

Broadband stochastic simulation of earthquake ground motions with multiple strong phases with an application to the 2023 Kahramanmaraş, Turkey (Türkiye), earthquake

Earthquake Spectra

1–37

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/87552930251331981

journals.sagepub.com/home/eqs

SM Sajad Hussaini¹ , Shaghayegh Karimzadeh¹ ,
Sanaz Rezaeian² , and Paulo B Lourenço¹

Abstract

Stochastic ground motion simulation models are often less accurate at lower frequencies than at higher frequencies when fitting recorded data unless supplemented by a deterministic forward directivity velocity pulse model. Moreover, time-modulated stochastic models, which adjust ground motion amplitudes over time, typically use functions that fail to capture multiple strong-motion phases. The February 2023 Turkey (Türkiye) earthquake exhibited diverse recordings, including near-fault and far-field motions with pulse-like and non-pulse-like characteristics, along with single and multiple strong-motion phases. To better represent such a diverse set of recordings, this study enhances a fully non-stationary site-based stochastic model without combining it with a deterministic model. Improvements include a new band-pass filter with upper- and lower-frequency limits, which refines the representation of the low-frequency content. Moreover, a time-modulating function that can represent energy arrival in multiple strong phases is introduced. The reference model's parameters are identified by fitting to the energy content, zero-level crossings, and cumulative counts of positive-minima and negative-maxima of a target accelerogram. This fitting procedure is modified to address the increased number of parameters. These improvements broaden the reference model's applicability while preserving its simplicity, a key aspect appealing to engineering practitioners. The improved model's applicability is demonstrated by simulating a dataset from the February

¹Department of Civil Engineering, Institute for Sustainability and Innovation in Structural Engineering, ARISE, University of Minho, Guimarães, Portugal

²United States Geological Survey, Golden, CO, USA

Corresponding author:

SM Sajad Hussaini, Department of Civil Engineering, Institute for Sustainability and Innovation in Structural Engineering, ARISE, University of Minho, Guimarães 4800-058, Portugal.

Email: hussaini@civil.uminho.pt

2023 Türkiye earthquake, and the accuracy is tested using a pulse-like Next Generation Attenuation Relationships for Western United States dataset. Validations are performed based on total energy, zero-level crossings, Fourier amplitude spectrum, elastic response spectra, and peak ground motion parameters. Validations are performed schematically in the time and frequency domains and quantitatively using goodness-of-fit scores, various validation-metrics errors, and inter-period correlations. Overall, the improved stochastic model can effectively simulate a set of diverse ground motion recordings, including near-fault pulse-like records, records with multiple strong phases, and far-field motions across a broad frequency range.

Keywords

Stochastic ground motion simulation, multiple strong phases, near-fault directivity pulse-like ground motions, far-field ground motions, site-based stochastic model, February 2023 Turkey (Türkiye) earthquake

Date received: 16 June 2024; accepted: 12 March 2025

Introduction

In recent years, advancements in numerical methods for structural analyses, coupled with enhanced computational capabilities, have made it easier to use ground motion time-series in both linear and non-linear dynamic analyses of structures. Despite the expansion of strong-motion networks worldwide, there are still notable gaps in recorded ground motions in many seismic regions. Consequently, the lack of availability of recorded strong ground motions remains a challenge for accurately characterizing seismic hazards in site-specific analysis for many seismic design scenarios (Rezaeian et al., 2024; Yamamoto and Baker, 2013). To acquire supplementary ground motions for a specific design scenario, engineers often must select recorded motions with seismological characteristics different from those of the site of interest and adjust them by scaling or spectrum-matching methods (Hancock et al., 2006; Watson-Lamprey and Abrahamson, 2006). These methods may alter the correlation between important ground motion characteristics and their original physical conditions. Consequently, these operations may provide motions with unrealistic characteristics (Luco and Bazzurro, 2007; Naeim and Lew, 1995). An alternative approach is to perform ground motion simulations that incorporate key features of earthquakes and are consistent with the physical conditions of interest. This is important for realistic estimation of seismic demand, particularly when non-linear structural models are used that account for features such as changes in the frequency content of the input motions. As such, ground motion simulations have been widely used worldwide in seismology and engineering disciplines (Arslan Kelam et al., 2022; Askan et al., 2017; Bernardo et al., 2024; Fayaz et al., 2021; Karimzadeh et al., 2020, 2023, 2024a, 2024b; Rezaeian and Der Kiureghian, 2012; Rezaeian et al., 2017, 2024).

Ground motion simulation techniques can be categorized into three primary classes (Douglas and Aochi, 2008). The first class, often referred to as physics-based or deterministic source-based (Rezaeian and Sun, 2014), uses a physics-informed approach by modeling the fault rupture and propagating the resulting waves to the site of interest. Substantial research efforts have been devoted to the development of deterministic source-based methods, particularly focusing on modeling the source rupture (Beresnev and Atkinson, 1997; Hartzell, 2005, 1978; Zeng et al., 1994) and wave propagation in a three-dimensional (3D)

medium using numerical methods (Komatitsch, 2004; Kristek, 2003). Nonetheless, these methods tend to be complicated and are not commonly used in engineering practice due to their high demand for computational resources and the necessity for a deep understanding of the ground medium, which in turn requires extensive seismological studies. The second class includes methods based on stochastic processes that represent ground motion time-series. Their parameters are sometimes theoretically determined and are sometimes empirically calibrated to fit a collection of recorded ground motions (Boore, 2003; Motazedian and Atkinson, 2005; Rezaeian and Der Kiureghian, 2008). Although deterministic physics-based models can better represent the low frequency of ground motions, stochastic models are often known to better represent the medium to high frequency contents of ground motions (Rezaeian and Sun, 2014). Hybrid methods, as the third class, combine deterministic physics-based and stochastic approaches to generate broadband synthetic ground motions (Graves and Pitarka, 2010) and can accurately represent all frequency ranges but are more complicated and their accuracy often suffers at the transition frequency.

Stochastic methods can be categorized into source-based and site-based approaches (Rezaeian and Sun, 2014). In the source-based stochastic approach, a theoretical spectrum of the earthquake source is shaped and scaled by incorporating the effects of path and site, which can be described in terms of seismological parameters (Boore, 2003, 2009; Motazedian and Atkinson, 2005). These parameters are often region-specific and can vary notably from one region to another, or they may be unavailable in a region of interest, thus limiting their applications in engineering practice (Chouet et al., 1978; Stafford et al., 2009). Given the complexity of simulation methods and the expertise and seismological knowledge they require, it is often advised for engineering practitioners to use a validated set of simulated ground motions (Rezaeian et al., 2024), which are not currently readily available for all seismic regions of interest to the engineering community. On the other hand, site-based stochastic approaches represent ground motions at a specific region by using a stochastic process capable of capturing the important characteristics of previously observed motions in that or similar region. They often require fewer parameters and notably lower computational costs compared to source-based or hybrid simulations. These advantages, along with their simplicity, are key aspects that make site-based stochastic models more appealing to engineering practitioners.

The key difference among various site-based stochastic models in the literature is how the underlying stochastic process is defined and enhanced to represent the characteristics of recorded ground motion time-series. Following Rezaeian and Sun (2014), existing site-based stochastic models can be classified into five categories: (a) models based on filtering and time-modulating a white noise process (Alamilla et al., 2001; Dabaghi and Der Kiureghian, 2017; Rezaeian and Der Kiureghian, 2008; Waezi and Balzadeh, 2022; Waezi et al., 2021), (b) models based on filtering a train of Poisson pulses (Lin, 1986), (c) autoregressive moving average models (Sgobba et al., 2011), (iv) direct spectral representations using short-time Fourier or wavelet transforms (Huang and Wang, 2017; Pousse et al., 2006; Vlachos et al., 2016; Yamamoto and Baker, 2013), and (v) data-driven machine learning methods such as Gaussian process regression (Alimoradi and Beck, 2015; Tamhidi et al., 2022).

Most existing site-based models in the literature rely on mathematical parameters rather than those with meaningful physical interpretations, although some have attempted to use physically meaningful parameters (Rezaeian and Der Kiureghian, 2008) or related the model parameters to other seismological parameters through empirical regressions (Rezaeian and Der Kiureghian, 2012; Sgobba et al., 2011). Improving these models to

capture physical characteristics such as low-frequency content and near-fault directivity effects entails adding complexity, more parameters, and tailoring them to a specific scenario. For instance, Dabaghi and Der Kiureghian (2017) and Waezi and Balzadeh (2022) added a deterministic forward directivity velocity pulse model to a site-based stochastic model to simulate ground motions with a low-frequency directivity velocity pulse. Several studies incorporated sequential filtering, including a stationary high-pass filter in their model to capture two or three peak frequencies in the evolutionary power spectral density (PSD) of an observed ground motion time-series (Broccardo and Dabaghi, 2019; Vlachos et al., 2016; Waezi and Balzadeh, 2022; Waezi et al., 2021). Nonetheless, additional post-simulation high-pass filtering is still required in most site-based stochastic models to remove undesired low frequencies. Filtering the low-frequency content after generating synthetics can lead to an implicit average bias that tends to underestimate the total energy of the synthetic record. Consequently, several studies have resorted to correction factors as a post-simulation solution for addressing energy bias in simulations (Broccardo and Dabaghi, 2019; Dabaghi and Der Kiureghian, 2017).

Stochastic approaches, despite their advancements, still face limitations in capturing low-frequency content and modeling multiple strong-motion phases without a deterministic component. On the other hand, real recorded ground motion sets, such as those observed during the February 6, 2023, Kahramanmaraş, Turkey (Türkiye), earthquake, can exhibit a variety of such features. The February 6, 2023, Kahramanmaraş, Türkiye, earthquake with a moment magnitude (M_w) of 7.8, caused extensive damage, including around 50,000 fatalities and the collapse of more than 19,000 buildings (Wang et al., 2024). This large-magnitude event exhibited a diverse range of recordings, including both near-fault and far-field motions, with directivity pulse-like and non-pulse-like characteristics, low-frequency basin amplifications, as well as single and multiple strong-motion phases. Prior to the earthquake, Arslan Kelam et al. (2022) used the stochastic finite-fault approach by (Motazedian and Atkinson, 2005) to assess the seismic vulnerability of buildings in the Gaziantep region, located near a major seismic data gap. Following the event, (Askan et al., 2025) performed a comprehensive study of the region, including pre-earthquake findings such as regional velocity models, stochastic ground motion simulations, building vulnerability assessments, and post-event damage modeling, providing valuable insights for seismic resilience. Wang et al. (2024) applied the stochastic finite-fault approach by (Motazedian and Atkinson, 2005) to simulate ground motions at eight stations for this event, achieving a good fit with real data for most of the considered stations. Despite their usefulness, these studies suffered from the limitations of current stochastic approaches in accurately representing low-frequency contents and multiple strong phases in time-series.

This study aims to enhance a fully non-stationary site-based stochastic ground motion simulation model developed by Rezaeian and Der Kiureghian (2008), hereafter called the *reference model*, to overcome the limitations of current stochastic approaches. This reference model is based on time-modulating a filtered Gaussian white noise process. It features a simple formulation, few physically informed model parameters, and computational efficiency. The model parameters are calibrated for a given seismic region by aligning the cumulative energy (CE), zero crossings, and positive-minima and negative-maxima of the model with those of the target accelerogram that have also been suggested as simulation validation metrics (Rezaeian et al., 2024). Parameters of the reference model were chosen based on functions that can accurately represent far-field ground motions where the total energy arrives in a single strong-motion phase, but they could not represent low-frequency

content observed in pulse-like and deep basin motions or temporal non-stationarities that are caused by multiple strong phases, without post-processing or combining with deterministic models. These shortcomings were recognized, and the use of alternative filters and parameters was recommended for future improvements to the model, which are undertaken in the current study.

The improvements presented in this study extend the applicability of the reference model to various types of ground motions, including the diverse set of recordings observed during February 6, 2023, Kahramanmaraş, Türkiye, earthquake, while maintaining the model's simplicity and ease of use by engineering practitioners. The improvements include: (a) adjusting the filtering order after time-modulating in the model formulation to avoid distortion of very low frequencies, (b) eliminating the post-simulation filtering to avoid correction factors for adjusting to the energy bias, (c) introducing a new band-pass filter function to represent the very low-frequency components, implicitly considering near-fault rupture directivity pulses and potential long-period basin effects without relying on supplementary deterministic models, and (d) introducing a new time-modulating function to account for the arrival of energy in multiple strong phases. The extended applicability of the improved model is first demonstrated for a diverse set of recordings that correspond to the February 6, 2023, Kahramanmaraş, Türkiye, earthquake. The accuracy is further tested by simulating a dataset of pulse-like motions from the Next Generation Attenuation Relationships for Western United States (NGA-West2) dataset and assessing the goodness-of-fit (GoF) scores (Olsen and Mayhew, 2010), the empirical cumulative distribution of various validation-metrics errors (Rezaeian et al., 2024), and the inter-period correlations (Burks and Baker, 2014). Although deterministic forward directivity velocity pulse models may still be more accurate approaches in representing near-fault directivity effects for specific seismic regions where detailed information on the fault rupture and the ground medium are available, the model improvements in this study present an acceptable and much simpler approach to simulate a diverse set of ground motions when our knowledge of seismological information is limited or when representing the variability in input ground motions is an objective of more advanced structural analyses.

Selected datasets

To simulate strong ground motions from the February 6, 2023, Kahramanmaraş, Türkiye, earthquake ($M_w = 7.8$), hereafter referred to as the *Kahramanmaraş* earthquake, we select a dataset of recorded motions with Joyner-Boore distances (R_{JB}) less than 80 km from the pan-European strong-motion (ESM) database (Luzi et al., 2016). Figure 1 shows the frequency distribution of the 41 selected recording pairs (82 components in the as-recorded directions) with respect to R_{JB} and the averaged shear wave velocity at the top 30 m of soil (V_{S30}). The selected records encompass a wide range of source-to-site distances and soil conditions. Wu et al. (2023) identified 18 pairs of these recordings as pulse-like motions, using the pulse index value at the orientation of the largest pulse determined through Baker's algorithm (Baker, 2007). Figure 2 shows a scatter plot of V_{S30} versus R_{JB} , distinguishing pulse-like and non-pulse-like motions, along with the period of the largest pulse for the identified pulse-like motions. Four sample time-series in the north-south (NS) and east-west (EW) directions are shown in Figure 3, showing a diverse set of characteristics, including pulse-like, non-pulse-like, single, and double strong phase motions at different source-to-site rupture distances.

