delineated in Somerville et al. (1997) and utilized in the empirical model by Bray and Rodriguez-Marek (2004). The variation of the Peak Interstory Drift Ratio (PIDR) as a function of the Peak Ground Velocity (PGV) of the fault-normal component of the ground motion demonstrates very close agreement between the two sets of motions.

To provide insight on the general regional agreement between simulated and Ground Motion Prediction Equation (GMPE) intensities across the SFBR, Figure 12 illustrates comparisons between simulated motions for all 50 Hayward Fault realizations and the mean of the four established Ground Motion Models (GMMs). Across all distances, the medians of simulations and GMPEs are in reasonable agreement. The standard deviation of the simulations is smaller in amplitude than the GMPE standard deviations; this is expected given that the simulations are based on one scenario earthquake in one regional

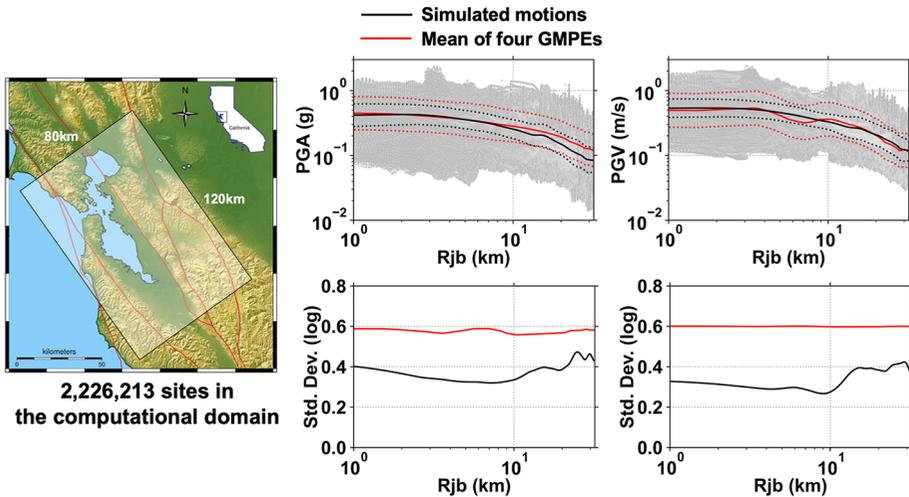


Figure 12. Comparison of RotD50 peak ground acceleration and velocity from Hayward Fault rupture simulations and GMPEs for the full SFBR domain, simulation data from sites at 62.5 m spacing and 50 rupture realizations for a total of 1.11×10^8 grid point acceleration and velocity values, black lines indicate the median and standard deviation of simulation model results, and red lines indicate the median of the mean and standard deviation of four GMPE values.

location, whereas the GMPEs reflect data from multiple earthquakes and multiple regions based on an ergodic process assumption.

Having assessed the realistic character of the simulated motions, even for cases where there are deviations from the prediction of classical empirical models, represented a fundamental step in these validation exercises. It was instrumental in building confidence in the utilization of the simulated motions in engineering applications, particularly for scenarios not adequately represented in the data sets of recorded motions yet highly relevant to risk assessments in regions of high seismicity.

The availability of the SGMD for the larger research and practitioner communities will be an important step toward additional understanding, improvements, and confidence building as additional experts independently utilize, critically evaluate, and provide feedback on the SGMD motions.

SGMD infrastructure—big data management and a framework for future database scalability

An ultimate long-term objective for the SGMD is to continue to expand the database over time with application to multiple geographic regions of interest and to continue to add simulation data sets tailored to future user application interests. In addition, pushing to higher frequency resolutions for engineered systems and model representation of softer near-surface soils should be an ultimate objective. To address the broadest set of earthquake engineering applications, a transition from ground surface data to data throughout a near-surface volume that allows engineers to rigorously include soil–structure interaction and fully account for complex incident body and surface waves is desirable. As regional simulations progress, the engineering application demands will continue to grow in terms

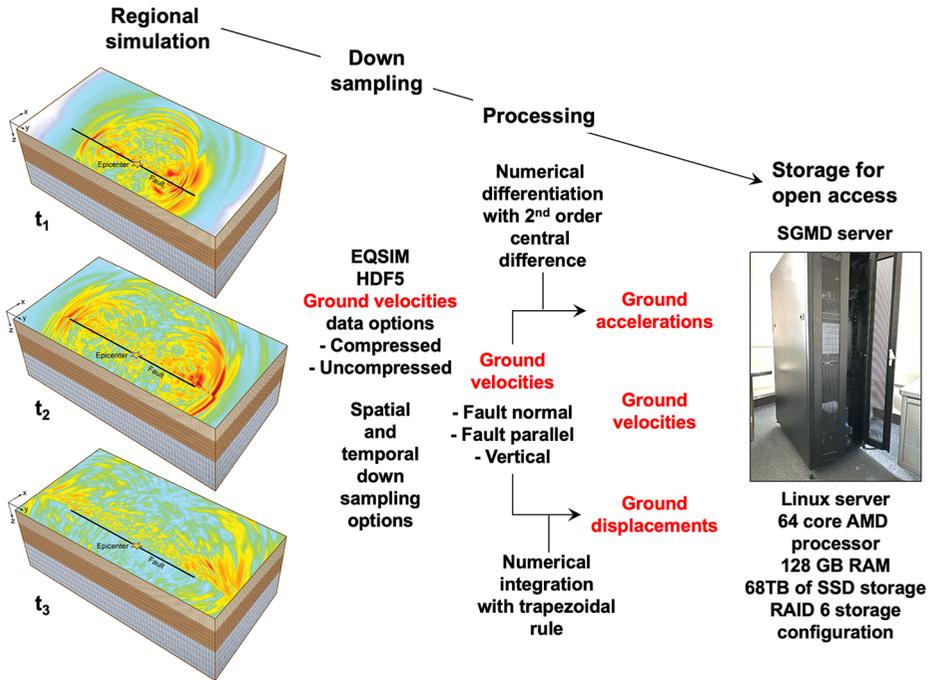


Figure 13. Simulated ground-motion data from generation to archiving.

of data set size and overall data volume. This places a premium on creating a database architecture that is scalable by design, can accommodate a multiplicity of anticipated user applications, and extract the most benefit from high-fidelity, spatially dense ground motions.

The current high-level workflow for generation of simulated motions for the SGMD is shown in Figure 13. Step 1 consists of a regional ground-motion simulation with an option for saving either uncompressed or compressed data sets of ground velocities (Tang et al., 2021). Depending on the specific application of interest, the resulting simulation data set can be downsampled both spatially and temporally, and the fundamental variables consisting of grid point velocities being generated by the SW4 program are numerically differentiated and integrated to get ground accelerations and displacements. The resulting data are then stored on the PEER SGMD server at the University of California Richmond Field Station.

For large-scale simulations, the steps in this workflow employ computational and network technologies developed under the US DoE Exascale Computing Project. The user-selected data compression methodology in the EQSIM framework employs the ZFP software (Diffenderfer et al., 2019; Lindstrom, 2014) to substantially compress data sets based on the elimination of storage of near-duplicate ground-motion numerical values at nearby grid points. In the EQSIM framework, all data from a regional simulation are stored in a single HDF5 data container (Byna et al., 2020) which efficiently includes both ground-motion data in compact binary form as well as relevant metadata in the same container.

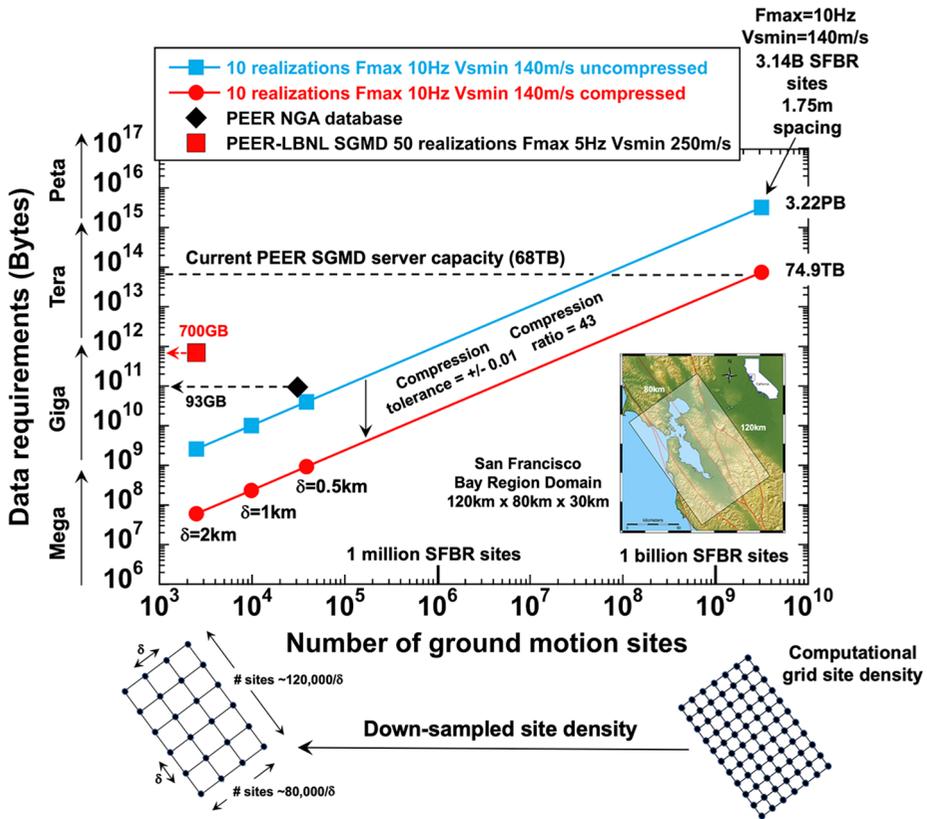


Figure 14. Storage requirements for large, high-fidelity regional simulations.