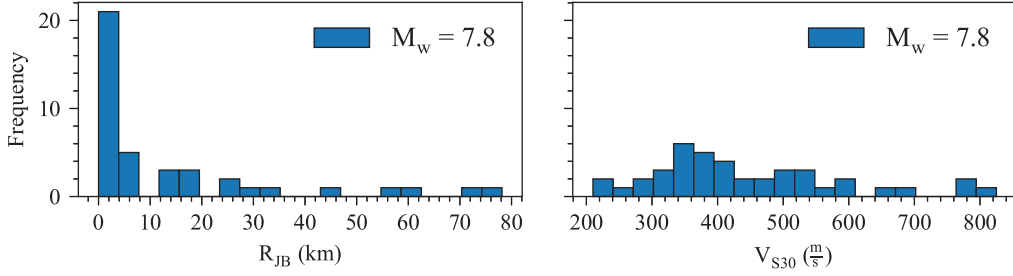


Figure 1. Frequency distribution of the selected ground motions from Kahramanmaraş, Türkiye, earthquake with moment magnitude (M_w) 7.8, with respect to Joyner-Boore distance (R_{JB}) and averaged shear wave velocity at the top 30 m of soil (V_{S30}).

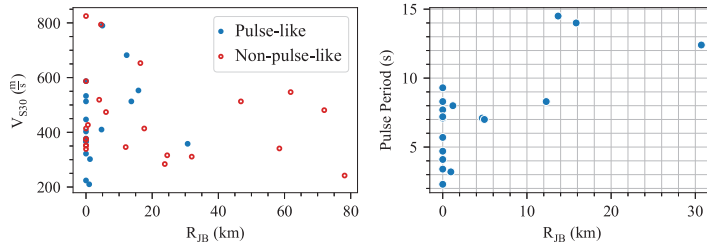


Figure 2. Scatter plot of V_{S30} versus R_{JB} for the Kahramanmaraş, Türkiye, earthquake dataset (Luzi et al., 2016), distinguishing pulse-like and non-pulse-like motions, along with the pulse period values for the identified pulse-like motions.

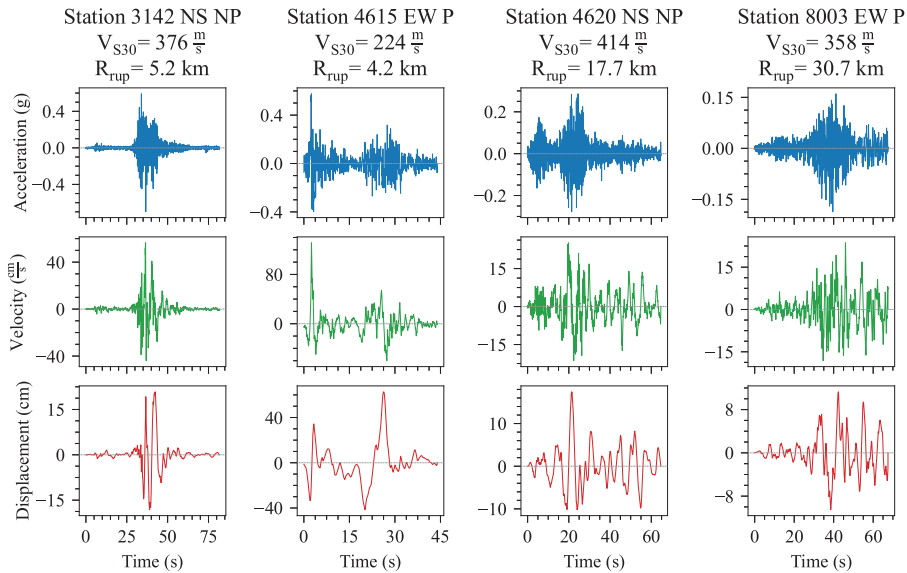


Figure 3. Sample ground motion time-series from the Kahramanmaraş, Türkiye, earthquake dataset (Luzi et al., 2016), at various V_{S30} , source-to-site rupture distances (R_{rup}), showing pulse-like (P), non-pulse-like (NP), single (left and right panels), and double strong-phase motions (mid panels).

To assess the ability of the improved model to accurately represent low-frequency pulse-like ground motions for a range of earthquake magnitudes, a subset of the NGA-West2 database (Ancheta et al., 2014) is also selected. This subset includes 97 recording pairs (194 components) of pulse-like motions with R_{JB} less than 10 km, selected without additional assumptions. Figure 4 shows the frequency distribution of the selected motions across R_{JB} , V_{S30} , and M_w , and Figure 5 shows a scatter plot of V_{S30} versus R_{JB} , along with estimated values of pulse periods for each motion as calculated by Shahi and Baker (2014) and reported in the NGA-West2 database.

The combined Kahramanmaraş earthquake and NGA-West2 datasets result in a total of 276 representative ground motion components, sourced from various regions and earthquakes. Information about the selected records regarding the ESM (<https://esm-db.eu>) and NGA-West2 (<https://ngawest2.berkeley.edu/site>) databases is summarized in Supplementary Table A2.1 and Table A2.2, respectively.

The four accelerograms, shown in Table 1, are selected for illustrative purposes and discussions in this study. These example accelerograms include the non-pulse-like component 90 of the Northridge-01 1994 (United States) earthquake recorded at the 116th Street School station with record sequence number (RSN) of NGA RSN 984, the pulse-like component 0 of the Imperial Valley-06 1979 (United States) earthquake recorded at El Centro-Meloland Geotechnical Array station (NGA RSN 171), the pulse-like horizontal EW component of the L'Aquila 2009 (Italy) earthquake recorded at station L'Aquila-V.Aterno-F. Aterno (NGA RSN 4482), and the pulse-like horizontal EW component of the Kahramanmaraş 2023 (Türkiye) earthquake recorded at station 4615 (ESM ID INT-

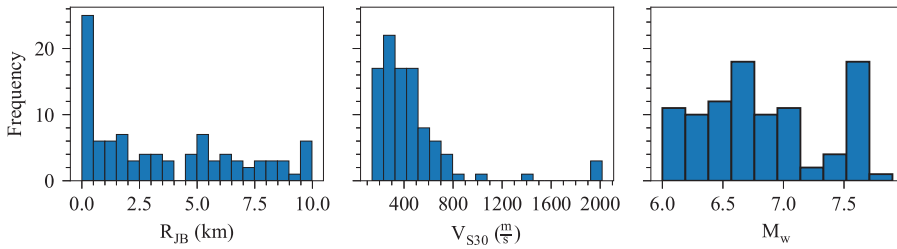


Figure 4. Frequency distribution of the selected ground motions from the Next Generation Attenuation Relationships for Western United States (NGA-West2) dataset (Ancheta et al., 2014), with respect to R_{JB} , V_{S30} , and M_w .

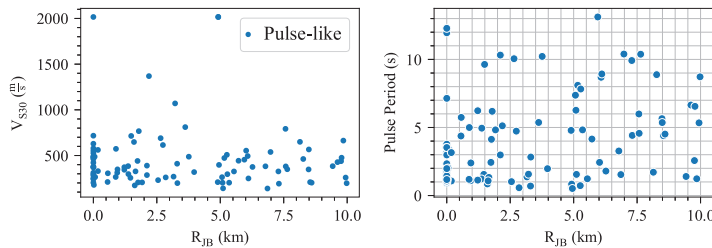


Figure 5. Scatter plot of V_{S30} versus R_{JB} for the selected pulse-like NGA-West2 dataset (Ancheta et al., 2014), along with the pulse period values of the motions.

Table 1. Seismological characteristics of example target accelerograms, including moment magnitude (M_w), Joyner-Boore distance (R_{JB}), Rupture distance (R_{rup}), and averaged shear wave velocity at the top 30 m of soil (V_{S30})

ID	NGA RSN / ESM ID	Event	M_w	Mechanism	R_{JB} (km)	R_{rup} (km)	V_{S30} ($\frac{m}{s}$)
NR	984	Northridge	6.69	Reverse	36.39	41.17	301
IV	171	Imperial Valley	6.53	Strike slip	0.07	0.07	264
LA	4482	L'Aquila	6.30	Normal	0.00	6.55	552
KM	INT-20230206-0000008	Kahramanmaraş	7.80	Strike slip	0.00	4.20	224

20230206-0000008). The latter two accelerograms exhibit double strong-motion phases. In this study, an identifier (ID) is assigned to each target recorded ground motion for ease of reference, as shown in Table 1.

The stochastic ground motion model

The reference model

This study builds upon the site-based stochastic model developed by Rezaeian and Der Kiureghian (2008). This reference model accounts for both the temporal and spectral non-stationarities of earthquake ground motion time-series. One of the key benefits of this model is the ease of parameter estimation, achieved by separating the temporal and spectral non-stationary features of the stochastic process. The ground acceleration process $x(t)$ is obtained by time-modulating a normalized filtered white noise process using an evolutionary (i.e. time-varying) filter. In the continuous form, the model is defined as:

$$\ddot{x}(t) = q(t, \alpha) \left\{ \frac{1}{\sigma_h(t)} \int_{-\infty}^t h(t - \tau, \lambda(\tau)) w(\tau) d\tau \right\} \quad (1)$$

where $q(t, \alpha)$ is the time-modulating function with parameters in vector α , $w(\tau)$ is the white noise process, $h(t - \tau, \lambda(\tau))$ is the shifted filter with time-varying parameters in vector $\lambda(\tau)$, and $\sigma_h^2(t) = \int_{-\infty}^t h^2(t - \tau, \lambda(\tau)) d\tau$ is the variance of the integral process. Normalization by $\sigma_h(t)$ results in a unit variance process denoted by the term inside the curved brackets. Thus, $q(t, \alpha)$ equals the standard deviation of the resulting process $\ddot{x}(t)$. In this formulation, the modulating function $q(t, \alpha)$ describes the temporal characteristics of the process, whereas the evolutionary filter defines the spectral characteristics. Rezaeian and Der Kiureghian (2008, 2010) used modified versions of the piecewise and Gamma modulating functions, capable of considering a single strong-motion phase (Housner and Jennings, 1964; Rodolfo Saragoni and Hart, 1973). For the evolutionary filter, the impulse response function (IRF) of the pseudo-acceleration response of an underdamped single degree of freedom (SDoF) linear system was used, which is characterized by a linearly time-varying predominant frequency and a damping ratio. This approach assumes that the bandwidth is centered around the predominant frequency, which is adopted by most site-based stochastic models for simplicity (e.g. Dabaghi and Der Kiureghian, 2017; Rezaeian and Der Kiureghian, 2010). In addition, a critically damped second-order oscillator was used as a high-pass filter to post-process the time-series and ensure zero residual velocity and displacement. The post-processed acceleration record, denoted by $\ddot{z}(t)$, was obtained from the solution of the differential equation:

$$\ddot{z}(t) + 2\omega_c \dot{z}(t) + \omega_c^2 z(t) = \ddot{x}(t) \quad (2)$$

where ω_c is the low-cut frequency of the filter, and $\ddot{x}(t)$ is the acceleration process defined in Equation 1.

The model formulation in Equation 1 and the corresponding high-pass filter in Equation 2 have served as the basis for the development of several improved non-stationary site-based stochastic models in the literature. These models have primarily aimed to broaden the applicability range of the model by enhancing its low-frequency content representation either by incorporating a deterministic forward directivity velocity pulse model or by refining the filter through sequential filtering techniques (e.g. Dabaghi and Der Kiureghian, 2017; Waezi and Balzadeh, 2022; Waezi et al., 2021). Moreover, the evolutionary frequency content of a ground motion time-series is affected by several factors, including the characteristics of the earthquake source, path, and site. These intricate non-stationary processes demonstrate multiple peak frequencies in their evolutionary PSDs. To better characterize this spectral non-stationarity, several studies have assumed two or three peak frequencies and corresponding bandwidths at the expense of additional parameters and complexity (Vlachos et al., 2016; Waezi and Balzadeh, 2022; Waezi et al., 2021). However, direct identification of multiple evolutionary peak frequencies is not straightforward and often requires non-stationary spectral estimation techniques such as short-time Fourier transform (Vlachos et al., 2016; Waezi and Balzadeh, 2022).

The limitations

The reference model in Equation 1, with the application of the high-pass filter in Equation 2, exhibits two main limitations. First, there could be a total energy deficit in the final process due to the high-pass filtering, and although this bias was negligible for the applicability range of the original records considered by Rezaeian and Der Kiureghian (2008), it could be noticeable for records that are near-fault or on deep basins and exhibit lower-frequency contents. Although the residual velocity and displacement are removed through a stationary high-pass filter, this post-processing step can eliminate a part of the low-frequency contents and the corresponding energy from the simulations. Thus, the final process can systematically underestimate the total energy with respect to the initial process. In the validation studies done by Rezaeian and Der Kiureghian (2008, 2012), such underestimations were considered nonsignificant for the selected set of strong far-field recorded ground motions; however, they may be of considerable interest in ground motions where a substantial portion of the energy comes from low-frequency contents, such as those observed in some of the records in the Kahramanmaraş earthquake.

Several studies have incorporated a stationary high-pass filter as a part of the main filter function within the reference model (Broccardo and Dabaghi, 2019; Waezi and Balzadeh, 2022; Waezi et al., 2021) or an evolutionary power spectral-based model (Vlachos et al., 2016). Although this inclusion may alleviate excessive low-frequency contents in the final process, several challenges persist. First, determining the optimal high-pass frequency parameter remains ambiguous, often requiring iterative adjustments to the parameter to achieve realistic velocity and displacement time-series (Rezaeian and Der Kiureghian, 2008; Vlachos et al., 2016), which has some degree of subjectivity. In a recent study (Su et al., 2024), this stationary high-pass corner frequency was optimized by matching the long-period ($T \geq 1$ s) linear acceleration response spectrum of synthetic ground motions to that of each target record. Second, the implementation of a stationary high-pass filter is

susceptible to generating undesired response amplitudes near its frequency parameter due to the cumulative effect of response contributions in the integral term in Equation 1. Consequently, a high-pass filter that attenuates frequencies below a stationary corner frequency, uniformly across all time instances, may be inadequate for simulating ground motions in which an important part of the non-stationary frequency content lies within the low-frequency range, such as in near-fault pulse-like motions and motions with deep basin effects.

As the second limitation, time-modulating after filtering a process as highlighted by Safak and Boore (1988) distorts the low-frequency content in the Fourier amplitude spectrum (FAS) of the process. This latter limitation indicates that models based on Equation 1, despite incorporating a stationary high-pass filter, could still require an additional post-simulation high-pass filtering. Consequently, this results in an inherent bias in the total energy, requiring a post-simulation energy correction factor to address the induced bias as described by Dabaghi and Der Kiureghian (2017) and Broccardo and Dabaghi (2019). Similarly, site-based stochastic models are typically fitted to target records based on criteria such as the rates of zero-level crossings and counts of local extrema. However, post-simulation high-pass filtering can introduce bias, causing the simulations to deviate from the original fitting criteria.

These two limitations restrict the application of the reference model to only far-field ground motions and motions without the presence of deep basin effects, in which low-frequency contents do not contribute to a substantial portion of the total energy. To broaden the applicability range of the reference model to near-fault pulse-like motions, previous studies have used a deterministic forward directivity velocity pulse model (Dabaghi and Der Kiureghian, 2017; Waezi and Balzadeh, 2022). This approach involves several steps to first identify and extract the largest pulse using a wavelet-based method, as outlined in Baker's algorithm (Baker, 2007), and then, a forward directivity velocity pulse model would be fitted to the extracted pulse in the velocity domain. The derivative of this model is then combined with a stochastic model of the residual motion, tailoring it to cases of forward directivity velocity pulses.

The proposed improved stochastic model

In this study, we first modify the order of time-modulating and filtering in the formulation of the reference model. The proposed modification is based on band-pass filtering of a time-modulated Gaussian white noise process. The band-pass filter is selected based on the response of a linear time-variant (LTV) SDoF system. Following the same terminology as the reference model, we define the IRF of the LTV filter as $h(t, \lambda(\tau))$ where $\lambda(\tau)$ indicates the vector of time-varying parameters. The filter is assumed to be causal $h(t, \lambda(\tau)) = 0$ for $t < 0$, stable $\int_{-\infty}^{\infty} h(t, \lambda(\tau)) dt < \infty$, and at least twice differentiable. The time domain representation of the improved site-based stochastic model in the continuous form is as follows:

$$\ddot{x}(t) = \int_{-\infty}^{\infty} q(\tau, \alpha) w(\tau) \frac{\delta(t - \tau) * h(t, \lambda(\tau))}{\gamma_h(\tau)} d\tau \quad (3)$$

where $\ddot{x}(t)$, $q(\tau, \alpha)$, and $w(\tau)$ are the same as those in the reference model, $\delta(t - \tau)$ is a shifted Dirac delta function convolved with $h(t, \lambda(\tau))$ which represents a band-pass filter as

a function of time (t), and $\gamma_h^2(\tau) = \int_{-\infty}^{\infty} h^2(t - \tau, \lambda(\tau)) dt$ is the energy normalization function of the filter. This normalization ensures that for each instance of τ , the response of the normalized filter has a unit of energy over time. Therefore, the modulating function solely controls the distribution of energy over time, controlling the temporal non-stationarity of the stochastic model (refer to Supplementary Appendix 1). Meanwhile, the normalized filter function at each instance of τ , adjusts the distribution of energy across frequencies, controlling the spectral non-stationarity of the process. Therefore, the non-stationarity terms continue to be defined separately, preserving the convenient identification and interpretation of the model parameters. The evolutionary band-pass filter will account for a broad frequency bandwidth and eliminate the need for post-simulation high-pass filtering. In other words, the frequency content is characterized by an evolutionary bandwidth, delimited by a lower and an upper dominant frequency, which is anticipated to encompass the primary frequencies of interest. Therefore, the total energy of the model is always controlled by $q(t, \alpha)$ without any systematic bias in the total energy of the simulated process, eliminating the need for further correction factors.

It is important to note that the term inside the integral in Equation 3 is a stationary Gaussian process at each instance of τ and uncorrelated from other instances. Thus, the convolution operation in time can be transformed into a simple multiplication in the frequency domain using the corresponding frequency response function (FRF) of the filter. Based on the linearity property of the Fourier transform and Parseval's theorem (refer to Supplementary Appendix 1), one can show that:

$$\ddot{x}(t) = \mathcal{F}^{-1}(\ddot{x}(\bar{\omega})) = \mathcal{F}^{-1} \left(\int_{-\infty}^{\infty} q(\tau, \alpha) w(\tau) \frac{e^{-i\bar{\omega}\tau} H(\bar{\omega}, \lambda(\tau))}{\gamma_H(\tau)} d\tau \right) \quad (4)$$

where $H(\bar{\omega}, \lambda(\tau))$ is the FRF of the filter as the function of frequency ($\bar{\omega}$), $\gamma_H^2(\tau) = \int_{-\infty}^{+\infty} |H(\bar{\omega}, \lambda(\tau))|^2 d\bar{\omega}$ is the equivalent energy normalization function of the filter in the frequency domain, and $\ddot{x}(\bar{\omega})$, defined by the term inside parentheses, is the model representation in the frequency domain. It is noteworthy that the term inside the integral in Equation 4 can be considered as the time-frequency representation of the model at each instance of τ . Figure 6 illustrates a schematic representation of the improved site-based stochastic model in both time and frequency domains.

The proposed filter function. For the filter function, we use the FRF of the filter $H(\bar{\omega}, \lambda(\tau))$ instead of its IRF $h(t, \lambda(\tau))$ to leverage the benefits of the frequency domain over the time domain for two reasons. First, the time domain convolutions are simplified to mere multiplications in the frequency domain, making the computations more straightforward. Second, the form of IRF of an SDoF system can vary depending on whether it is under, over, or critically damped. This requires modification of the IRF's form to account for the appropriate damping ratio. In contrast, the FRF of an SDoF system maintains the same fundamental form regardless of the damping ratio, resulting in a stable formulation for determining its response. In this study, the proposed band-pass filter is constructed based on cascading the displacement and acceleration IRF of an LTV SDoF system. In the time domain, the filter is expressed as follows:

$$h(t, \lambda(\tau)) = h_l(t, \lambda_l(\tau)) * h_u(t, \lambda_u(\tau)) \quad (5)$$

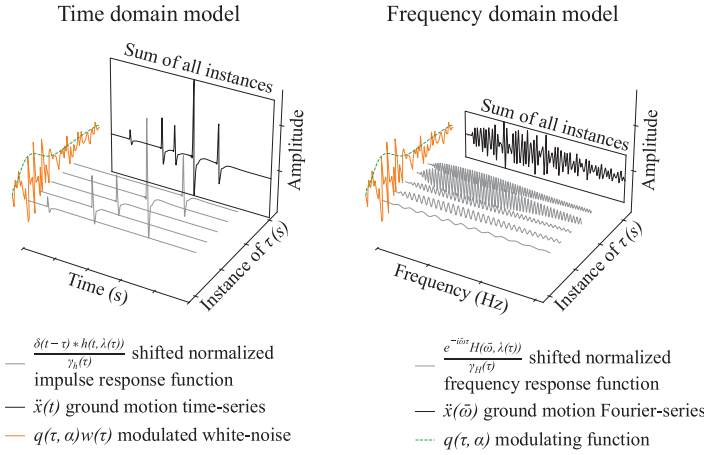


Figure 6. Schematic representation of the improved stochastic model in the time and frequency domains.

$$h_l(t, \lambda_l(\tau)) = \frac{d^2}{dt^2} [h_u(t, \lambda_l(\tau))] \quad (6)$$

where subscripts l and u indicate association with the lower and upper dominant frequencies, respectively. Here, $h_u(t, \lambda(\tau))$ is the displacement IRF, expressed as follows:

$$h_u(t, \lambda(\tau)) = \frac{1}{\omega(\tau) \sqrt{1 - \zeta^2(\tau)}} e^{-\zeta(\tau) \omega(\tau) t} \sin\left(\omega(\tau) \sqrt{1 - \zeta^2(\tau)} t\right) \quad (7)$$

where $\lambda(\tau) = [\omega(\tau), \zeta(\tau)]$ is the set of filter parameters corresponding to its natural frequency and damping ratio, respectively. In the frequency domain, the corresponding FRF can be expressed as:

$$H(\bar{\omega}, \lambda(\tau)) = H_l(\bar{\omega}, \lambda_l(\tau)) H_u(\bar{\omega}, \lambda_u(\tau)) \quad (8)$$

$$H_l(\bar{\omega}, \lambda_l(\tau)) = \frac{-\bar{\omega}^2}{\omega_l^2(\tau) - \bar{\omega}^2 + i2\zeta_l(\tau)\omega_l(\tau)\bar{\omega}} \quad (9)$$

$$H_u(\bar{\omega}, \lambda_u(\tau)) = \frac{1}{\omega_u^2(\tau) - \bar{\omega}^2 + i2\zeta_u(\tau)\omega_u(\tau)\bar{\omega}} \quad (10)$$

here, $H_l(\bar{\omega}, \lambda_l(\tau))$ and $H_u(\bar{\omega}, \lambda_u(\tau))$ control low- to intermediate-frequency and intermediate- to high-frequency content, respectively, and eliminate the zero-frequency amplitude. The proposed filter offers a simpler functional form and requires fewer parameters than existing filters (Vlachos et al., 2016; Waezi and Balzadeh, 2022). Moreover, it ensures consistent behavior for ζ_l and ζ_u such that identical values for these parameters yield equal filter response amplitudes at ω_l and ω_u . This feature is important because ζ_l and ζ_u are expected to respectively control the bandwidth from low to intermediate and intermediate to high frequencies, thereby associating filter parameters to characteristics of a ground motion time-series.

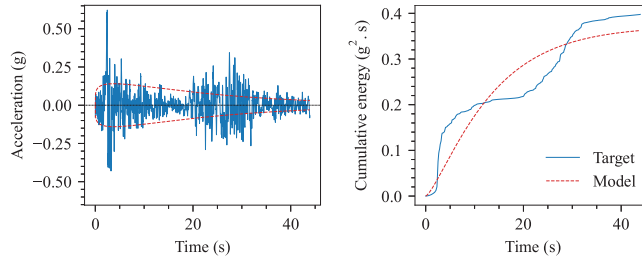


Figure 7. The Kahramanmaraş accelerogram with multiple strong phases, along with the fitted Gamma function and their cumulative energy.

The functional forms of the filter parameters $\omega_l(\tau)$, $\omega_u(\tau)$, $\zeta_l(\tau)$, $\zeta_u(\tau)$ depend on the time-variation of the frequency content of an earthquake ground motion. For simplicity, we opt for a linear function as below:

$$p(\tau) = p_0 - \frac{p_0 - p_n}{D} \tau \quad (11)$$

where $p(\tau)$ represents each of the filter parameters (i.e. $\omega_l(\tau)$, $\omega_u(\tau)$, $\zeta_l(\tau)$, $\zeta_u(\tau)$), D is the total duration of motion, p_0 and p_n are the initial and final filter parameters, respectively.

The proposed time-modulating function. Existing time-modulating functions include the piecewise function (Amin and Ang, 1968; Housner and Jennings, 1964), the double-exponential function (Shinozuka and Sato, 1967), and the Gamma function (Rodolfo Saragoni and Hart, 1973). Stafford et al. (2009) proposed a modulating function based on the Lognormal distribution, which models the energy buildup in ground motion time-series. These parameterized time-modulating functions describe the arrival of energy in a single strong phase, preceded by an energy buildup and followed by a decay, which may not necessarily reach zero at the end of the acceleration process. Consequently, for accelerograms characterized by the arrival of energy in multiple strong phases, these functions require adjustments. Figure 7 illustrates an example of such ground motions, featuring the Kahramanmaraş (KM) accelerogram (refer to Table 1) along with a corresponding fitted Gamma function as used by Rezaeian and Der Kiureghian (2010). Notably, the fitted Gamma function is unable to capture the realistic arrival of energy.

Moreover, existing time-modulating functions rely on mathematical parameters that often lack physical interpretation. One way to relate these parameters to the seismological characteristics of a scenario earthquake is through empirically developed predictive models (Rezaeian and Der Kiureghian, 2010; Stafford et al., 2009). Modulating functions often require adjustments to account for various shapes of ground motion time-series, such as near-fault pulse-like accelerograms, by introducing additional parameters in the function (e.g. Dabaghi and Der Kiureghian, 2017). Numerical interpolation methods like monotonic cubic Hermite splines can require a varying number of input parameters depending on the target motion and desired accuracy. Broccardo and Dabaghi (2019) recommended using seven interpolation points to model the modulating function. These numerical methods may necessitate supplementary post-processing to achieve a smooth modulating function and circumvent issues like overshooting (e.g. a local abrupt jump), particularly when neighboring interpolation points are spaced very closely. This can be challenging for

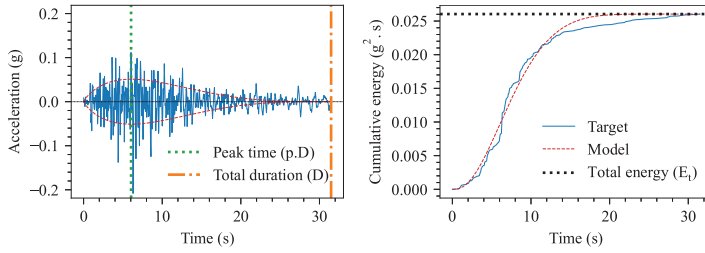


Figure 8. Parameters of the basic beta function fitted to the Northridge accelerogram. $E_t = 0.026037 \text{ g}^2\text{s}$, $D = 31.48 \text{ s}$, $p.D = 6.04 \text{ s}$, $c = 8.65$.

recorded motions with rapid energy arrivals, requiring a choice of interpolation points based on the specific characteristics of the motion.

To overcome the limitations of existing time-modulating functions, here we use concave Beta distributions to describe the energy buildup in ground motions. Let $a(t)$ be an accelerogram with its total expected energy defined as $E_t = \int_0^D a^2(t)dt$, arriving in a single strong phase with a total recorded duration of D . A basic Beta time-modulating function can be proposed as below:

$$q(t, E_t, D, c, p) = \sqrt{E_t \cdot \frac{t^{cp} (D-t)^{c(1-p)}}{\beta(1+cp, 1+c(1-p)) D^{1+c}}} \quad (12)$$

where p is the ratio of the time of peak amplitude to the total duration (i.e. $p = \frac{\text{peak time}}{D}$), c is a dimensionless parameter that describes the distribution of energy around the peak time, and β is the Beta function. Figure 8 illustrates the total duration (D), total energy (E_t), and peak time (i.e. $p.D$) using an example application of the basic Beta modulating function as the model on the Northridge (NR) accelerogram (refer to Table 1), featuring a single strong-motion phase.