To provide perspective into the data requirements that could be envisioned in the future, Figure 14 illustrates the data set size that would be associated with ten realizations of an M7 Hayward earthquake with a frequency resolution of 10 Hz and a V_{min} of 140 m/s, resulting in a computational grid spacing of 1.75 m at the ground surface (McCallen et al., 2024b). Without data down-sampling or compression, the ground surface data for the associated 3.14 billion surface grid points in the SFBR model and 90 s of earthquake motion would yield 3.22 PB of data. Application of ZFP data compression can reduce the data set to 74.9 TB in size based on a compression ratio of 43. Data size for spatially down-sampled sites at 0.5, 1.0, and 2 km spacing is shown in the lower left of the plot. Finally, to provide some context, the data required for all of PEER’s existing Next-Generation Attenuation (NGA) database is shown by the black diamond (93 GB) and the data required for the current 50 realizations with sites at 2 km spacing in the SGMD, in ASCII format, is shown by the red rectangle (700 GB). Ideally, it would be desirable to save the surface data at spatial grid points on the order of 2–3 m to allow coupled Soil-Structure Interaction (SSI) simulations whereby the regional geophysics model and local soil–structure model are coupled through the Domain Reduction Method (McCallen et al., 2022b), which, as the number of high-fidelity realizations continues to grow with time, would tax the practical limits of any envisioned data storage system. A final consideration is the ability to have a user interface which is efficient and allows users to interactively browse large data sets with reasonable interactivity latency. A well-designed combined

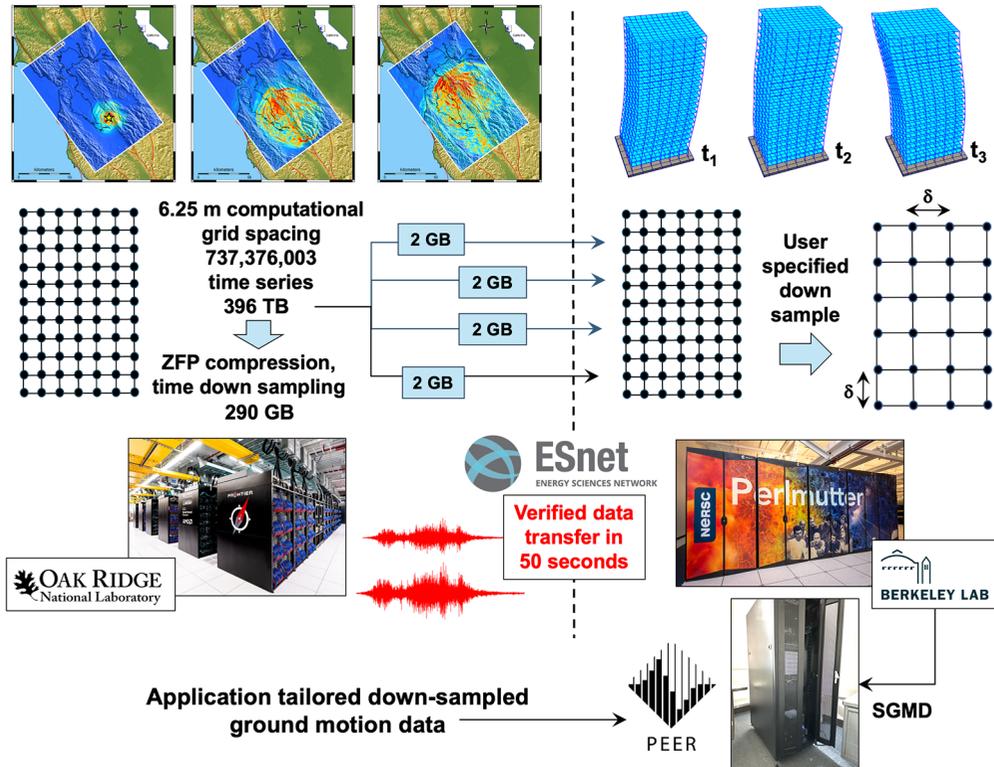


Figure 15. Big data transmission strategy for large-scale simulations within the EQSIM workflow: data partitioned into 2 GB chunks for transfer over the ESnet optical data pipe network; all data files are contained in an HDF5 container; and metrics shown for an F_{max} 5 Hz and V_{smin} 250 m/s regional simulation.

strategy of data compression and application-specific spatial- and temporal-down sampling will be required to extract the full benefit of future simulations for earthquake engineering applications.

In addition to large data storage challenges, it is necessary to have the capacity to move large data sets between supercomputer centers and insert the data into efficient user-interfacing, open-access storage systems. As shown in Figure 15, the data sets associated with the very large runs executed on the Summit and Frontier exaflop computer at ORNL in Tennessee are chunked into 2 GB blocks and transferred across the DOE Energy Sciences Network (ESnet) optical fiber ultrafast data transfer system to Lawrence Berkeley Lab for long-term storage. Subsequently, selectively down-sampled data sets are transferred from Berkeley Lab to the PEER SGMD server. The ESnet data transfer is based on a fiber optic backbone with multiple 100–400 Gbits/s optical channels. As an important data integrity feature, ESnet performs a confirmatory check to ensure the entire large data set was successfully transferred.

For the current Hayward Fault simulations completed for the initial SGMD, two data sets have been created for each Hayward Fault realization. Dense spatial resolution data with temporal-down sampling and data compression are stored at the LBNL NERSC

supercomputing facility, and spatially down-sampled data are transferred to PEER for storage on the SGMD server, as summarized in Table 4.

An efficient and intuitive user interface for data access

For expedient user open access to simulated ground motions, PEER has developed a web-based user interface that allows browsing and selective downloading of simulated ground-motion data (McCallen et al., 2024a). The interface was designed to be simple and intuitive and is accessed through a user-selected three-step process including a user login with PEER-granted credentials (Figure 16), selection of the region of interest (Figure 17c), and data browsing and selection of three-component (fault normal, fault parallel, vertical) simulated data to download to the user's computer system (Figure 17d). Currently, the SFBR is available as a region, and the Greater Los Angeles Area and New Madrid Seismic Zone are two regions that will be included in the future. Finally, a user resource window (Figure 18) can be entered to view the user guide, metadata flat files and technical background papers and information.

Data search parameter options include:

- Selection of individual fault rupture realizations associated with event magnitude, hypocenter location, fault rupture plane geometry, and fault rupture parameters. Each realization represents a unique earthquake in the region.
- Geographical coordinates of a ground surface site of interest in longitude and latitude, which can be a single pair of longitude and latitude coordinates or a range of longitudes and latitudes.
- Site selection based on the time-averaged shear-wave velocity to 30 m depth at the location of the surface site (V_{s30}).
- Site selection based on the distance from the surface site to the fault plane including options for the Joyner–Boore distance—the shortest horizontal distance from the surface site to the vertical projection of the rupture (R_{jb}), the closest distance from a surface site to the fault rupture (R_{rup}), and the horizontal distance from the surface projection of the top edge of the fault rupture to the surface site (R_x).
- Site selection based on site Peak Ground Acceleration, Velocity, and Displacement (RotD50 PGA, RotD50 PGV, RotD50 PGD). The RotD50 metric represents the median across 180 orientations of the horizontal ground motion rotated from 1° to 180° in 1° increments.

Two types of metadata are available for the user, both provided in flat files. The first type is metadata for the rupture realizations, and the second type is metadata for the ground-motion time series. The metadata for the realizations includes three sets of information:

- A general description of the simulated earthquake including the region name, corresponding region code, and the realization number.
- Fault rupture parameters, including fault geometry characterized by the fault name; rupture geometry defined by fault length and width, depth to top of rupture, dip, strike, and rake; earthquake magnitude and hypocenter location; the rupture model utilized in the regional simulations (e.g. the Graves–Pitarka kinematic rupture model); and slip characteristics. A visual graphic that displays the slip, slip-rate and

Table 4. Ground-motion simulation data for one Hayward Fault event realization, F_{max} 5 Hz, V_{smin} 250 m/s, 120 km \times 80 km surface domain

Data set	Ground surface site spacing (number of surface sites)	Time step (number of time steps)	Data set size (HDF5 file)	ESnet data transfer time (number of ground-motion time series)
Spatially dense raw SW4 simulation output ^a	6.25 m (245,792,001)	0.001216 s (73,972)	396 TB	14.6 h (737,376,003)
Down sampled in time and ZFP compressed ^b	6.25 m (245,792,001)	0.0194 s (4,623)	290 GB	50 s (737,376,003)
Down sampled in space, uncompressed ^c	2 km (2301)	0.001216 s (73,972)	2.2 GB	10 s (6903)

SW: seismic wave.

^aThese raw data sets are not permanently stored for all 50 realizations due to size.

^bThese data sets are stored at NERSC at LBNL.

^cThese data sets are stored on the PEER SGMD server.

rise time across the fault rupture is also provided to assist in interpretation (Figure 3).

- Simulation model parameters, which include maximum frequency resolved, minimum shear-wave velocity included in the model (V_{smin}), surface grid spacing that defines the distance between computational nodes in the simulation model, output spacing that provides distance between the down-sampled grid points where ground-motion data are available, and the geologic velocity model utilized in the simulation (e.g. the USGS velocity model for the SFBR).

For each region, the metadata for the ground motions of all realizations are provided in a separate flat file (Figure 19). The metadata includes two sets of information. The first set includes the name, latitude, and longitude coordinates of the grid point, the vertical elevation of the grid point from sea level, the 2D Cartesian coordinates of the grid point in the computational model domain (with X- and Y-axes in the fault normal and parallel directions, respectively), the V_{s30} , and the depth at which the shear-wave velocity reaches 1.0 km/s and 2.5 km/s (Z1.0 and Z2.5) at the location of the grid point. The second set consists of the distance parameters (Rjb, Rrup, and Rx) and peak ground-motion values (RotD50 for PGA, PGV, and PGD).

Site-specific regional ground-motion examples for an M7 Hayward Fault earthquake

A principal objective behind the development of the EQSIM fault-to-structure simulation framework was the creation of the ability to gain increased insight into the complex regional distribution of site-specific earthquake ground motions and corresponding infrastructure earthquake demand and risk. Regional simulations offer the potential to rigorously account for source, path, and site effects within the context of one comprehensive numerical model without the imposition of problem partitioning based on simplifying idealizations between different domains and boundaries. The SFBR SGMD will provide a unique data set allowing researchers and practitioners to explore many aspects of regional

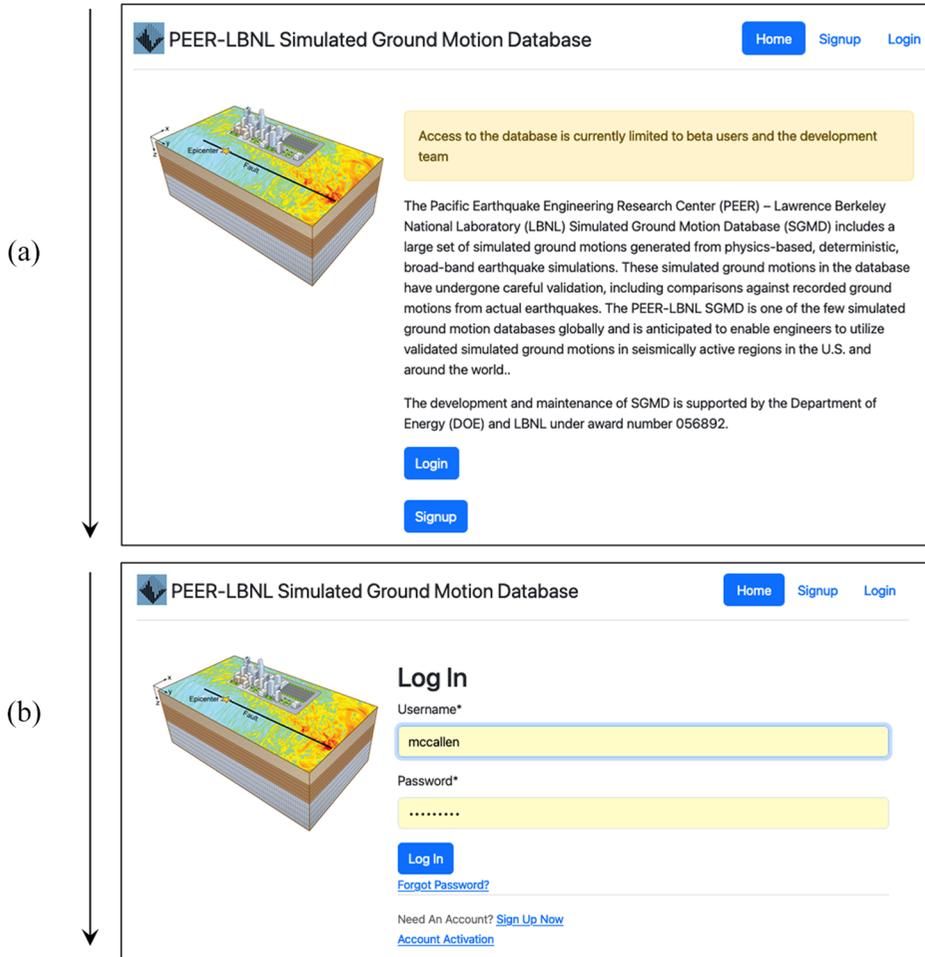


Figure 16. User interface windows for the SGMD: (a) website description landing page and (b) user login window.

ground motions with applications to both earthquake hazard and infrastructure risk assessments. As an example of the information derivable from such data, ground motions for all 50 Hayward Fault rupture realizations for the selected sites shown in Figure 20 are illustrated in Figures 21 and 22. In Figures 21 and 22, the site metadata shows two V_{s30} values from USGS at a subset of sites: the first value is computed from the existing USGS velocity model V21.1, which was used for all EQSIM ground-motion simulations, and a second value that is based on more recent data from (Thompson, 2018). It is noted the V21.1 data has very low near-surface V_s values around the immediate San Francisco Bay margins, and Thompson (2018) appears to be more consistent with available borehole data. It is also noted that the V_{smin} cutoff of 250 m/s utilized in the EQSIM simulations looks more representative of the near-surface sediments based on the more recent data from (Thompson, 2018).

(c)

(d)

Figure 17. User interface windows for the SGMD: (c) region selection window and (d) data selection window.

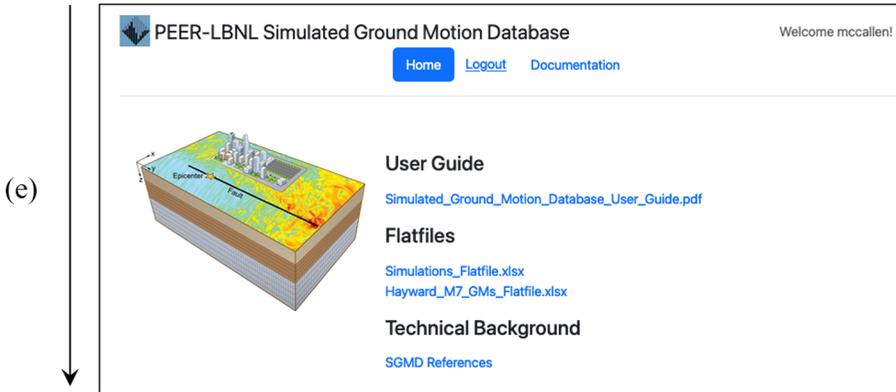


Figure 18. User interface windows for the SFBR database: (e) User resource window.

The site-specific spectra for all 50 realizations are shown in grayscale along with the median and $\pm 1\sigma$ spectra. Significant site-to-site variability of ground-motion amplitude and frequency content is evident, and for selected sites in Oakland, San Leandro, and Fremont (OAK, SLN, FMT), a pronounced site response is observed. These data illustrate intra-event variability with dependency on the rupture characteristics and site location. Although the concept of intra-event variability is known and expected, it can only be systematically quantified and characterized with validated simulated ground motions at regularly spaced grid points as opposed to sparsely and irregularly located instrumented sites from recorded earthquakes. Inter-event variability for five selected Hayward Fault rupture realizations, each of M 7.0, is illustrated in Figures 23 to 25. In these figures, the very complex distribution of ground motions is evident, with significant variability over short spatial separations. The figures also provide insight into the degree to which ground motions from an individual realization can vary substantially from the median ground motion from all 50 realizations. For realization R1, for example, the ground motions in selected areas immediately west of the Hayward Fault exceed the median ground motions by substantial margins. This raises some key questions with respect to risk evaluations for infrastructure systems, if risk evaluations are based on homogenized median simulated ground motion (or median plus some increase factor), there could be a realistic rupture realization that results in substantially larger demand on the infrastructure system. It is noted this type of variation was evident in building simulations with a canonical fault system, as shown in Figure 9. This suggests it could be advisable to explore the tails of the distribution of risk to understand the limits of what damage could occur and to ensure an unacceptable system performance is not reached should the extreme individual earthquake event occur.