The arrival of energy in multiple strong phases can be replicated by a weighted sum of basic Beta time-modulating functions. The proposed Beta modulating function which explicitly starts from and ends in zero amplitude, can be expressed as follows:

$$q(t, E_t, D, \vec{\alpha}, \vec{c}, \vec{p}) = \sqrt{E_t \cdot \sum_{i=0}^n \left(\alpha_i \frac{t^{c_i p_i} (D-t)^{c_i(1-p_i)}}{\beta(1+c_i p_i, 1+c_i(1-p_i)) D^{1+c_i}} \right)} \quad (13)$$

where $\vec{\alpha}$, $\vec{c}$, $\vec{p}$ are vectors of size $n + 1$, with p_i representing the i^{th} strong phase peak-time ratio, α_i denoting the ratio of energy contributed by the i^{th} strong phase to the total energy, and c_i quantifying the energy concentration around the peak time within that i^{th} strong phase. The term corresponding to $i=0$ is considered as a baseline energy level, to capture the very low-amplitude tails of a ground motion. Therefore, 5% of the total energy ($\alpha_0 = 5\%$) is distributed over the total duration with parameters $c_0 = 2.0$ and $p_0 = 0.5$. Setting $n=1$ allows the modulating function to account for a single strong phase using four parameters (E_t, D, p_1, c_1). By setting $n=2$, the proposed modulating function accounts for up to two strong phases ($i=1, 2$) with seven parameters ($E_t, D, \alpha_1, p_1, c_1, p_2, c_2$). Note that α_2 can be determined by specifying α_1 (i.e. $\alpha_2 = 1 - 0.05 - \alpha_1$). With this

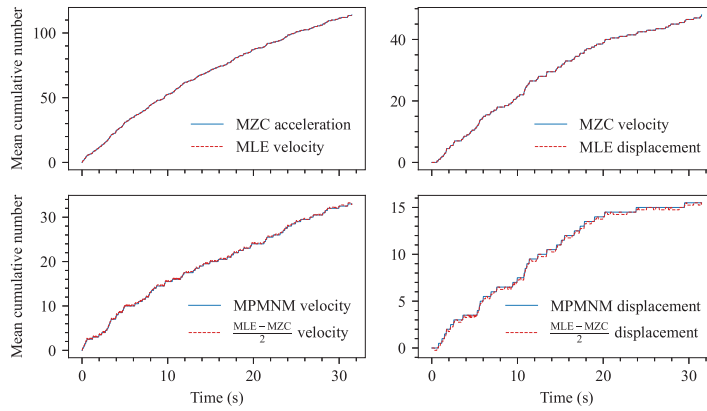


Figure 9. Relationships between time-series characteristics of recorded ground motions, using the Northridge accelerogram. Characteristics include mean zero crossing (MZC), mean local extrema (MLE), and mean positive-minima and negative-maxima (MPMNM).

modulating function, adding more strong phases only requires dividing the total energy and specifying the corresponding parameters (i.e. α_i, p_i, c_i).

Therefore, the improved stochastic model is fully defined by eight parameters ($\omega_{l0}, \omega_{ln}, \omega_{u0}, \omega_{un}, \zeta_{l0}, \zeta_{ln}, \zeta_{u0}, \zeta_{un}$) that specify the filter function, and seven parameters ($E_t, D, \alpha_1, p_1, c_1, p_2, c_2$) that specify the time-modulating function for up to two strong phases.

Fitting to the evolutionary frequency and energy contents of the ground motion process

Time-varying frequency content

To characterize the predominant frequency and bandwidth, many site-based stochastic models have used the rate of zero-level up-crossings and positive-minima and negative-maxima of the acceleration process, respectively (e.g. Dabaghi and Der Kiureghian, 2017; Rezaeian and Der Kiureghian, 2008; Rodolfo Saragoni and Hart, 1973; Waezi et al., 2021; Yeh and Wen, 1990). To explore the characteristics of a ground motion time-series, we define mean zero crossing (MZC), mean local extrema (MLE), and mean positive-minima and negative-maxima (MPMNM) as follows; MZC represents the average number of zero-level up-crossings and down-crossings, MLE denotes the average number of local minima and maxima, and MPMNM indicates the average number of local positive-minima and negative-maxima. Analysis of these three features within recorded ground motions demonstrates that the MZC of acceleration equals the MLE of velocity, and the MZC of velocity equals the MLE of displacement, which is supported by the theoretical definitions of acceleration, velocity, and displacement processes. Similarly supported by their definitions, the MPMNM of velocity is half of the difference between its MLE and MZC. Likewise, the MPMNM of displacement is half of the difference between its MLE and MZC. Figure 9 illustrates these relationships for the example NR ground motion (refer to Table 1).

As such, the MPMNM of a process can be replaced with the difference between the MZC of that process and its time-derivative. In other words, given that the MZC of a process (i.e. acceleration, velocity, or displacement) is related to its dominant frequencies, the

MPMNM of the process is associated with the bandwidth between the dominant frequencies of that process and its time-derivative. Therefore, the MZC of acceleration, velocity, and displacement processes can be used as proxies for high, intermediate, and low dominant frequencies, respectively. Similarly, the MPMNM of the displacement and velocity processes can be used as proxies for the bandwidth between low to intermediate and intermediate to high dominant frequencies. As the MPMNM of the acceleration process acts as a proxy for the bandwidth between the acceleration and its time-derivative, it may not provide any advantages for modeling the acceleration, velocity, and displacement time-series. Consequently, it is not used as a fitting criterion in this study. Subsequently, we formulate these features for the acceleration, velocity, and displacement processes of the stochastic model.

Let $g(t, \tau)$ be the term inside the integral in Equation 3. Because $g(t, \tau)$ represents a stationary Gaussian process at each instance of τ and is uncorrelated from other instances, its first time-derivative $\frac{d}{dt}g(t, \tau)$ is independent of both the process $g(t, \tau)$ and its second time-derivative $\frac{d^2}{dt^2}g(t, \tau)$ at each instance of τ (Lutes and Sarkani, 2004). However, the process and the second time-derivative are not mutually independent. Therefore, the MZC rate of acceleration ($\nu_a(\tau)$), velocity ($\nu_v(\tau)$), and displacement ($\nu_d(\tau)$) processes can be respectively simplified as (Lutes and Sarkani, 2004):

$$\nu_a(\tau) = \frac{1}{2\pi} \frac{\sigma_{\dot{g}}(\tau)}{\sigma_g(\tau)} \quad (14)$$

$$\nu_v(\tau) = \frac{1}{2\pi} \frac{\sigma_{\dot{g}}(\tau)}{\sigma_{\bar{g}}(\tau)} \quad (15)$$

$$\nu_d(\tau) = \frac{1}{2\pi} \frac{\sigma_{\ddot{g}}(\tau)}{\sigma_{\bar{g}}(\tau)} \quad (16)$$

where $\sigma_g(\tau)$ and $\sigma_{\dot{g}}(\tau)$ are the standard deviation of $g(t, \tau)$ and its first time-derivative at the instance of τ , respectively. $\sigma_{\bar{g}}(\tau)$ and $\sigma_{\ddot{g}}(\tau)$ are the standard deviation of the first and second integrals of $g(t, \tau)$, respectively. Furthermore, the analytical expression for the MLE rate of acceleration ($\eta_a(\tau)$), velocity ($\eta_v(\tau)$), and displacement ($\eta_d(\tau)$) processes can be described as (Lutes and Sarkani, 2004):

$$\eta_a(\tau) = \frac{1}{2\pi} \frac{\sigma_{\ddot{g}}(\tau)}{\sigma_{\dot{g}}(\tau)} \quad (17)$$

$$\eta_v(\tau) = \frac{1}{2\pi} \frac{\sigma_{\ddot{g}}(\tau)}{\sigma_g(\tau)} \quad (18)$$

$$\eta_d(\tau) = \frac{1}{2\pi} \frac{\sigma_{\ddot{g}}(\tau)}{\sigma_{\bar{g}}(\tau)} \quad (19)$$

where $\sigma_{\ddot{g}}(\tau)$ is the standard deviation of the second time-derivative of the process $g(t, \tau)$. It is noteworthy that the MPMNM rate of a process (i.e. acceleration, velocity, or displacement) is also formulated as half of the difference between the MLE and MZC rate of the process using Equations 14–19. In the above expressions, $\nu_a(\tau) = \eta_v(\tau)$ and $\nu_v(\tau) = \eta_d(\tau)$, which are also observed numerically in recorded ground motions (refer to the example in Figure 9). Moreover, $\eta_a(\tau)$ is related to the first and second time-derivatives of the process

$g(t, \tau)$. As discussed earlier, the inclusion of this relationship provides no benefits to the model; therefore, it is omitted from the fitting procedures.

This study uses Equations 14–16 as the fitting criterion to describe the time-varying frequency content of a target recorded motion. As a result, the MZC of acceleration and displacement processes serve as proxies for the upper and lower dominant frequencies, respectively. To avoid redundancy, the MZC of the velocity process (alongside acceleration and displacement MZC) adequately accounts for the bandwidth of the process, implicitly capturing the MPMNM of the velocity and displacement processes.

In Equation 3, the term $q(\tau, \alpha)w(\tau)$ at each instance of τ has the same value for $\sigma_g(\tau)$, $\sigma_{\dot{g}}(\tau)$, $\sigma_{\ddot{g}}(\tau)$, and $\sigma_{\bar{g}}(\tau)$ and can be ignored while using Equations 14–16. This can also be interpreted physically, as scaling a process is not expected to affect the count of its zero crossings or local extrema. A simple alternative for determining the standard deviations of the process $g(t, \tau)$, its derivatives, and integrals in Equations 14–16 is to use the expected evolutionary PSD of the process $g(t, \tau)$. Because $g(t, \tau)$ is equivalent to the term inside the integral in Equation 4, the expected PSD (ϕ_E) can be expressed as follows:

$$\phi_E(\bar{\omega}, \tau, \alpha, \lambda(\tau)) = q(\tau, \alpha)^2 \frac{|H(\bar{\omega}, \lambda(\tau))|^2}{\int_{-\infty}^{+\infty} |H(\bar{\omega}, \lambda(\tau))|^2 d\bar{\omega}} \quad (20)$$

Therefore, the variance of the process $g(t, \tau)$, its time-derivatives, and integrals can be expressed as follows:

$$\sigma_g^2(\tau) = \int_{-\infty}^{\infty} \phi_E(\bar{\omega}, \tau, \lambda(\tau)) d\bar{\omega} \quad (21)$$

$$\sigma_{\dot{g}}^2(\tau) = \int_{-\infty}^{\infty} \bar{\omega}^2 \phi_E(\bar{\omega}, \tau, \lambda(\tau)) d\bar{\omega} \quad (22)$$

$$\sigma_{\ddot{g}}^2(\tau) = \int_{-\infty}^{\infty} \bar{\omega}^{-2} \phi_E(\bar{\omega}, \tau, \lambda(\tau)) d\bar{\omega} \quad (23)$$

$$\sigma_{\bar{g}}^2(\tau) = \int_{-\infty}^{\infty} \bar{\omega}^{-4} \phi_E(\bar{\omega}, \tau, \lambda(\tau)) d\bar{\omega} \quad (24)$$

Because τ is a dummy variable of time, the MZC rate of the acceleration process $\ddot{x}(t)$ and its corresponding velocity and displacement processes are essentially the same as in Equations 14–16 by substituting τ with t . Note that when using Equations 21–24 for a discrete system corresponding to a recorded accelerogram, the bound of the integral is limited to the corresponding Nyquist frequency to avoid aliasing (Shannon, 1949).

Time-varying energy content

Fitting to the time-varying energy content follows the same procedure used by Rezaeian and Der Kiureghian (2008) but with an added number of parameters for the time-

modulating function. For a ground motion time-series $a(t)$, CE can directly represent the time-varying energy content and is expressed as follows:

$$CE_a(t) = \int_0^t a^2(\tau) d\tau \quad (25)$$

Similarly, the expected CE of the process $\ddot{x}(t)$ defined in Equation 3 or 4 can be formulated as follows:

$$CE_{\ddot{x}}(t) = E \left[\int_0^t \ddot{x}(\tau)^2 d\tau \right] = \int_0^t q^2(\tau, E_t, D, \vec{\alpha}, \vec{c}, \vec{p}) d\tau \quad (26)$$

This means the modulating function can fully and independently describe the energy content of the process.

Parameter identification

Similar to Rezaeian and Der Kiureghian (2008), the parameters of both the filter and the time-modulating function can be independently identified by matching the characteristics outlined in the previous section with those of a target accelerogram. Unlike Rezaeian and Der Kiureghian (2008), in which filter parameters were identified by matching to the MZC and MPMNM of the acceleration process, we identify the parameters by matching to the MZC of the acceleration, velocity, and displacement processes. This adjustment to the objective functions is necessary to converge to an optimized set of model parameters given their increased number and is more intuitive given their segregation into lower and upper limits of the dominant frequency in Equations 9–10.

Parameter identification for the filter function

The filter parameters (ω_{l0} , ω_{ln} , ω_{u0} , ω_{un} , ζ_{l0} , ζ_{ln} , ζ_{u0} , ζ_{un}) have interrelated effects. For instance, lower damping ratios enable oscillation with specific dominant frequencies to persist in time, leading to dominant frequencies with higher amplitude and a narrower bandwidth in the proposed filter. Hence, it is reasonable to concurrently identify the filter frequencies and damping ratios. This is achieved by matching the cumulative expected MZC of the acceleration, velocity, and displacement processes of the model with those of the corresponding target motion. The squared differences between the corresponding terms are minimized as expressed below:

$$\begin{aligned} & [\omega_{l0}, \omega_{ln}, \omega_{u0}, \omega_{un}, \zeta_{l0}, \zeta_{ln}, \zeta_{u0}, \zeta_{un}] \\ &= \underset{\text{argmin}}{\left[\int_0^D \left[\int_0^t \nu_a(\tau) d\tau - N_{\nu_a}(t) \right]^2 dt + \int_0^D \left[\int_0^t \nu_v(\tau) d\tau - N_{\nu_v}(t) \right]^2 dt \right.} \\ & \quad \left. + \int_0^D \left[\int_0^t \nu_d(\tau) d\tau - N_{\nu_d}(t) \right]^2 dt \right] \end{aligned} \quad (27)$$

where $N_{\nu_a}(t)$, $N_{\nu_v}(t)$, and $N_{\nu_d}(t)$ are the cumulative MZC of the acceleration, velocity, and displacement time-series of the target motion, respectively. In this study, we use the SciPy library in Python to identify model parameters. One advantage of the proposed filter over existing filters (e.g. Vlachos et al., 2016; Waezi et al., 2021) is that $\omega_u(\tau)$ and $\omega_l(\tau)$ are

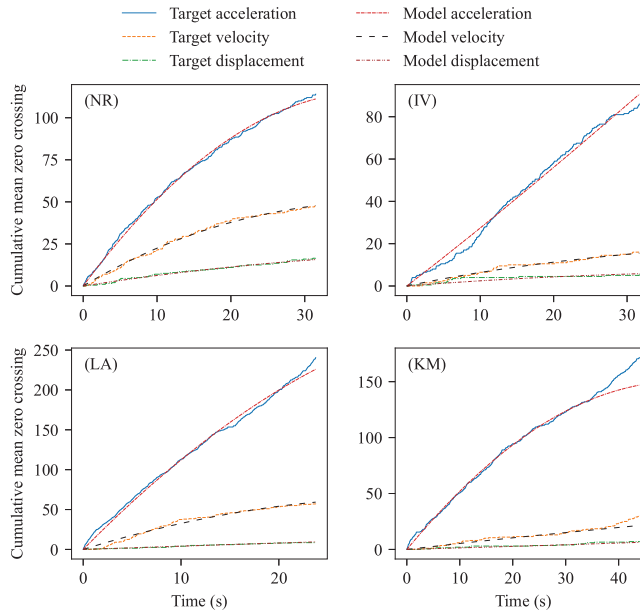


Figure 10. Cumulative mean zero crossings of the target acceleration, velocity, and displacement time-series, along with their corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions.

primarily influenced by the rate of MZC of the target acceleration and displacement time-series, respectively. This underlying relation facilitates understanding the variation of these parameters from the corresponding MZC curves.