In light of the fact that all realizations vary substantially from the median ground motion, considerations of the seismic performance of an interconnected infrastructure network, for example, an energy distribution, water distribution or transportation system, and response evaluations with individual site motions based on median regional motions may provide a potentially misleading picture of the integrated system performance and risk during an actual Hayward Fault event. Evaluating the system performance for a set of rupture realizations can provide a clearer, and more realistic, picture of system resilience. This has implications for PBEE. At a given location, the ground motions from 50

Name	Station				Realization 1										
	Longitude	Latitude	Elevation (m)	Y (m)	Vs30 (m/sec)	z1_0(km)	z2_5(km)	Vs30_sim_model (m/sec)	Rjb (km)	Rx (km)	Rrup (km)	RotD50_FGA (g)	RotD50_PGV (m/sec)	RotD50_PGD (m)	
S 01 01	-122.1836	38.3523	257.5	2000.0	605.0	0.34	3.51	784.4	42.98	35.82	44.58	0.05	0.07	0.04	
S 01 02	-122.2022	38.3417	439.8	2000.0	4000.0	0.23	0.86	784.4	41.43	33.80	42.99	0.03	0.07	0.04	
S 01 03	-122.2207	38.3312	342.2	2000.0	6000.0	0.18	0.87	784.4	39.93	31.80	41.48	0.04	0.06	0.04	
S 01 04	-122.2392	38.3206	57.7	2000.0	8000.0	0.47	0.87	784.4	38.46	29.79	40.03	0.06	0.08	0.04	
S 01 05	-122.2578	38.3101	52.2	2000.0	10000.0	0.27	0.87	784.4	37.05	27.78	38.59	0.05	0.07	0.04	
S 01 06	-122.2763	38.2995	6.2	2000.0	12000.0	0.29	0.87	784.4	35.70	25.77	37.19	0.04	0.08	0.04	
S 01 07	-122.2948	38.2889	7.6	2000.0	14000.0	0.47	2.10	784.4	34.40	23.75	35.85	0.05	0.10	0.05	
S 01 08	-122.3133	38.2784	79.4	2000.0	16000.0	0.13	0.87	784.4	33.19	21.75	34.57	0.04	0.06	0.04	
S 01 09	-122.3318	38.2678	33.9	2000.0	18000.0	0.04	0.87	784.4	32.05	19.74	33.36	0.05	0.07	0.04	
S 01 10	-122.3503	38.2572	41.3	2000.0	20000.0	0.25	0.87	784.4	31.00	17.73	32.22	0.05	0.08	0.05	
S 01 11	-122.3688	38.2466	82.1	2000.0	22000.0	0.38	0.87	784.4	30.04	15.72	31.16	0.04	0.08	0.06	
S 01 12	-122.3873	38.2360	37.7	2000.0	24000.0	0.47	1.13	784.4	29.20	13.71	30.20	0.05	0.10	0.08	
S 01 13	-122.4058	38.2255	2.0	2000.0	26000.0	0.77	1.13	784.4	28.48	11.70	29.36	0.04	0.11	0.09	
S 01 14	-122.4243	38.2149	1.6	2000.0	28000.0	0.47	1.66	784.4	27.87	9.69	28.61	0.05	0.13	0.10	
S 01 15	-122.4427	38.2043	18.4	2000.0	30000.0	0.47	1.57	784.4	27.40	7.69	27.98	0.06	0.14	0.11	
S 01 16	-122.4612	38.1937	201.9	2000.0	32000.0	0.46	1.63	784.4	27.07	5.68	27.47	0.05	0.12	0.10	
S 01 17	-122.4797	38.1831	45.3	2000.0	34000.0	0.09	0.87	784.4	26.89	3.67	27.11	0.03	0.08	0.06	
S 01 18	-122.4981	38.1725	88.1	2000.0	36000.0	0.14	0.87	784.4	26.83	1.66	26.88	0.05	0.10	0.06	
S 01 19	-122.5166	38.1619	2.5	2000.0	38000.0	0.24	0.87	349.5	26.81	-0.35	26.88	0.09	0.12	0.06	
S 01 20	-122.5350	38.1513	0.0	2000.0	40000.0	0.16	0.87	250.0	26.87	-2.35	26.88	0.10	0.15	0.06	
S 01 21	-122.5534	38.1406	0.1	2000.0	42000.0	0.14	0.87	504.8	27.08	-4.36	27.08	0.07	0.12	0.05	
S 01 22	-122.5719	38.1300	56.0	2000.0	44000.0	0.00	0.87	1028.1	27.45	-6.37	27.45	0.05	0.08	0.04	
S 01 23	-122.5903	38.1194	29.4	2000.0	46000.0	0.02	0.87	734.0	27.95	-8.38	27.95	0.06	0.07	0.04	
S 01 24	-122.6087	38.1088	41.0	2000.0	48000.0	0.11	0.87	504.8	28.58	-10.38	28.59	0.10	0.11	0.04	
S 01 25	-122.6271	38.0982	107.8	2000.0	50000.0	0.00	0.87	1028.1	29.34	-12.38	29.34	0.05	0.06	0.04	
S 01 26	-122.6456	38.0875	417.9	2000.0	52000.0	0.00	0.86	1028.1	30.22	-14.40	30.22	0.06	0.07	0.04	
S 01 27	-122.6640	38.0769	92.2	2000.0	54000.0	0.00	0.87	1028.1	31.19	-16.40	31.19	0.06	0.07	0.04	
S 01 28	-122.6824	38.0663	232.7	2000.0	56000.0	0.00	0.87	1028.1	32.26	-18.41	32.26	0.06	0.07	0.03	
S 01 29	-122.7008	38.0557	94.0	2000.0	58000.0	0.00	0.87	1028.1	33.42	-20.41	33.42	0.06	0.06	0.03	
S 01 30	-122.7192	38.0450	237.6	2000.0	60000.0	0.00	0.87	1028.1	34.65	-22.42	34.65	0.06	0.06	0.03	
S 01 31	-122.7375	38.0344	139.7	2000.0	62000.0	0.00	0.87	1028.1	35.95	-24.42	35.95	0.05	0.07	0.03	
S 01 32	-122.7559	38.0237	175.4	2000.0	64000.0	0.00	0.87	1028.1	37.31	-26.43	37.31	0.05	0.06	0.03	
S 01 33	-122.7743	38.0131	150.1	2000.0	66000.0	0.20	0.71	363.0	38.73	-28.43	38.73	0.10	0.12	0.04	
S 01 34	-122.7927	38.0024	135.2	2000.0	68000.0	0.41	0.71	504.8	40.20	-30.44	40.20	0.07	0.11	0.04	

Figure 19. Metadata and ground-motion intensities flat file for site simulated ground motions, realization R1 shown.

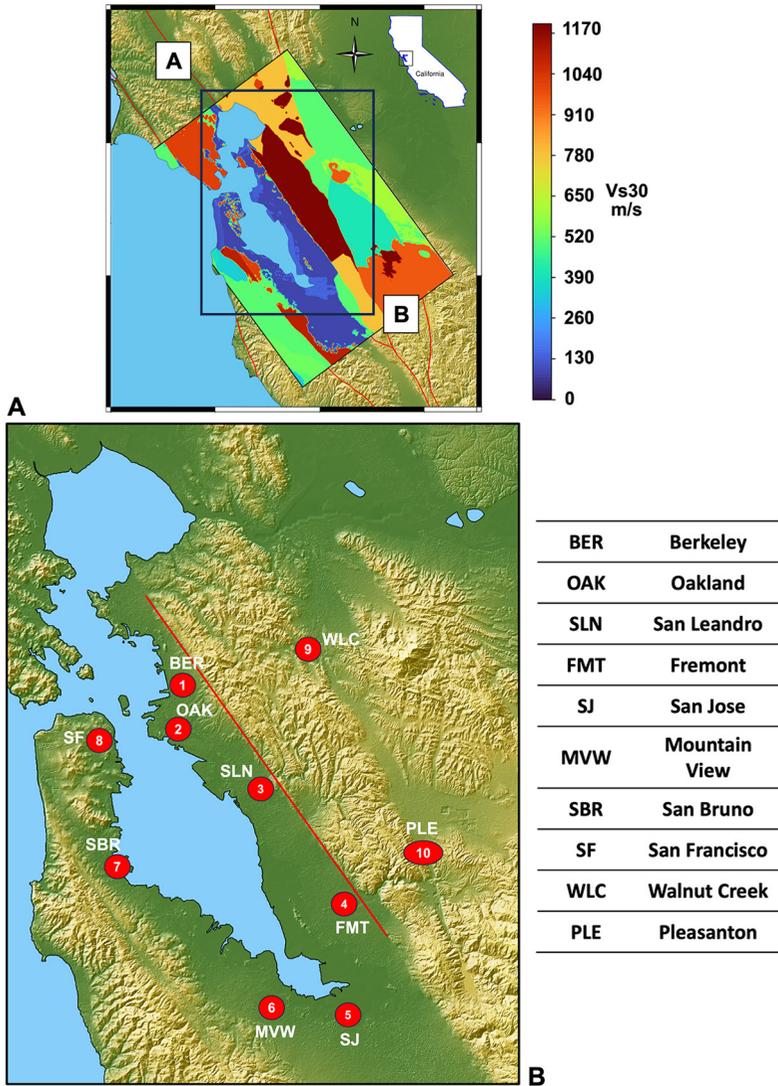


Figure 20. Map of USGS Vs30 and ground-motion sites, for example simulated motions.

realizations can be directly employed in the structural analysis phase for utilization in damage and consequence analyses. A study that performs and quantifies this comparison at multiple locations is not part of this data paper; however, this study would be important and is in the planning stages. The SGMD will provide the earthquake community with a new resource and toolset for exploring precisely these types of questions.

Summary and conclusions

This article describes significant progress toward the development and construction of a regional-scale simulated earthquake ground-motion database for earthquake engineering applications. The overall effort required multiple achievements including executing a large

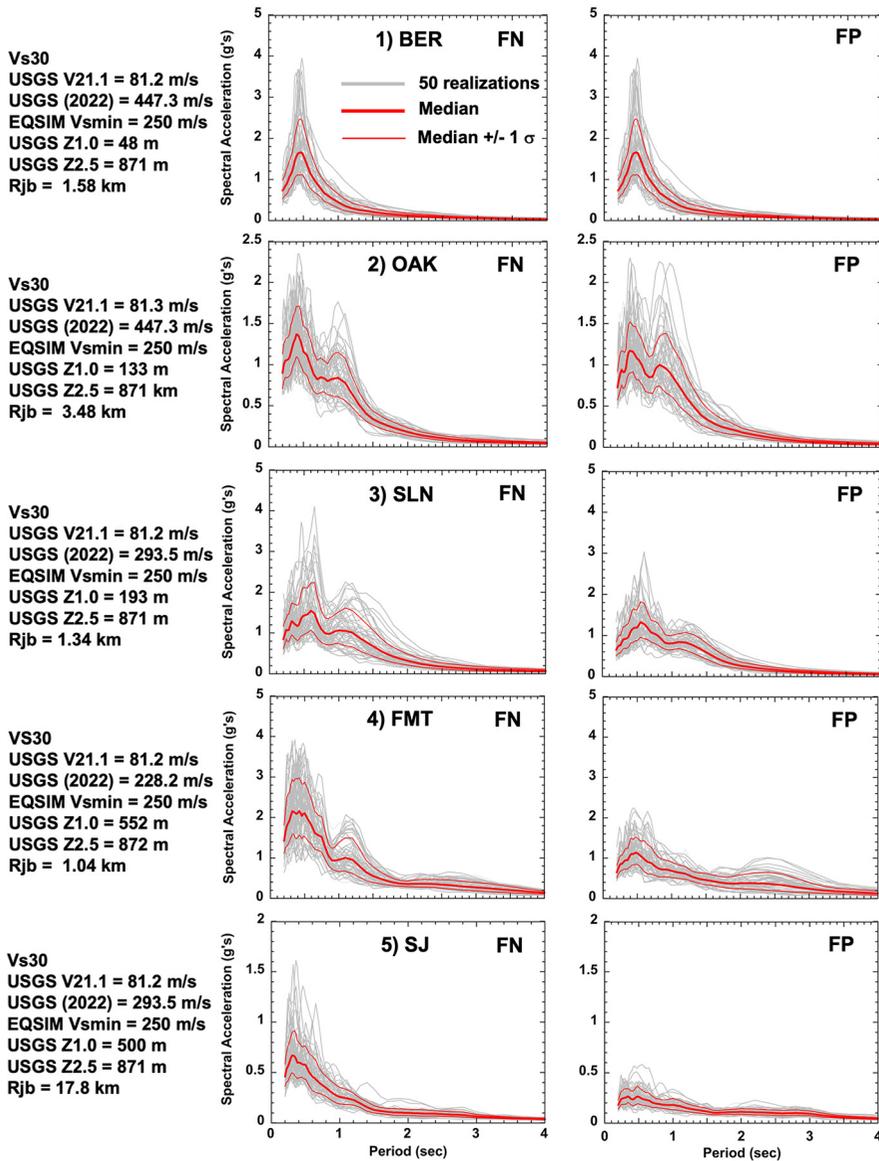


Figure 21. Fault-normal and fault-parallel response spectra from simulated motions at sites 1–5 in Figure 20 (5% damping).

number of massively parallel regional earthquake simulations, software verification and simulation results validation, the development of a schema for transfer and storage of big data, and the creation of a practical web-based user interface for accessing big data. For fifty 5 Hz resolution simulations of the SFBR, the database currently consists of 1,021,500 uncompressed ASCII time series at 2 km spacing stored on the PEER SGMD open-access server, backed by 36,868,800,150 compressed HDF5 time series at 6.25 m spacing stored at the National Energy Research Scientific Computing Center at LBNL. In combination, these developments have yielded a practical and accessible database. The SGMD is ready

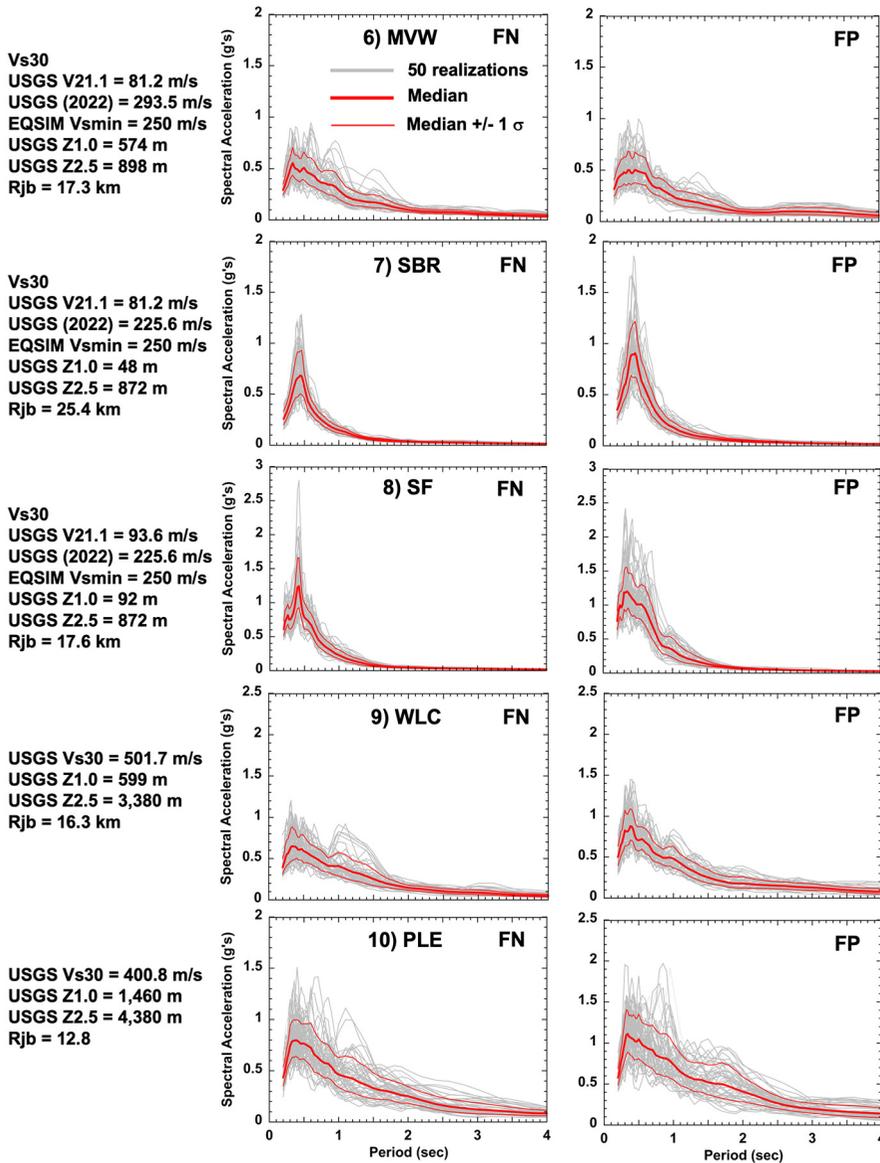


Figure 22. Fault-normal and fault-parallel response spectra from simulated motions at sites 6–10 in Figure 20 (5% damping).

for open access and community use, and we look forward to feedback and recommendations for future refinements.

Current practice in seismic hazard assessment relies heavily on empirical GMMs for developing hazard maps and designing structures. As a result of the sparsity of ground-motion observations, an ergodic assumption is typically invoked and ground motions from other regions are incorporated to augment data for the region of interest. Recognizing earthquake motions from other regions are in general not fully translatable, recent efforts