Figure 10 illustrates the cumulative MZC of the example target accelerograms, their velocity and displacement time-series, and the corresponding fitted models. Identified filter parameters are shown in Table 2. The MZC rate of the target motions (i.e. acceleration, velocity, and displacement), which is the slope of the corresponding curves in Figure 10, fluctuates over time for all selected accelerograms. This means that the dominant frequencies in the target motions vary with time. In the case of the Imperial Valley (IV) (refer to Table 1) accelerogram with a shorter rupture distance ($R_{rup} = 0.07$ km) and a soft soil ($V_{S30} = 264 \frac{m}{s}$), the fluctuations are more apparent than in other cases. The value of the upper dominant frequency (ω_u), which follows the acceleration MZC rate, generally exhibits a downward trend. However, it shows an upward trend in the IV case. The higher MZC in the acceleration time-series signifies the prevalence of higher frequencies. On the other hand, a reduced MZC in the displacement time-series indicates the predominance of lower frequencies. Thus, the MZC rate in each acceleration or displacement time-series is related to the corresponding upper or lower dominant frequencies, respectively. The L'Aquila (LA) accelerogram shows an elevated rate of MZC in the acceleration time-series compared to other cases. Consequently, their upper dominant frequency value surpasses that of other cases as shown in Table 2. The lower dominant frequency (ω_l) corresponds to a decrease in rate. It is noteworthy that damping ratios (i.e. ζ_u , ζ_l) regulate the bandwidth to accurately capture the intermediate frequencies, as shown by the rate of MZC for velocity. To verify the adequacy of the considered MZC rates as criteria, Figure 11 compares the MPMNM of the target velocity and displacement time-series to those of the

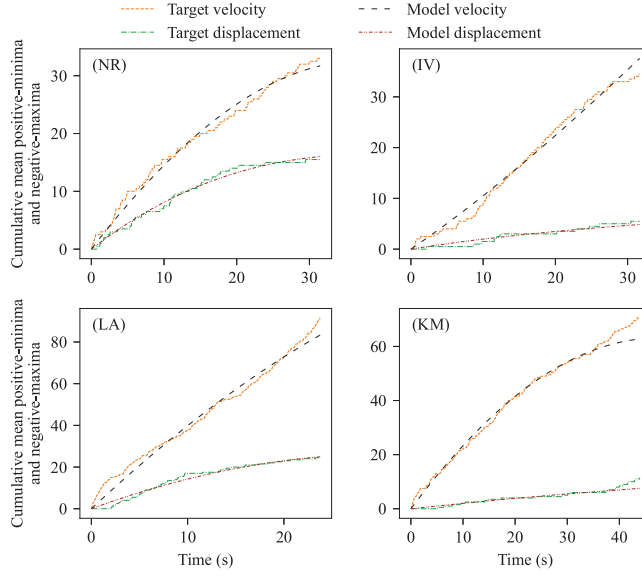


Figure 11. Cumulative mean positive-minima and negative-maxima of the target velocity and displacement time-series, along with their corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions.

corresponding models. Figure 11 shows that the MPMNM of displacement as a proxy for low to intermediate frequencies and the MPMNM of velocity as a proxy for intermediate to high frequencies are effectively incorporated in the fitted model.

Errors for various validation metrics (e.g. MZC of acceleration) can be calculated using the following expression:

$$\epsilon_M = \frac{\int_0^D |M_{Target}(t) - E[M_{Model}(t)]| dt}{\int_0^D M_{Target}(t) dt} \quad (28)$$

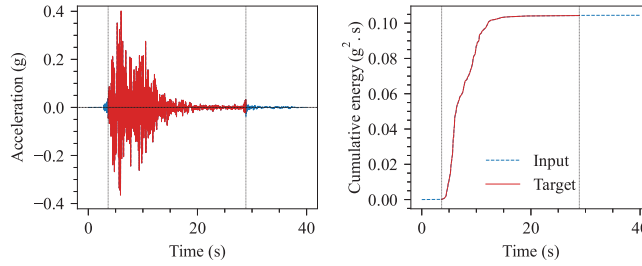
where M refers to the relevant validation metric for both the target and expected model processes, and ϵ_M is the error value for that specified validation metric M . Therefore, to quantify the errors associated with the cumulative MZC of the acceleration (ϵ_{MZC_a}), velocity (ϵ_{MZC_v}), and displacement (ϵ_{MZC_d}), as well as MPMNM of the velocity (ϵ_{MPMNM_v}) and displacement (ϵ_{MPMNM_d}), we use Equation 28 by substituting these respective metrics for the variable M for both target and model processes. These error measures are reported in Table 2 for each of the example target motions. The increased error values associated with the MZC of displacement time-series stem from the staggered shape of their curves. These curves are not as smooth as those for acceleration time-series due to the reduced count of MZC, resulting in increased relative errors.

Parameter identification for the time-modulating function

The parameters of the time-modulating function (E_t , D , α_1 , p_1 , c_1 , p_2 , c_2) are identified by matching the expected CE of the process with that of the target accelerogram over the

Table 2. Parameters of the filter function fitted to target accelerograms

ID	Frequency parameters (Hz)				Damping ratio parameters				Errors				
	ω_{u0}	ω_{un}	ω_{l0}	ω_{ln}	ζ_{u0}	ζ_{un}	ζ_{l0}	ζ_{ln}	ϵ_{MZC_a}	ϵ_{MZC_v}	ϵ_{MZC_d}	ϵ_{MPMNM_v}	ϵ_{MPMNM_d}
NR	5.88	1.21	0.69	0.32	0.50	0.52	1.00	0.36	0.02	0.03	0.05	0.04	0.04
IV	2.70	3.01	0.29	0.10	0.70	2.43	0.47	0.64	0.05	0.06	0.19	0.06	0.16
LA	12.35	6.78	0.41	0.37	0.53	0.58	1.70	0.27	0.03	0.06	0.06	0.06	0.06
KM	5.77	0.94	0.14	0.16	2.88	0.10	1.03	0.26	0.04	0.12	0.16	0.04	0.13

**Figure 12.** The L'Aquila (LA) target motion corresponding to 0.1%–99.9% of the total energy of the input LA accelerogram.

total duration. This can be done by minimizing the squared difference between the two CE terms, expressed as follows:

$$(\alpha_1, p_1, c_1, p_2, c_2) = \operatorname{argmin} \int_0^D [CE_{\ddot{x}}(t) - CE_a(t)]^2 dt \quad (29)$$

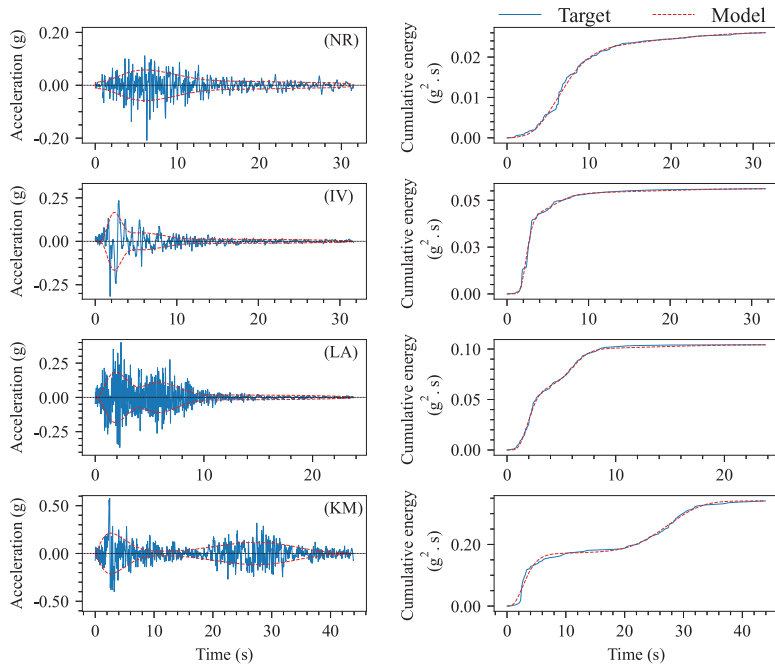
To ensure a unique solution for cases with two strong phases, it is assumed that the first phase precedes the second, imposing the constraint $p_1 \leq p_2$ on the parameters. For $p_1 = p_2$, a single strong phase can be selected by setting $n = 1$ in Equation 13. However, in simulating our selected datasets, we use $n = 2$ to maintain consistency in the results.

Recorded ground motions often exhibit leading or trailing zeros or negligible amplitudes, resulting in different durations. To avoid these tails that have a negligible contribution to the total energy, the target motion is defined as that corresponding to 0.1%–99.9% of the total energy of the input recorded accelerogram. Figure 12 illustrates the target motion and the corresponding duration for LA as an example (refer to Table 1).

Errors associated with CE are determined by substituting CE for the variable M in Equation 28. The identified parameters of the modulating function and the corresponding errors are shown in Table 3. Figure 13 illustrates the modulating function and CE of the model when fitting to the corresponding target accelerograms. The fit with slight error demonstrates the capability of the time-modulating function in modeling various shapes of ground motions, including pulse-like and double strong phase motions. The arrival of energy in a single strong phase is observable in NR and IV with a more abrupt arrival in the latter (refer to Table 1). In NR, 76.6% of the total energy is smoothly distributed

Table 3. Parameters of the modulating function fitted to target accelerograms

ID	Modulating function parameters							Errors
	E_t ($g^2 \cdot s$)	D (s)	α_1	p_1	c_1	p_2	c_2	E_{CE}
NR	0.026037	31.48	0.766	0.193	22.8	0.193	0.7	0.013
IV	0.056155	31.67	0.754	0.074	182.5	0.178	41.9	0.008
LA	0.104165	23.76	0.602	0.079	74.6	0.247	69.3	0.011
KM	0.341184	43.90	0.497	0.060	49.5	0.625	19.0	0.022

**Figure 13.** Fitted modulating function to target accelerograms and their respective cumulative energy. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions.

around the first peak time ($D.p_1 = 6.1$ s), featuring a low value for the energy concentration ($c_1 = 22.8$). In contrast, IV exhibits a higher energy concentration ($c_1 = 182.5$) around the first peak time ($D.p_1 = 2.3$ s) that carries 75.4% of the total energy. In the cases of LA and KM, two distinct strong phases are evident. The first phase in LA carries 60.2% of the total energy with a peak time of $D.p_1 = 1.9$ s. In KM, the first phase carries 49.7% of the total energy with a peak time of $D.p_1 = 2.6$ s and the energy concentration of $c_1 = 49.5$. As can be observed, the second strong phase with a peak time of $D.p_2 = 27.4$ s is less concentrated ($c_2 = 19.0$).

Validation

Before simulations can be used in engineering applications, it is crucial to thoroughly validate the simulation methodologies and their outcomes by assessing their effectiveness in

predicting both seismological and engineering demand parameters (Karimzadeh, 2019; Rezaeian et al., 2024; Smerzini et al., 2024). Such validation ensures that the model accurately represents the main characteristics of recorded ground motions, thus demonstrating the reliability of the synthetic motions for engineering applications. Validation procedures in the literature are often based on characteristics that are expected to affect engineering systems, such as elastic response spectra or other ground motion validation metrics (Arslan Kelam et al., 2022; Askan et al., 2017; Burks and Baker, 2014; Goulet et al., 2015; Karimzadeh et al., 2020, 2023, 2024a; Olsen and Mayhew, 2010; Rezaeian and Der Kiureghian, 2010). In this study, we first perform a visual comparison between the ground motion time-series, CE, Fourier amplitude, and elastic response spectra of the four example target accelerograms and those of simulations generated from the fitted stochastic model parameters described in previous sections. Next, we quantitatively assess the efficiency of the improved stochastic model using 276 horizontal components from the selected datasets.

Visual qualitative comparisons of target and simulated motions

Given the identified model parameters in the previous section, the improved stochastic model can directly simulate ground motion time-series without added deterministic components or any additional post-processing. Figures 14–16 show the acceleration, velocity, and displacement time-series for the four example target motions listed in Table 1, along with two realizations for each, obtained from stochastic simulations as described in previous sections. In Figure 14, the stochastic simulations resemble target accelerograms in terms of the overall waveforms, intensity levels, durations, and frequency contents. In particular, simulations of LA and KM exhibit multiple strong phases like their respective target motions. This illustrates the effectiveness of the proposed modulating function in accurately accounting for the arrival of energy in multiple strong phases. The simulated velocity time-series in Figure 15 are also like those of target motions, capturing the overall waveforms, pulse timings, and pulse intensities without considering additional explicit constraints on these features. For example, in IV, a velocity pulse is observed at $t = 2.73$ s with a peak ground velocity (PGV) of $72.96 \frac{\text{mm}}{\text{s}}$. Consequently, the stochastic simulations that aim to provide possible realizations of the target motion exhibit a velocity pulse with comparable timing and intensity. Thus, the proposed filter has adequately captured the low-frequency contents, successfully representing near-fault pulse-like ground motions. The displacement time-series, illustrated in Figure 16 further underscores these observations, revealing analogous waveforms and intensity characteristics.

Figure 17 shows the CE of the target accelerograms and 20 random realizations from simulations. CEs of the target accelerograms lie within one standard deviation of the corresponding simulated motions. This indicates that the target accelerogram can be regarded as one reasonable realization of the stochastic model. It is noteworthy that CE is non-linear with respect to the amplitude of each simulation, and the amplitude is influenced by the random component of modulated white noise, $q(\tau, \alpha)w(\tau)$, as shown in Equations 3 and 4. This explains the varying levels of simulation variability across different target records. For instance, in the case of IV, where energy arrives rapidly, the modulating function exhibits a sharper and higher peak amplitude. This causes $q(\tau, \alpha)w(\tau)$ to vary more widely, leading to an increased variability in the total energy. If a closer match between the total energy of each synthetic motion and the target is desired, simulations with total energy less than half or greater than twice (or any other threshold) of the target can be discarded as also suggested by Dabaghi and Der Kiureghian (2017). Although the model

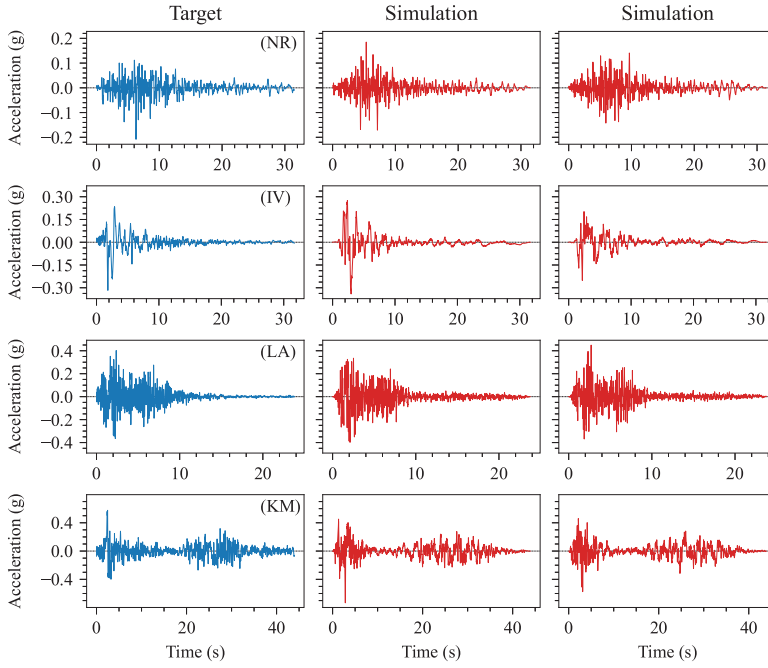


Figure 14. Target accelerograms and two alternative simulations using the corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions.

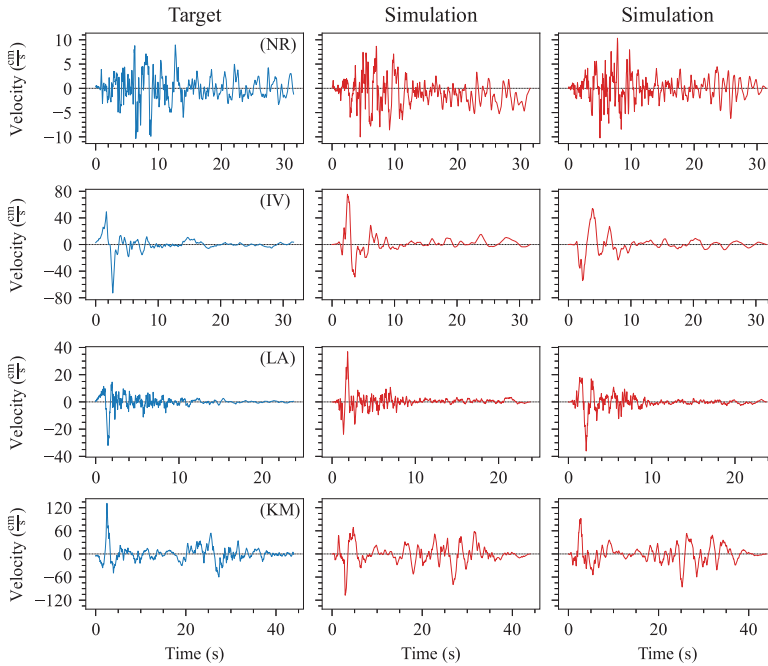


Figure 15. Velocity time-series and two alternative simulations using the corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions.

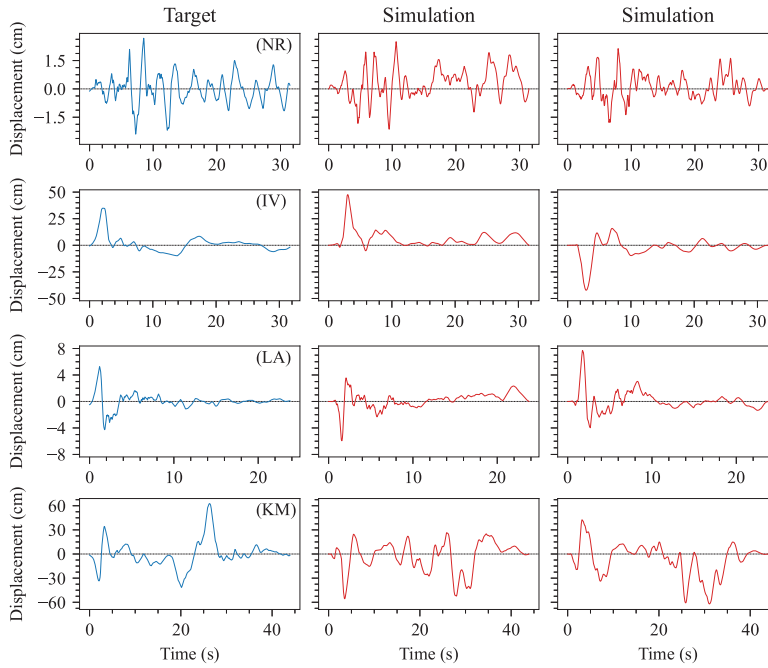


Figure 16. Displacement time-series and two alternative simulations using the corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions.

provides variability, the average of simulations matches the target motion. The slight mismatch between the mean CE of the simulations (Figure 17) and the fitted modulating functions (Figure 13) is attributed to the underlying white noise, which diminishes by considering a higher number of simulations. This indicates the absence of average bias, achieved by omitting the post-processing required in other studies (e.g. Dabaghi and Der Kiureghian, 2017; Rezaeian and Der Kiureghian, 2008; Waezi and Balzadeh, 2022).

For further comparison, the FAS and 5%-damped response spectral acceleration (SA) for the target accelerograms and 20 random realizations from their simulations are shown in Figures 18 and 19, respectively. The FAS is smoothed using a simple moving average with a window size of 9 points solely for visual comparison. FAS for simulations can be represented directly using Equation 4. The FAS and SA of the target motions fall within the ranges spanned by those of the simulated motions nearly across all frequency and period ranges. Therefore, the target motions can be considered reasonable realizations of the stochastic model for the given parameters. The rather smooth shape of the mean FAS and SA is attributed to the linear functional form of filter parameters, aiming to capture evolutionary dominant frequencies and bandwidth parameters in an average manner. Consequently, the linear functions do not capture all local fluctuations of these parameters. For instance, in the IV, a local peak in SA is evident around the period of 0.7 s, and the mean SA does not exhibit this distinct peak. Yet, such peaks occur within the range of simulations due to the variability. Nevertheless, given the intricate nature of ground motions, different functional forms can be used to represent the very local

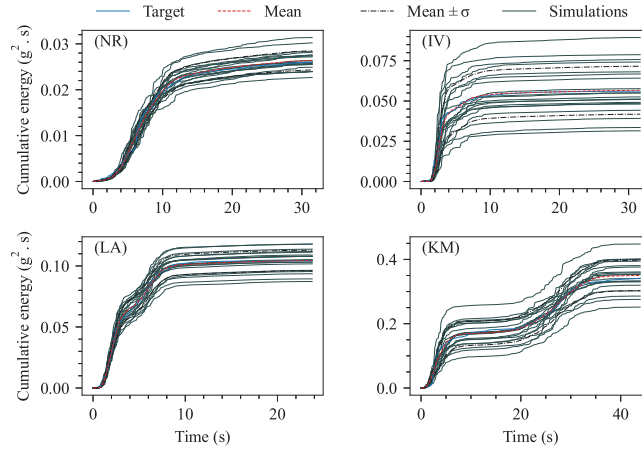


Figure 17. Cumulative energy of the target accelerograms and 20 random simulations using the corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions. The standard deviation is denoted as σ .

Table 4. Errors associated with Fourier amplitude spectrum (FAS) and spectral acceleration (SA) for Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions

Error metric	Identifier			
	NR	IV	LA	KM
E_{FAS}	0.53	0.48	0.46	0.49
E_{SA}	0.17	0.15	0.33	0.29

variation of these filter parameters. However, this is not pursued in this study and will likely lead to more parameters and more complex models.

Next, quantitative error measures for the FAS (ϵ_{FAS}) and SA (ϵ_{SA}) are respectively determined using the following formulas:

$$\epsilon_{FAS} = \frac{\int_{0.1}^{25 \text{ Hz}} |FAS_{Target}(\bar{f}) - E[FAS_{Model}(\bar{f})]| d\bar{f}}{\int_0^{25 \text{ Hz}} FAS_{Target}(\bar{f}) d\bar{f}} \quad (30)$$

$$\epsilon_{SA} = \frac{\int_{0.04}^{10 \text{ s}} |SA_{Target}(T) - E[SA_{Model}(T)]| dT}{\int_0^{10 \text{ s}} SA_{Target}(T) dT} \quad (31)$$

These errors are reported in Table 4, and alongside error measures in previous sections, they can provide an understanding of the relationship between the quantitative errors and visual assessments. These errors exhibit increased values due to the abrupt amplitude variations in the target FAS and SA. Nonetheless, based on visual assessment, the simulation model effectively captures the overall behavior.

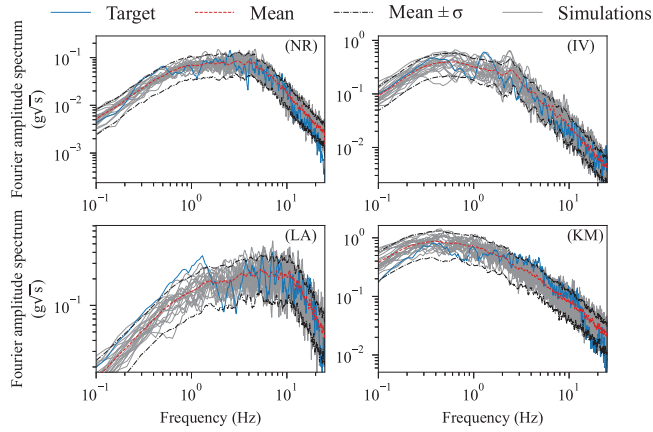


Figure 18. Fourier amplitude spectrum of the target accelerograms and 20 alternative simulations using the corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions. The standard deviation is denoted as σ .

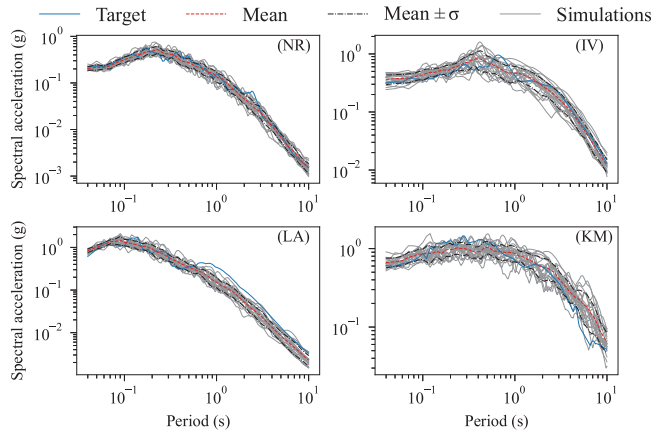


Figure 19. Spectral acceleration of the target accelerograms and 20 alternative simulations using the corresponding fitted models. The panels correspond to Northridge (NR), Imperial Valley (IV), L'Aquila (LA), and Kahramanmaraş (KM) ground motions. The standard deviation is denoted as σ .

Quantitative model evaluation for the selected datasets

Empirical cumulative distribution function of errors. Quantitative evaluation of the improved model is performed separately for the Kahramanmaraş earthquake ($M_w = 7.8$) and pulse-like NGA-West2 datasets by comparing the characteristics of time-series, Fourier amplitude, elastic response spectra, and peak ground motion parameters of 20 random simulations with those of recorded motions. Equation 28 is used for the comparisons in terms of CE, MZC of acceleration, velocity, and displacement, as well as MPMNM of velocity and displacement. Equations 30 and 31 are used to compare the FAS (0.1–25 Hz) and SA (0.04–10 s), respectively. Similarly, Equation 31 is used for the elastic response spectra of velocity (SV) and displacement (SD) by substituting them for SA, respectively. In addition,

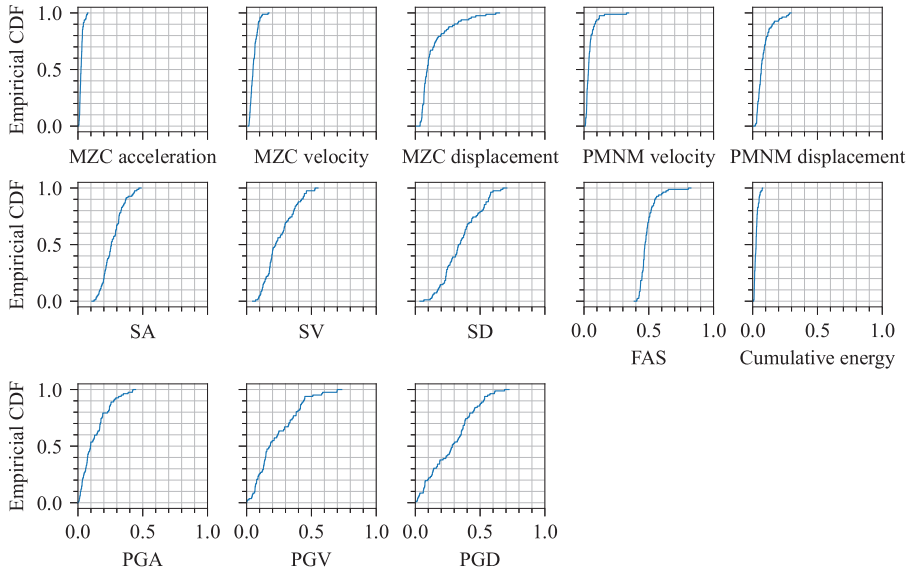


Figure 20. Empirical cumulative distribution (CDF) function of various metrics errors for the Kahramanmaraş, Türkiye, earthquake ($M_w = 7.8$) dataset.

PGA, PGV, and peak ground displacement (PGD) are compared using Equation 32 by substituting each for the variable K :

$$\epsilon_K = \frac{|K_{target} - E[K_{sim}]|}{K_{target}} \quad (32)$$

Figures 20 and 21 illustrate the empirical cumulative distribution function (CDF) of the error values for these 13 metrics, corresponding to the Kahramanmaraş earthquake and NGA-West2 datasets, respectively. These figures illustrate the percentage of simulations (vertical axis) that fall below each error value (horizontal axis), providing an overview of the model's performance across the two datasets. Based on the previous assessments of the four example target motions, several observations are noteworthy. The very low error value based on CE indicates the accuracy of the modulating function in capturing the energy buildup in recorded motions of various earthquake scenarios within the datasets. Similarly, the low error values associated with the MZC of acceleration, velocity, and displacement for a large percentage of simulations indicate that the high, intermediate, and low-frequency content of the simulations aligns well with the target records in both datasets. Nonetheless, the error is higher for displacement MZC. This may be attributed to the more staggered curves caused by a reduced count of zero crossings or the non-linear variation of lower dominant frequencies in a small subset of the recordings. The empirical CDF of MPMNM curves indicates that the bandwidth of the simulations aligns with the recorded datasets. For SA and SV, 90% of simulations have errors below 0.5, and SD errors are higher, consistent with the corresponding MZC errors. A similar observation is made for errors associated with PGA and PGV, with higher errors observed in PGD. For FAS, 90% of simulations have errors below 0.6, reflecting consistently higher error values due to abrupt amplitude variations, as noted in the previous section (e.g. Table 4 and

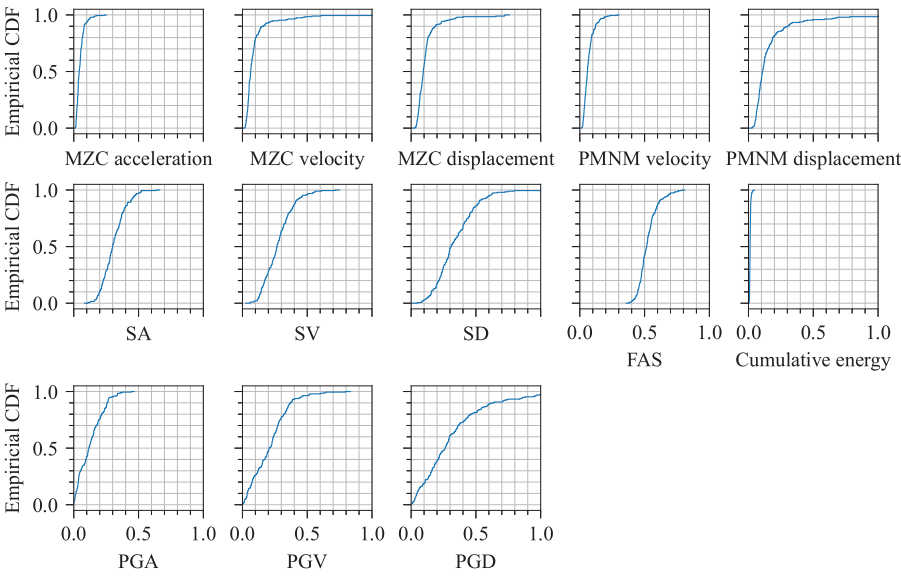


Figure 21. Empirical cumulative distribution (CDF) function of various metrics errors for the NGA-West2 dataset.