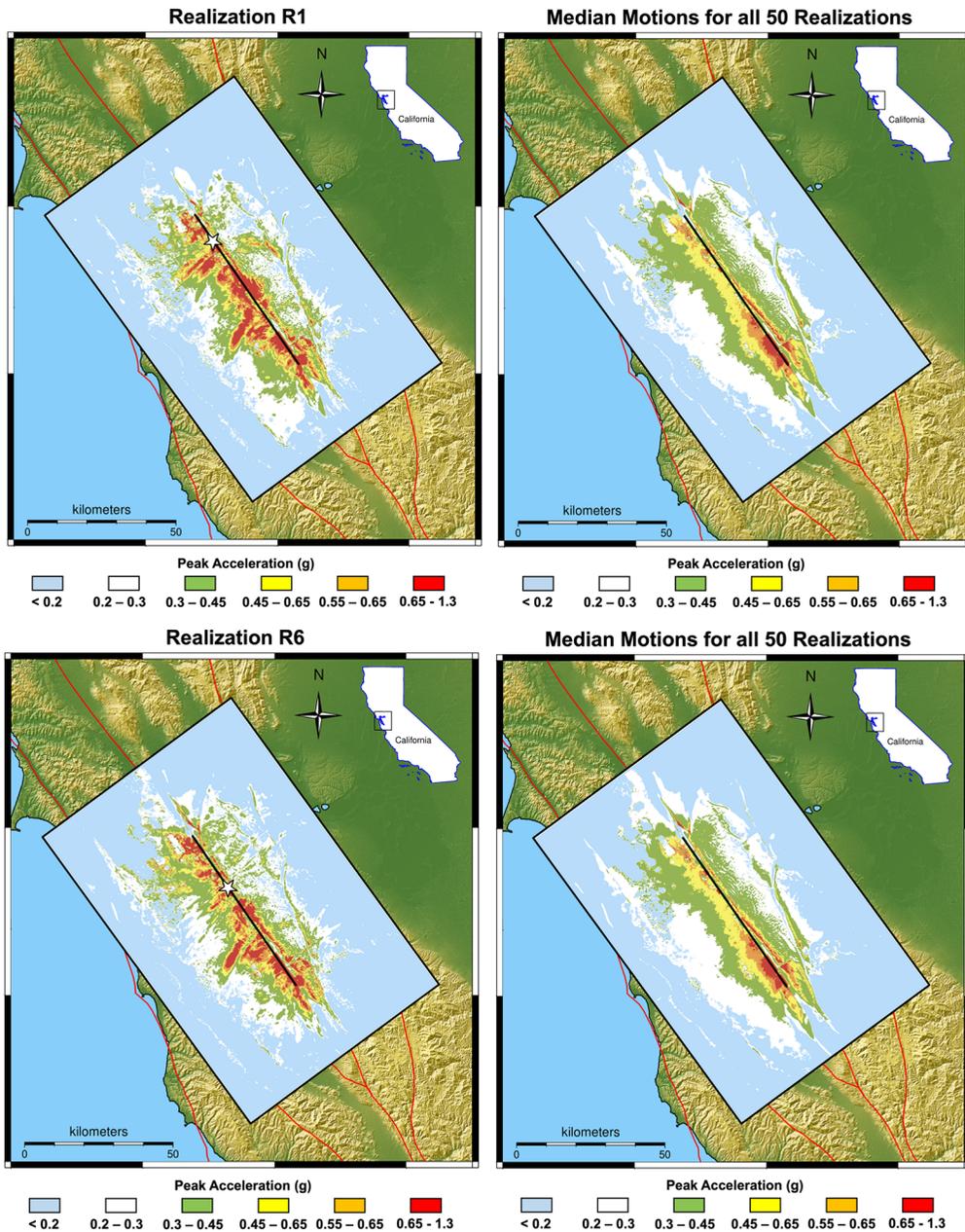


Figure 23. Site-dependent ground motions and inter- and intra-event variability: peak ground acceleration contours for R1 and R6 rupture realizations compared to the median accelerations for all 50 rupture realizations (0–5 Hz simulation resolution).

have focused on moving toward non-ergodic, region-specific GMMs (Lavrentiadis and Abrahamson, 2023). However, ergodic GMMs remain the standard approach.

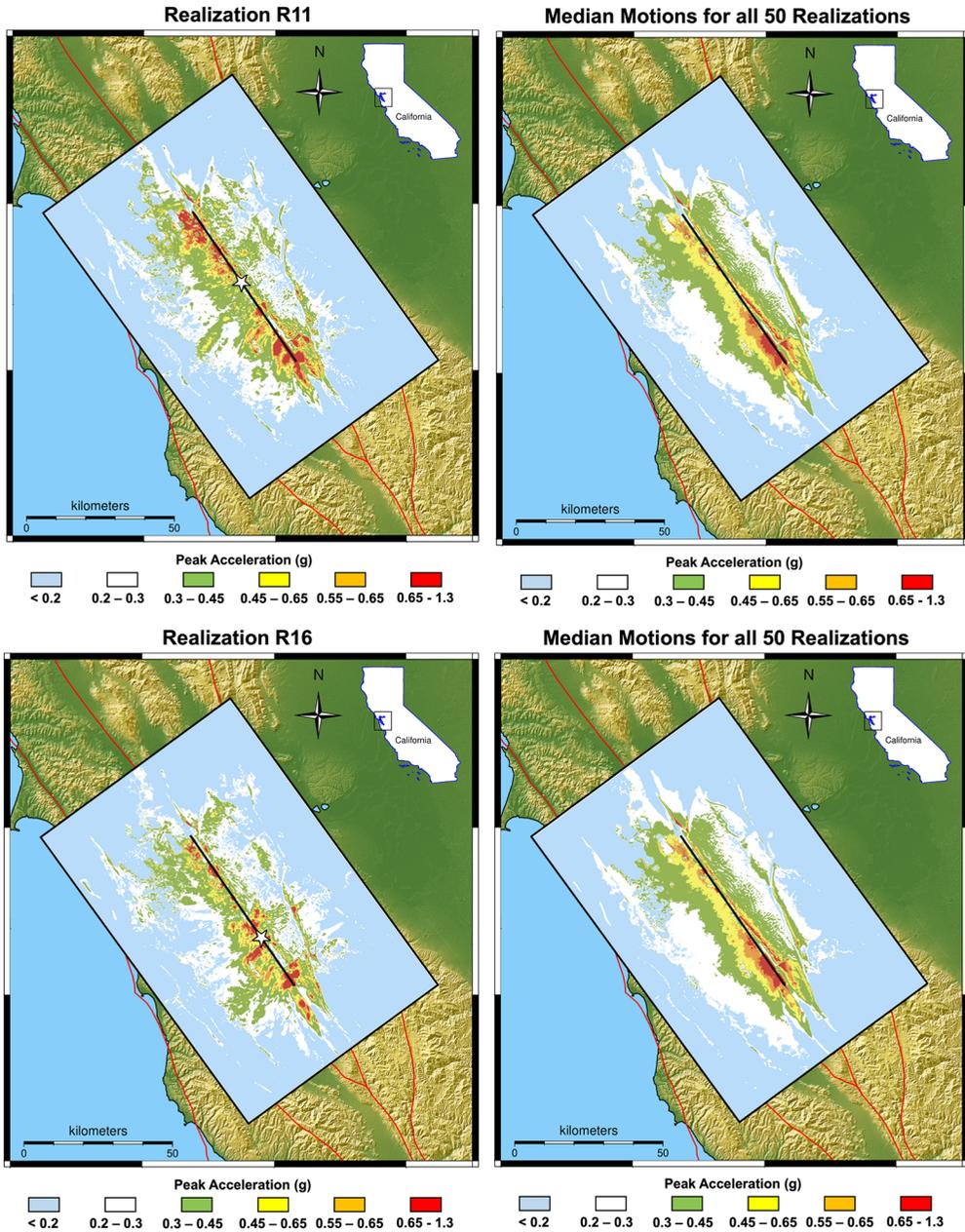


Figure 24. Site-dependent ground motions and inter- and intra-event variability: peak ground acceleration contours for R11 and R16 rupture realizations compared to the median accelerations for all 50 rupture realizations (0–5 Hz simulation resolution).

The important influence of fault rupture directivity on structural demand and risk has come into clearer focus in recent years. Although some GMMs include directivity terms (e.g. Bray and Rodriguez-Marek, 2004), widely used state-of-the-art models such as the NGA-West2 GMMs (Abrahamson et al., 2014; Boore et al., 2014; Campbell and

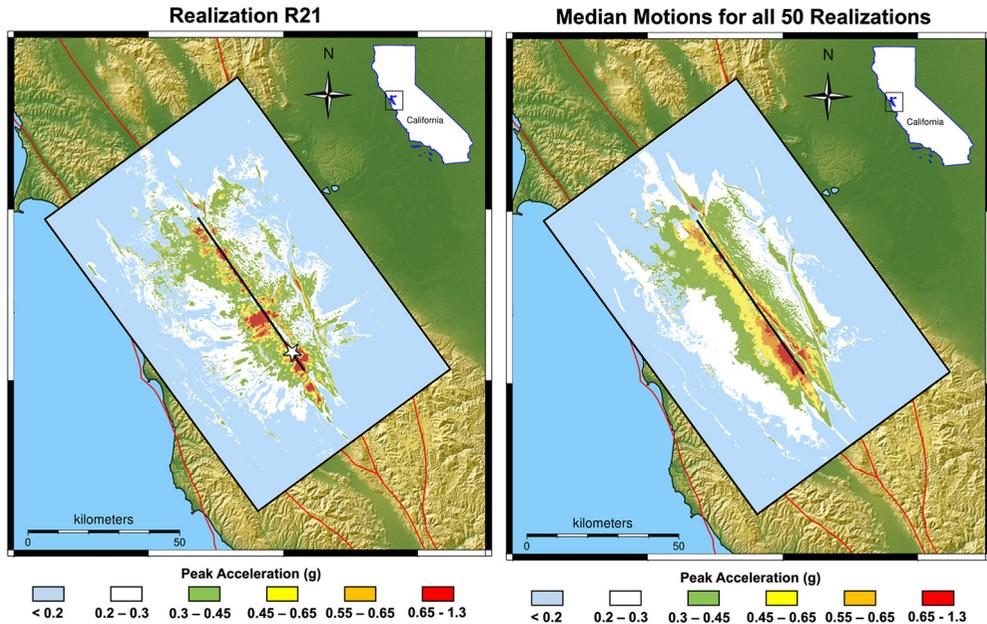


Figure 25. Site-dependent ground motions and inter- and intra-event variability: peak ground acceleration contours for the R21 rupture realization compared to the median accelerations for all 50 rupture realizations (0–5 Hz simulation resolution).

Bozorgnia, 2014; Chiou and Youngs, 2014), employed in both the 2023 USGS National Seismic Hazard Model (NSHM23; Moschetti et al., 2024) and USGS ShakeMaps (Wald et al., 2022), generally do not explicitly incorporate these terms or do so in an average sense. Directivity effects remain insufficiently understood and integrated into seismic hazard models, largely due to a lack of near-fault data from large-magnitude earthquakes (Withers et al., 2024).

Physics-based simulations hold the potential to complement these efforts toward improving GMMs by providing region-specific ground-motion time series and corresponding intensity measures and providing deeper insight into the factors that most influence ground motions. The validated ground motions in the SGMD offer a transformative opportunity to begin addressing such gaps. This data set provides robust information for updating GMMs with critical missing features, such as directivity and site effects, and enables the development of AI-based models that estimate ground-motion intensities with reduced uncertainty.