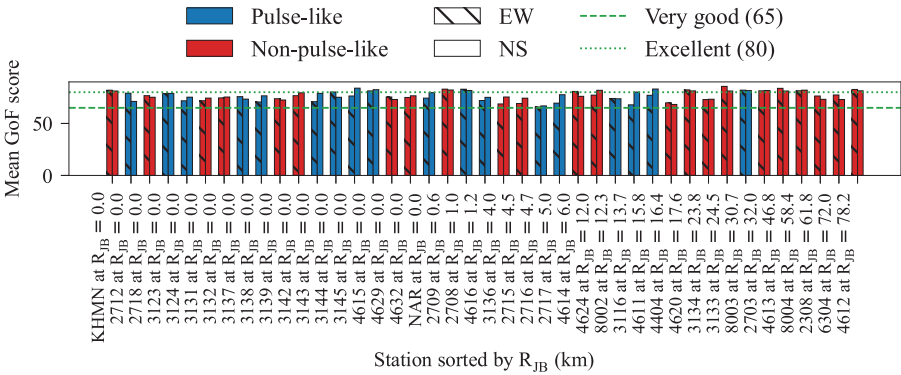


Figure 22. Mean goodness-of-fit (GoF) scores for the east-west (EW) and north-south (NS) components across stations for the Kahramanmaraş, Türkiye, earthquake dataset.

Figure 18). The results indicate that the improved simulation model performs consistently and effectively for both selected datasets.

Goodness-of-fit score. For further evaluation, the GoF score (Olsen and Mayhew, 2010) is used to quantitatively compare a random simulation with its corresponding target record from the selected datasets. GoF scores are computed for the 13 considered metrics in the previous section. Figure 22 shows the mean GoF scores of the 13 metrics for each component at each station for the Kahramanmaraş earthquake dataset. The GoF scores are consistently above 65, indicating a very good fit for each target record (Olsen and Mayhew, 2010). The consistent GoF scores across different R_{JB} values, as well as for pulse-like and

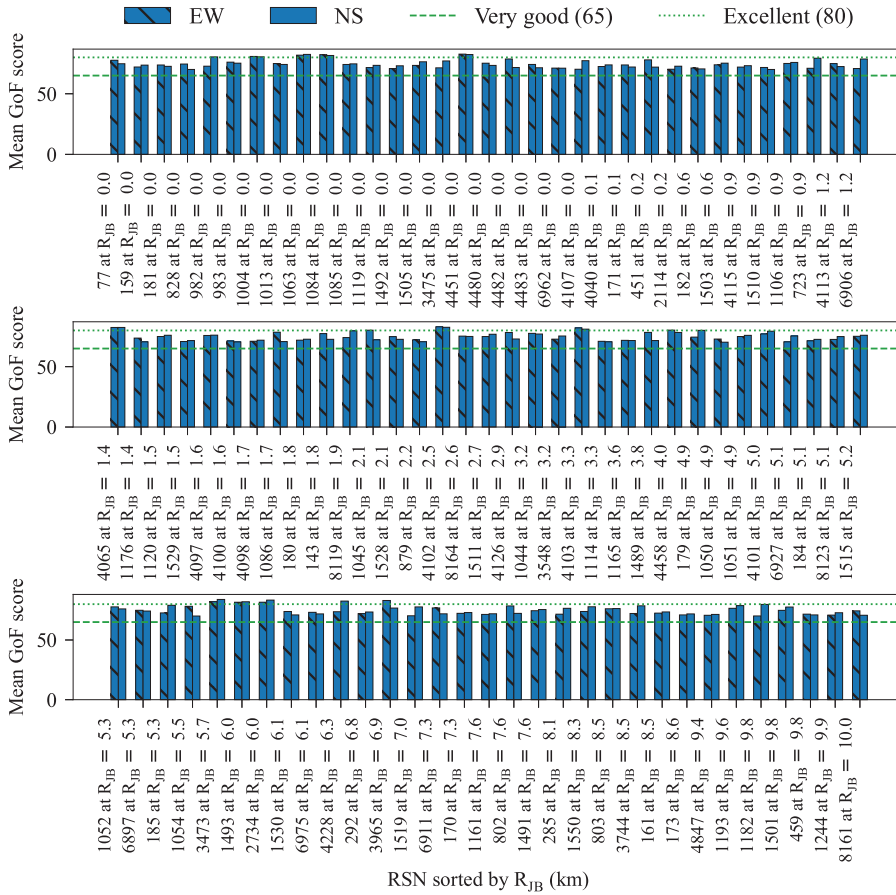


Figure 23. Mean goodness-of-fit (GoF) scores for the east-west (EW) and north-south (NS) components across record sequence number (RSN) for the NGA-West2 dataset.

non-pulse-like motions of this event, indicate that the improved model performs efficiently and consistently across the dataset, from near-fault pulse-like to far-field non-pulse-like motions. Figure 23 shows similar consistent performance for the NGA-West2 dataset, indicating that the improved model effectively captures ground motions with notable low-frequency content, typical of pulse-like motions.

To evaluate the performance of the improved model for each metric individually, the mean GoF score for each metric is calculated within each dataset. Figures 24 and 25 show the mean GoF scores of each metric, averaged across all stations for the Kahramanmaraş earthquake and pulse-like NGA-West2 datasets, respectively. The results show that the simulations achieve a very good fit for all metrics, except for FAS, which shows a fair fit ($\text{GoF} > 45$), mostly due to the abrupt amplitude variations discussed previously.

Inter-period correlation of spectral acceleration ordinates. Furthermore, the correlations between the spectral acceleration ordinates of the recorded motions and the corresponding simulated motions within periods of 0.04–10 s are compared. These correlations can be an important indicator of the seismic demand in multi-degree-of-freedom (MDOF) systems

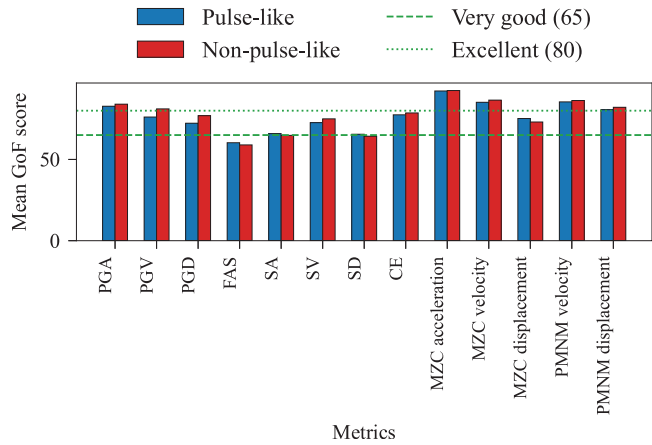


Figure 24. Mean goodness-of-fit (GoF) scores for each metric averaged across all stations, for pulse-like and non-pulse-like motions from the Kahramanmaraş, Türkiye, earthquake dataset.

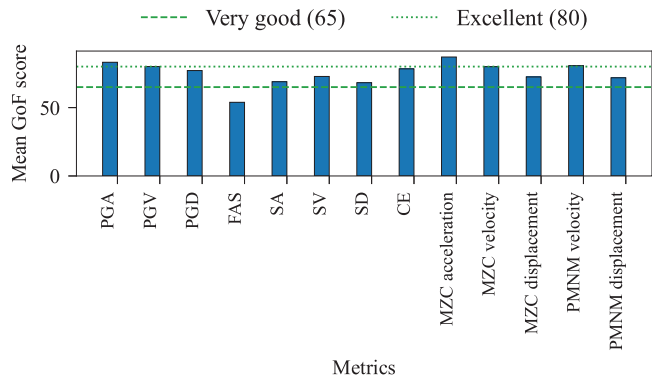


Figure 25. Mean goodness-of-fit (GoF) scores for each metric, averaged across all stations, for pulse-like motions from the NGA-West2 dataset.

that are excited at various vibration periods (Burks and Baker, 2014). Figure 26 illustrates these correlations separately for the Kahramanmaraş earthquake and NGA-West2 datasets. It includes recorded motions and individual simulated motions (one random simulation per target record). It is noteworthy that considering a specific set of model parameters corresponding to a station (a recorded motion), a set of simulations in that station does not show a significant correlation due to the stochasticity introduced by underlying white noise. However, when different stations (different sets of parameters) are considered, even with a single random simulation, correlations emerge that align with those observed in real records.

Overall, the simulated and target ground motions exhibit similar evolutionary features. The identified model parameters implicitly consider the properties of the earthquake and site corresponding to a target motion. Thus, the improved model can generate a set of

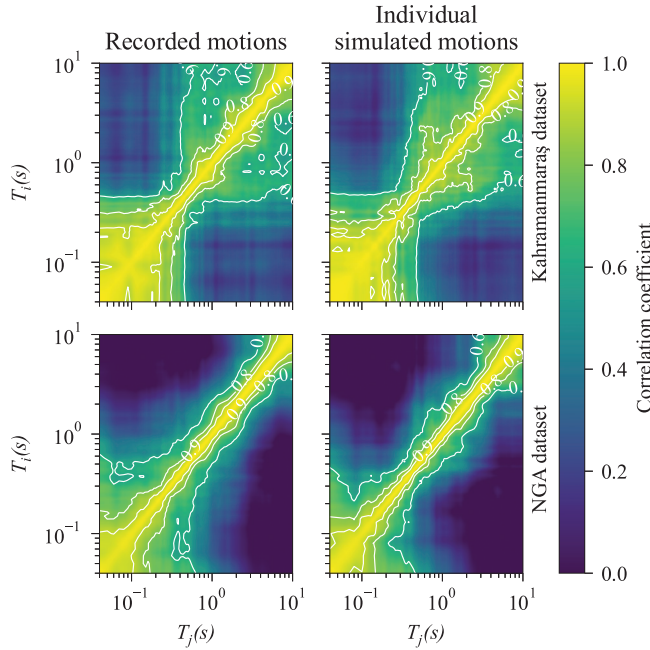


Figure 26. Correlations between spectral acceleration ordinates of recorded and individual simulated motions for the pulse-like and non-pulse-like Kahramanmaraş earthquake dataset (top panel) and pulse-like NGA dataset (bottom panel).

ground motions suitable for designing and evaluating a structure for that specific earthquake and site characteristics.

In future studies, a database of model parameters can be obtained by fitting to many rotated accelerograms with known earthquake and site characteristics. However, further studies are necessary to determine whether these model parameters exhibit a meaningful physical relationship with those characteristics. If such a relationship is established, predictive equations for the model parameters can be developed based on the considered earthquake and site characteristics. The outcome enables the simulation of horizontal orthogonal ground motions for a specified earthquake and site characteristics, using the improved stochastic model following similar procedures in previous studies (Dabaghi and Der Kiureghian, 2018; Rezaeian and Der Kiureghian, 2012; Sgobba et al., 2011).

Conclusion

This study improves the fully non-stationary site-based stochastic ground motion simulation model developed by Rezaeian and Der Kiureghian (2008). The model is improved to have a broader applicability range by representing multiple strong phases in ground motions as well as low-frequency contents which might be due to near-fault directivity effects or potential long-period deep basin effects without adding deterministic velocity pulse model or post-processing filters. An adjustment to the formulation is proposed and is presented in both time and frequency domains. A new modulating function is proposed that can account for the arrival of energy in multiple strong phases. Moreover, a new band-pass filter with additional parameters representing evolutionary bandwidths and

dominant frequencies of the ground motion is proposed. This enhancement aims to better represent the lower-frequency content of ground motions, thereby implicitly accounting for the near-fault rupture directivity pulses and potential long-period basin effects. The improved formulation of the model, along with the proposed filter, omits the need for additional post-processing. This consequently eliminates the average bias in the total energy of synthetics and the need for correction factors present in previous site-based stochastic models. On the downside, the additional filter parameters necessitate an adjustment to the objective functions used in parameter identification. However, the simplicity of the model is still preserved through separable temporal and spectral non-stationarities and all parameters can be identified by matching to the time domain CE of the acceleration and the mean zero-level crossings of the acceleration, velocity, and displacement time-series of a target motion.

The proposed simulation model's performance was evaluated, and the expanded applicability range was tested by simulating the February 6, 2023, Kahramanmaraş (Türkiye) earthquake ($M_w = 7.8$). The model was further validated for its ability to represent low-frequency content using a near-fault pulse-like dataset from the NGA-West2 database. These datasets represent a diverse set of earthquakes and site characteristics, exhibiting near-fault, far-field, pulse-like, non-pulse-like, single, and multiple strong phase motions. To discuss the model in more detail, four example target accelerograms were selected from the Northridge 1994, Imperial Valley 1979, L'Aquila 2009, and Kahramanmaraş 2023 earthquake events. The validation of the improved stochastic model was performed through both visual and quantitative comparisons. The visual comparisons focused on the four selected records, and quantitative assessments considered both datasets, evaluating various validation metric errors, GoF scores, and inter-period correlations. The results demonstrated an overall good agreement between the simulated and observed motions, with favorable inter-period correlations, and satisfactory errors across the datasets. The GoF scores consistently ranged from very good to excellent for all stations, regardless of ground motion characteristics.

These results indicate that, given the identified parameters for specific earthquake and site characteristics, with the proposed improvements the stochastic model can efficiently simulate a wide range of ground motions, including near-fault pulse-like and multiple strong phase motions, without requiring additional deterministic models and post-processing procedures.

Future research could investigate the potential for scaling model parameters with specific earthquake and site characteristics. In addition, the model's variability could be explored through comparisons to the standard deviations of empirical ground motion models. However, this requires either a complete set of observed records or the development of a scaling law for model parameters to effectively validate the model's variability.

Authors' note

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

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 partly financed by FCT/MCTES through national funds (PIDDAC) under the R&D Unit Institute for Sustainability and Innovation in Structural Engineering (ISISE), under reference UIDB/04029/2020 (doi.org/10.54499/UIDB/04029/2020), and under the Associate Laboratory Advanced Production and Intelligent Systems (ARISE) under reference LA/P/0112/2020. This study has been partly funded by the STAND4HERITAGE project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 833123), as an Advanced Grant. This work is partly financed by national funds through FCT (Foundation for Science and Technology), under grant agreement UI/BD/153379/2022 attributed to the first author.

ORCID iDs

SM Sajad Hussaini  <https://orcid.org/0000-0001-9841-5498>

Shaghayegh Karimzadeh  <https://orcid.org/0000-0003-3753-1676>

Sanaz Rezaeian  <https://orcid.org/0000-0001-7589-7893>

Data and resources

The recorded ground motions and their associated information were obtained from the NGA-West2 website accessible at <https://ngawest2.berkeley.edu/site> and ESM databases website accessible at <https://esm-db.eu>. A Python package implemented in object-oriented programming, along with the necessary instructions for the proposed improved stochastic ground motion simulation model, is available at: <https://github.com/Sajad-Hussaini/sgsim> with DOI: <https://doi.org/10.5281/zenodo.14565922>.

Supplemental material

Supplemental material for this article is available online.

References

- Alamilla J, Esteve L, García-Pérez J and Díaz-López O (2001) Evolutionary properties of stochastic models of earthquake accelerograms: Their dependence on magnitude and distance. *Journal of Seismology* 5(1): 1–21.
- Alimoradi A and Beck JL (2015) Machine-learning methods for earthquake ground motion analysis and simulation. *Journal of Engineering Mechanics* 141(4): 04014147.
- Amin M and Ang AH-S (1968) Nonstationary stochastic models of earthquake motions. *Journal of the Engineering Mechanics Division* 94(2): 559–584.
- Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T and Donahue JL (2014) NGA-West2 Database. *Earthquake Spectra* 30(3): 989–1005.
- Arslan Kelam A, Karimzadeh S, Yousefivavil K, Akgün H, Askan A, Erberik MA, Koçkar MK, Peçkan O and Ciftci H (2022) An evaluation of seismic hazard and potential damage in Gaziantep, Turkey using site specific models for sources, velocity structure and building stock. *Soil Dynamics and Earthquake Engineering* 154: 107129.
- Askan A, Altindal A, Aydin MF, Erberik MA, Koçkar MK, Tun M, Senol Balaban M, Uygucgil H, Kop A, Karimzadeh S, Mutlu S, Koska H, Pekkan E, Erkmen C, Celik A and Kılıç N (2025) Assessment of urban seismic resilience of a town in Eastern Türkiye: Turkoglu, Kahramanmaraş before and after 6 February 2023 M 7.8 Kahramanmaraş earthquake. *Earthquake Spectra* 41(1): 146–175.

- Askan A, Karimzadeh S and Bilal M (2017) Seismic intensity maps for the Eastern Part of the North Anatolian Fault Zone (Turkey) based on recorded and simulated ground-motion data. In: Çemen I and Yilmaz Y (eds), *Active Global Seismology: Neotectonics and Earthquake Potential of the Eastern Mediterranean Region, Geophysical Monograph Series*. New York: Wiley, pp. 273–287.
- Baker JW (2007) Quantitative classification of near-fault ground motions using wavelet analysis. *Bulletin of the Seismological Society of America* 97(5): 1486–1501.
- Beresnev IA and Atkinson GM (1997) Modeling finite-fault radiation from the ω spectrum. *Bulletin of the Seismological Society of America* 87(1): 67–84.
- Bernardo V, Karimzadeh S, Caicedo D, Hussaini SMS and Lourenço PB (2024) Fragility-based seismic assessment of traditional masonry buildings on Azores (Portugal) using simulated ground-motion records. *Earthquake Spectra* 40(4): 2836–2861.
- Boore DM (2003) Simulation of ground motion using the stochastic method. *Pure and Applied Geophysics* 160(3): 635–676.
- Boore DM (2009) Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bulletin of the Seismological Society of America* 99(6): 3202–3216.
- Broccardo M and Dabaghi M (2019) Preliminary validation of a spectral-based stochastic ground motion model with a non-parametric time-modulating function. In: *13th international conference on applications of statistics and probability in civil engineering (ICASPI3)*, Seoul, South Korea, 26–30 May.
- Burks LS and Baker JW (2014) Validation of ground-motion simulations through simple proxies for the response of engineered systems. *Bulletin of the Seismological Society of America* 104(4): 1930–1946.
- Chouet B, Aki K and Tsujiura M (1978) Regional variation of the scaling law of earthquake source spectra. *Bulletin of the Seismological Society of America* 68(1): 49–79.
- Dabaghi M and Der Kiureghian A (2017) Stochastic model for simulation of near-fault ground motions. *Earthquake Engineering & Structural Dynamics* 46(6): 963–984.
- Dabaghi M and Der Kiureghian A (2018) Simulation of orthogonal horizontal components of near-fault ground motion for specified earthquake source and site characteristics. *Earthquake Engineering & Structural Dynamics* 47(6): 1369–1393.
- Douglas J and Aochi H (2008) A survey of techniques for predicting earthquake ground motions for engineering purposes. *Surveys in Geophysics* 29(3): 187–220.
- Fayaz J, Rezaeian S and Zareian F (2021) Evaluation of simulated ground motions using probabilistic seismic demand analysis: CyberShake (ver. 15.12) simulations for Ordinary Standard Bridges. *Soil Dynamics and Earthquake Engineering* 141: 106533.
- 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(1): 17–26.
- 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.
- Hancock J, Watson-Lamprey J, Abrahamson NA, Bommer JJ, Markatis A, McCoyh E and Mendis R (2006) An improved method of matching response spectra of recorded earthquake ground motion using wavelets. *Journal of Earthquake Engineering* 10(Suppl. 001): 67–89.
- Hartzell S (2005) Calculation of broadband time histories of ground motion, Part II: Kinematic and dynamic modeling using theoretical green's functions and comparison with the 1994 Northridge earthquake. *Bulletin of the Seismological Society of America* 95(2): 614–645.
- Hartzell SH (1978) Earthquake aftershocks as Green's functions. *Geophysical Research Letters* 5(1): 1–4.
- Housner GW and Jennings PC (1964) Generation of artificial earthquakes. *Journal of the Engineering Mechanics Division* 90(1): 113–150.
- Huang D and Wang G (2017) Energy-compatible and spectrum-compatible (ECSC) ground motion simulation using wavelet packets. *Earthquake Engineering & Structural Dynamics* 46(11): 1855–1873.
- Karimzadeh S (2019) Seismological and engineering demand misfits for evaluating simulated ground motion records. *Applied Sciences* 9(21): 4497.

- Karimzadeh S, Funari MF, Szabó S, Hussaini SMS, Rezaeian S and Lourenço PB (2024a) Stochastic simulation of earthquake ground motions for the seismic assessment of monumental masonry structures: Source-based vs site-based approaches. *Earthquake Engineering & Structural Dynamics* 53(1): 303–330.
- Karimzadeh S, Kadasa K, Askanb A, Erberikb MA and Yakutb A (2020) Derivation of analytical fragility curves using SDOF models of masonry structures in Erzincan (Turkey). *Earthquakes and Structures* 18(2): 249–261.
- Karimzadeh S, Mohammadi A, Hussaini SMS, Caicedo D, Askan A and Lourenço PB (2023) ANN-based ground motion model for Turkey using stochastic simulation of earthquakes. *Geophysical Journal International* 236(1): 413–429.
- Karimzadeh S, Mohammadi A, Salahuddin U, Carvalho A and Lourenço PB (2024b) Backbone ground motion model through simulated records and XGBoost machine learning algorithm: An application for the Azores plateau (Portugal). *Earthquake Engineering & Structural Dynamics* 53(2): 668–693.
- Komatitsch D (2004) Simulations of ground motion in the Los Angeles basin based upon the spectral-element method. *Bulletin of the Seismological Society of America* 94(1): 187–206.
- Kristek J (2003) Seismic-wave propagation in viscoelastic media with material discontinuities: A 3D fourth-order staggered-grid finite-difference modeling. *Bulletin of the Seismological Society of America* 93(5): 2273–2280.
- Lin YK (1986) On random pulse train and its evolutionary spectral representation. *Probabilistic Engineering Mechanics* 1(4): 219–223.
- Luco N and Bazzurro P (2007) Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses? *Earthquake Engineering & Structural Dynamics* 36(13): 1813–1835.
- Lutes LD and Sarkani S (2004) Frequency, bandwidth, and amplitude. In: Lutes LD and Sarkani S (eds), *Random Vibrations*. Butterworth-Heinemann, pp. 261–306.
- Luzi L, Puglia R, Russo E, D'Amico M, Felicetta C, Pacor F, Lanzano G, Çeken U, Clinton J, Costa G, Duni L, Farzanegan E, Gueguen P, Ionescu C, Kalogeras I, Özener H, Pesaresi D, Sleeman R, Strollo A and Zare M (2016) The engineering strong-motion database: A platform to access pan-European accelerometric data. *Seismological Research Letters* 87(4): 987–997.
- Motazedian D and Atkinson GM (2005) Stochastic finite-fault modeling based on a dynamic corner frequency. *Bulletin of the Seismological Society of America* 95(3): 995–1010.
- Naeim F and Lew M (1995) On the use of design spectrum compatible time histories. *Earthquake Spectra* 11(1): 111–127.
- Olsen KB and Mayhew JE (2010) Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 mw 5.4 Chino Hills, California, earthquake. *Seismological Research Letters* 81(5): 715–723.
- Pousse G, Bonilla LF, Cotton F and Margerin L (2006) Nonstationary stochastic simulation of strong ground motion time histories including natural variability: Application to the K-Net Japanese database. *Bulletin of the Seismological Society of America* 96(6): 2103–2117.
- Rezaeian S and Der Kiureghian A (2008) A stochastic ground motion model with separable temporal and spectral nonstationarities. *Earthquake Engineering & Structural Dynamics* 37(13): 1565–1584.
- Rezaeian S and Der Kiureghian A (2010) Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthquake Engineering & Structural Dynamics* 39(10): 1155–1180.
- Rezaeian S and Der Kiureghian A (2012) Simulation of orthogonal horizontal ground motion components for specified earthquake and site characteristics. *Earthquake Engineering & Structural Dynamics* 41(2): 335–353.
- Rezaeian S and Sun X (2014) Stochastic ground motion simulation. In: Beer M, Kougiumtzooglou IA, Patelli E and Siu-Kui Au I (eds), *Encyclopedia of Earthquake Engineering*. Cham: Springer, pp. 1–15.

- Rezaeian S, Hartzell S, Sun X and Mendoza C (2017) Simulation of earthquake ground motions in the eastern United States using deterministic physics-based and site-based stochastic approaches. *Bulletin of the Seismological Society of America* 107(1): 149–168.
- 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.
- Rodolfo Saragoni G and Hart GC (1973) Simulation of artificial earthquakes. *Earthquake Engineering & Structural Dynamics* 2(3): 249–267.
- Safak E and Boore DM (1988) On low-frequency errors of uniformly modulated filtered white-noise models for ground motions. *Earthquake Engineering & Structural Dynamics* 16(3): 381–388.
- Sgobba S, Stafford PJ, Marano GC and Guaragnella C (2011) An evolutionary stochastic ground-motion model defined by a seismological scenario and local site conditions. *Soil Dynamics and Earthquake Engineering* 31(11): 1465–1479.
- Shahi SK and Baker JW (2014) An efficient algorithm to identify strong-velocity pulses in multicomponent ground motions. *Bulletin of the Seismological Society of America* 104(5): 2456–2466.
- Shannon CE (1949) Communication in the presence of noise. *Proceedings of the IRE* 37(1): 10–21.
- Shinozuka M and Sato Y (1967) Simulation of nonstationary random process. *Journal of the Engineering Mechanics Division* 93(1): 11–40.
- 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.
- Stafford PJ, Sgobba S and Marano GC (2009) An energy-based envelope function for the stochastic simulation of earthquake accelerograms. *Soil Dynamics and Earthquake Engineering* 29(7): 1123–1133.
- Su M, Dabaghi M and Broccardo M (2024) The importance of corner frequency in site-based stochastic ground motion models. *Earthquake Engineering & Structural Dynamics* 53(10): 3318–3329.
- Tamhidi A, Kuehn N, Ghahari SF, Rodgers AJ, Kohler MD, Taciroglu E and Bozorgnia Y (2022) Conditioned simulation of ground-motion time series at uninstrumented sites using Gaussian process regression. *Bulletin of the Seismological Society of America* 112(1): 331–347.
- Vlachos C, Papakonstantinou KG and Deodatis G (2016) A multi-modal analytical non-stationary spectral model for characterization and stochastic simulation of earthquake ground motions. *Soil Dynamics and Earthquake Engineering* 80: 177–191.
- Waezi Z and Balzadeh S (2022) Simulation of near-field pulse-like ground motions using a correlated bimodal fractional stochastic model. *Soil Dynamics and Earthquake Engineering* 161: 107434.
- Waezi Z, Rofooei FR and Hashemi MJ (2021) A multi-peak evolutionary model for stochastic simulation of ground motions based on time-domain features. *Journal of Earthquake Engineering* 25(2): 343–379.
- Wang F, Zhang Y, Yang B, Lin X and Ba Z (2024) Stochastic finite fault simulation of 2023 Mw 7.8 and Mw 7.5 Turkey earthquakes and its application to regional buildings damage estimation at Kahramanmaraş City. *Bulletin of Earthquake Engineering* 23(3): 867–892.
- Watson-Lamprey J and Abrahamson N (2006) Selection of ground motion time series and limits on scaling. *Soil Dynamics and Earthquake Engineering* 26(5): 477–482.
- Wu F, Xie J, An Z, Lyu C, Taymaz T, Irmak TS, Li X, Wen Z and Zhou B (2023) Pulse-like ground motion observed during the 6 February 2023 MW7.8 Pazarcık Earthquake (Kahramanmaraş, SE Türkiye). *Earthquake Science* 36(4): 328–339.
- Yamamoto Y and Baker JW (2013) Stochastic model for earthquake ground motion using wavelet packets. *Bulletin of the Seismological Society of America* 103(6): 3044–3056.
- Yeh C-H and Wen YK (1990) Modeling of nonstationary ground motion and analysis of inelastic structural response. *Structural Safety* 8(1–4): 281–298.
- Zeng Y, Anderson JG and Yu G (1994) A composite source model for computing realistic synthetic strong ground motions. *Geophysical Research Letters* 21(8): 725–728.