In addition, SGMD motions can allow for more realistic, site-dependent evaluation of structural responses, particularly when multi-component waveform time series are required for nonlinear analyses. Unlike the ergodic assumptions inherent to recorded motions, the SGMD includes a wide variety of simulated motions (currently, 50 realizations per site) that account for regional source, path, and site effects and can help inform site-specific variability. When coupled with detailed structural models and damage functions, these motions can improve estimates of Engineering Demand Parameters (EDPs) and expected damage (Matinrad and Petrone, 2023). Furthermore, the diversity of SGMD motions supports the training of AI models capable of accurately predicting structural responses under varied seismic excitations (Muin and Mosalam, 2021). In terms of risk assessments for

regionally distributed, interconnected infrastructure systems, simulated motions can define appropriate time phasing and coherency of earthquake shaking at all system component sites, allowing better insight into the overall system response (Taslimi and Petrone, 2024).

In summary, the SGMD data set can not only support the refinement of GMMs but also facilitate innovative approaches to seismic hazard assessment and structural response modeling and be particularly enabling for region-wide hazard and risk assessments. By leveraging these simulations, the field can begin to address existing knowledge gaps, reduce uncertainties, and better prepare for the seismic challenges of the future. As correctly noted by participants at the PEER International Pacific Rim Forum in 2021, getting simulated motion data into the hands of the broad engineering, scientific, and emergency response communities will be essential to advancing this emerging technology. Expert SGMD consumers can identify compelling use cases, as well as define best practice approaches for addressing key issues such as nonlinear response in near-surface soils (Kuncar et al., 2024), region-dependent uncertainties in geologic data, and the exploitation of rapidly advancing data analytics methodologies to extract the most useful information from very large data sets.

Acknowledgments

The authors thank Dr. Grigorios Lavrentiadis and Dr. Elnaz Seylabi for discussions related to recent information on San Francisco Bay Region near-surface velocity data. Arben Pitarka's work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by the US DOE Office of Cyber Security, Energy Security and Emergency Response (CESER) toward ensuring resilience of US energy infrastructure; this support and the encouragement of CESER Program Manager Joseph Dygert are gratefully acknowledged. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a Department of Energy Office of Science User Facility, and resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. Access, bank allocations, and exceptional user support at both facilities were essential to the successful completion of this work. Gabriel Vargas supported the design and configuration of the SGMD server and Amarnath Kasalanati organized the PEER Pacific Rim International Forum from which the SGMD concepts emerged.

ORCID iDs

David McCallen  <https://orcid.org/0000-0003-0386-0324>

Arben Pitarka  <https://orcid.org/0000-0001-6026-8859>

Floriana Petrone  <https://orcid.org/0000-0003-4545-6303>

Claudio Perez  <https://orcid.org/0000-0001-8156-2651>

Research data availability

The SFBR Hayward Fault event ground-motion data are available through the Simulated Ground Motion Database on the PEER website (McCallen et al., 2024a). Initially, a set of 15 expert beta users are being provided access to test the ground motions and the user interface. Based on feedback from these users, the intent is to open the SGMD to the broad community in April 2025.

Supplemental material

Supplemental material for this article is available online.

Note

1. <https://github.com/geodynamics/sw4/tree/master/pytest>.

References

- Aagaard BT and Hirakawa ET (2021) San Francisco Bay region 3D seismic velocity model v21.1. US Geological Survey data release. Available at: <https://doi.org/10.5066/P9TRDCHE> (accessed 2024).
- Abrahamson NA, Silva WJ and Kamai R (2014) Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra* 30(3): 1025–1055.
- Akinci A, Pitarka A, Harris PA, De Gori P and Buttinelli M (2024) Impact of the earthquake rupture on ground-motion variability of the 24 August 2016 M_w 6.2 Amatrice, Italy, Earthquake. *Bulletin of the Seismological Society of America* 114(6): 2823–2845.
- Akkar S and Özen Ö (2005) Effect of peak ground velocity on deformation demands for SDOF systems. *Earthquake Engineering & Structural Dynamics* 34(13): 1551–1571.
- American Society of Civil Engineers (ASCE)/SEI 7-22 (2021) Minimum design loads and associated criteria for buildings and other structures.
- Beckingsale DA, Burmark J, Hornung R, Jones H, Killian W and Kunen AJ (2019) RAJA: Portable performance for large-scale scientific applications. In: *Proceedings of the 2019 IEEE/ACM international workshop on performance, portability and productivity in HPC (P3HPC)*, Denver, CO, 22 November, pp. 71–81. New York: IEEE.
- Boore DM, Stewart JP, Seyhan E and Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra* 30(3): 1057–1085.
- Bradley BA, Razafindrakoto H and Polak V (2017) Ground-Motion Observations from the 14 November 2016 M_w 7.8 Kaikoura, New Zealand, Earthquake and Insights from Broadband Simulations. *Seismological Research Letters* 88(3): 740–756.
- Bray JD and Rodriguez-Marek A (2004) Characterization of forward-directivity ground motions in the near-fault region. *Soil Dynamics and Earthquake Engineering* 24(11): 815–828.
- Byna S, Breitenfeld MS, Dong B, Koziol Q, Pourmal E, Robinson D, Soumagne J, Tang H, Vishwanath V and Warren R (2020) ExaHDF5: Delivering efficient parallel I/O on exascale computing systems. *Journal of Computer Science and Technology* 35: 145–160.
- Campbell KW 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): 1087–1114.
- Chiou BS-J and Youngs RR (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.
- Diffenderfer J, Fox AL, Hittinger J, Sanders G and Lindstrom P (2019) Error analysis of ZFP compression for floating-point data. *SIAM Journal on Scientific Computing* 41(3): A1867–A1898.

- European Standard EN 1998-1:2004 (2004) Eurocode 8: Design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings (Brussels: Comité Européen de Normalisation (CEN)).
- Goulet CA, Abrahamson NA, Somerville PG and Wooddell KE (2015) The SCEC broadband platform validation exercise: Methodology for code validation in the context of seismic-hazard analyses. *Seismological Research Letters* 86: 17–26.
- Graves RW (2022) Using a grid-search approach to validate the Graves–Pitarka broadband simulation method. *Earth, Planets and Space* 74: 186.
- Graves RW and Day SM (2003) Stability and accuracy analysis of coarse-grain viscoelastic simulations. *Bulletin of the Seismological Society of America* 93(1): 283–300.
- Graves RW and Pitarka A (2010) Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America* 100(5A): 2095–2123.
- Graves RW and Pitarka A (2015) Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method. *Seismological Research Letters* 86(1): 75–80.
- Graves RW 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 RW, Jordan TH, Callaghan S, Deelman E, Field E, Juve G, Kesselman C, Maechling P, Mehta G, Milner K, Okaya D, Small P and Vahi K (2011) CyberShake: A physics-based seismic hazard model for southern California. *Pure and Applied Geophysics* 168: 367–381.
- Günay S and Mosalam KM (2013) PEER performance-based earthquake engineering methodology, revisited. *Journal of Earthquake Engineering* 17(6): 829–858.
- Irikura K and Miyake H (2011) Recipe for predicting strong ground motion from crustal earthquake scenarios. *Pure and Applied Geophysics* 168: 85–104.
- Kuncar F, Bradley BA, De la Torre CA, Rodriguez-Marek A, Zhu C and Lee RL (2024) Methods to account for shallow site effects in hybrid broadband ground-motion simulations. *Earthquake Spectra* 41: 1272–1313.
- Lavrentiadis G and Abrahamson NA (2023) A non-ergodic spectral acceleration ground motion model for California developed with random vibration theory. *Bulletin of Earthquake Engineering* 21: 5265–5291.
- Lee RL, Bradley BA, Stafford PJ, Graves RW and Rodriguez-Marek A (2020) Hybrid broadband ground motion simulation validation of small magnitude earthquakes in Canterbury, New Zealand. *Earthquake Spectra* 36(2): 673–699.
- Lee RL, Bradley BA, Stafford PJ, Graves RW and Rodriguez-Marek A (2022) Hybrid broadband ground-motion simulation validation of small magnitude active shallow crustal earthquakes in New Zealand. *Earthquake Spectra* 38(4): 2548–2579.
- Lindstrom P (2014) Fixed-rate compressed floating-point arrays. *IEEE Transactions on Visualization and Computer Graphics* 20(12): 2674–2683.
- Liu P, Archuleta RJ and Hartzell SH (2006) Prediction of broadband ground-motion time histories: Hybrid low/high-frequency method with correlated random source parameters. *Bulletin of the Seismological Society of America* 96(6): 2118–2130.
- McCallen D, Petersson NA, Rodgers A, Pitarka A, Miah M, Petrone F, Sjogreen B, Abrahamson NA 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, Petrone F, Miah M, Pitarka A, Rodgers A and Abrahamson NA (2021b) EQSIM—A multidisciplinary framework for fault-to-structure earthquake simulations on exascale computers, Part II: Regional simulations of building response. *Earthquake Spectra* 37: 736–761.
- McCallen D, Petrone F, Seylabi EE, Pitarka A, Abrahamson NA and Elfass S (2022a) *The PEER International Pacific Rim Forum 2021: Regional-scale simulation of earthquake ground motions and infrastructure response for performance-based earthquake engineering*. PEER report no. 2022/04. Berkeley, CA: Pacific Earthquake Engineering Research (PEER) Center, University of California, Berkeley. Available at: https://apps.peer.berkeley.edu/publications/peer_reports/reports_2022/PEERReportNo2022_04.pdf (accessed 2024).

- McCallen D, Pitarka A, Tang H, Nakata R, Mosalam KM, Petrone F, Gunay S and Perez C (2024a) *The PEER-LBNL Simulated Ground Motion Database*. Available at: <https://doi.org/10.55461/SGMD12232024> (accessed 2024).
- McCallen D, Pitarka A, Tang H, Pankajakshan R, Petersson NA, Miah M and Huang J (2024b) Regional-scale fault-to-structure earthquake simulations with the EQSIM framework: Workflow maturation and computational performance on GPU-accelerated exascale platforms. *Earthquake Spectra* 40(3): 1615–1652.
- McCallen D, Tang H, Wu S, Eckert E, Huang J and Petersson NA (2022b) Coupling of regional geophysics and local soil-structure models in the EQSIM fault-to-structure earthquake simulation framework. *The International Journal of High-Performance Computing Applications* 36(1): 78–92.
- Maechling P, Silva F, Callaghan S and Jordan TH (2015) SCEC broadband platform: System architecture and software implementation. *Seismological Research Letters* 86(1): 27–38.
- Matinrad P and Petrone F (2023) ASCE/SEI 7-compliant site-specific evaluation of the seismic demand posed to reinforced concrete buildings with real and simulated ground motions. *Earthquake Engineering & Structural Dynamics* 52(15): 4987–5009.
- Moschetti MP, Aagaard BT, Ahdi SK, Altekruze J, Boyd OS, Frankel AD, Herrick J, Petersen MD, Powers PM, Rezaeian S, Shumway AM, Smith JA, Stephenson WJ, Thompson EM and Withers KB (2024) The 2023 US National Seismic Hazard Model: Ground-motion characterization for the conterminous United States. *Earthquake Spectra* 40(2): 1158–1190.
- Muin S and Mosalam KM (2021) Structural health monitoring using machine learning and cumulative absolute velocity features. *Applied Sciences* 11(12): 5727.
- Paolucci R, Infantino M, Mazzieri I, Özcebe AG, Smerzini C and Stupazzini M (2018) 3D physics-based numerical simulations: Advantages and current limitations of a new frontier to earthquake ground motion prediction. The Istanbul case study. In: Ptilakis K (ed.) *Recent Advances in Earthquake Engineering in Europe* (Proceedings of the 16th European conference on earthquake engineering (ECEE 2018), Thessaloniki). Cham: Springer International Publishing, pp. 203–223.
- 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.
- Petersson NA and Sjögreen B (2010) Stable grid refinement and singular source discretization for seismic wave simulations. *Communications in Computational Physics* 8: 1074–1110.
- Petersson NA and Sjögreen B (2012) Stable and efficient modeling of anelastic attenuation in seismic wave propagation. *Communications in Computational Physics* 12: 193–225.
- Petersson NA and Sjögreen B (2014) Super-grid modeling of the elastic wave equation in semi-bounded domains. *Communications in Computational Physics* 16: 913–955.
- Petersson NA and Sjögreen B (2015) Wave propagation in anisotropic elastic materials and curvilinear coordinates using a summation-by-parts finite difference method. *Journal of Computational Physics* 299: 820–841.
- Petrone F, Abrahamson NA, McCallen D and Miah M (2021a) Validation of (not-historical) large-event near-fault ground-motion simulations for use in civil engineering applications. *Earthquake Engineering & Structural Dynamics* 50: 116–134.
- Petrone F, Abrahamson NA, McCallen D, Pitarka A and Rodgers A (2021b) Engineering evaluation of the EQSIM simulated ground-motion database: The San Francisco Bay Area region. *Earthquake Engineering & Structural Dynamics* 50(15): 3939–3961.
- Petrone F, Taslimi A, Nia MM, McCallen D and Pitarka A (2024) Ground-motions site and event specificity: Insights from assessing a suite of simulated ground motions in the San Francisco Bay Area. *Earthquake Spectra* 40(4): 2712–2736.
- Pinilla-Ramos C, Pitarka A, McCallen D and Nakata R (2024) Performance evaluation of the USGS velocity model for the San Francisco Bay Area. *Earthquake Spectra* 41: 457–494.
- Pitarka A, Akinci A, De Gori P and Buttinelli M (2021a) Deterministic 3D ground-motion simulations (0–5 Hz) and surface topography effects of the 30 October 2016 M_w 6.5 Norcia, Italy, Earthquake. *Bulletin of the Seismological Society of America* 112: 262–286.

- Pitarka A, Graves RW, Irikura K, Miyake H and Rodgers A (2017) Performance of Irikura recipe rupture model generator in earthquake ground motion simulations with Graves and Pitarka hybrid approach. *Pure and Applied Geophysics* 174: 3537–3555.
- Pitarka A, Graves RW, Irikura K, Miyakoshi K and Rodgers A (2020) Kinematic rupture modeling of ground motion from the M7 Kumamoto, Japan earthquake. *Pure and Applied Geophysics* 177: 2199–2221.
- Pitarka A, Graves RW, Irikura K, Miyakoshi K, Wu C, Kawase H, Rodgers A and McCallen D (2021b) Refinements to the Graves–Pitarka kinematic rupture generator, including a dynamically consistent slip-rate function, applied to the 2019 M_w 7.1 Ridgecrest earthquake. *Bulletin of the Seismological Society of America* 112: 287–306.
- Rezaeian S, Stewart JP, Luco N and Goulet CA (2024) Findings from a decade of ground motion simulation validation research and a path forward. *Earthquake Spectra* 40(1): 346–378.
- Rodgers A, Pitarka A, Pankajakshan R, Sjögreen B and Petersson NA (2020) Regional-scale 3D ground-motion simulations of Mw 7 earthquakes on the Hayward fault, Northern California resolving frequencies 0–10 Hz and including site-response corrections. *Bulletin of the Seismological Society of America* 110: 2862–2881.
- Smerzini C, Amendola C, Paolucci R and Bazrafshan A (2024) Engineering validation of BB-SPEEDset, a data set of near-source physics-based simulated accelerograms. *Earthquake Spectra* 40(1): 420–445.
- Somerville PG, Smith NF, Graves RW and Abrahamson NA (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.
- Standards New Zealand (NZS) 1170.5:2004 (2004) Structural design actions. Part 5: Earthquake actions–New Zealand incorporating amendment 1, 2016 (NZS, Wellington, New Zealand, 76 pp).
- Stewart JP, Chiou BS-J, Bray JD, Graves RW, Somerville PG and Abrahamson NA (2001) *Ground motion evaluation procedures for performance-based design*. PEER report no. 2001-09. Berkeley, CA: Pacific Earthquake Engineering Research (PEER) Center, University of California, Berkeley. Available at: https://peer.berkeley.edu/sites/default/files/0109_j._stewart_s._chiou_j._bray_r._graves_p._somerville_n._abrahamson.pdf (accessed 2024).
- Taborda R and Bielak J (2011) Large-scale earthquake simulation: Computational seismology and complex engineering systems. *Computing in Science & Engineering* 13(4): 14–27.
- Tang H, Byna S, Petersson NA and McCallen D (2021) Tuning parallel data compression and I/O for large-scale earthquake simulation. In: *Proceedings of the 2021 IEEE international conference on big data*, Orlando, FL, 15–18 December, pp. 2992–2997. New York: IEEE.
- Taslimi A and Petrone F (2024) Vulnerability of suspension bridges to spatially variable vertical ground motions. *Earthquake Spectra* 40(4): 2737–2760.
- Thompson EM (2018) An updated Vs30 Map for California with geologic and topographic constraints (version 2.0, July 2022). US Geological Survey data release. Available at: <https://doi.org/10.5066/F7JQ108S> (accessed 2024).
- Wald DJ, Worden CB, Thompson EM and Hearne M (2022) ShakeMap operations, policies, and procedures. *Earthquake Spectra* 38(1): 756–777.
- Wang F, Zhang Y, Lin X and Ba Z (2024) City-scale buildings damage estimation based on broadband physics-based ground motion simulation of 2021 M_s 6.4 Yangbi, China, earthquake. *Earthquake Spectra* 40(1): 446–468.
- Wang S and Petersson NA (2019) Fourth order finite difference methods for the wave equation with mesh refinement interfaces. *SIAM Journal on Scientific Computing* 41(5): A3246–A3275.
- Wills CJ, Gutierrez CI, Perez FG and Branum DM (2015) A next generation V_{s30} map for California based on geology and topography. *Bulletin of the Seismological Society of America* 105: 3083–3091.
- Withers K, Moschetti M, Powers P, Petersen M, Graves R, Aagaard BT, Baltay A, Luco N, Wirth E, Rezaeian S and Thompson E (2024) Integration of rupture directivity models for the US National Seismic Hazard Model. *Earthquake Spectra* 40(2): 1066–1098.

- Yeh T-Y and Olsen KB (2024) Simulation of 0–7.5 Hz physics-based nonlinear ground motions for maximum credible earthquake scenarios at the Long Valley Dam, CA. *Earthquake Spectra* 40(2): 1479–1506.
- Zhang L, Wang S and Petersson NA (2021a) Elastic wave propagation in curvilinear coordinates with mesh refinement interfaces by a fourth order finite difference method. *SIAM Journal on Scientific Computing* 43(2): A1472–A1496.
- Zhang W, Restrepo D, Crempien JGF, Erkmén B, Taborda R, Kurtulus A and Taciroglu E (2021b) A computational workflow for rupture-to-structural-response simulation and its application to Istanbul. *Earthquake Engineering & Structural Dynamics* 50(1): 177–196.