

# Ground motion simulation and validation of the 2008 Chino Hills earthquake in scattering media

W.H. Savran and K.B. Olsen

Department of Geological Sciences, San Diego State University, San Diego, CA, USA. E-mail: [wsavran@usc.edu](mailto:wsavran@usc.edu)

Accepted 2019 September 7. Received 2019 August 26; in original form 2019 January 8

## SUMMARY

We simulate 0–2.5 Hz deterministic wave propagation in 3-D velocity models for the 2008 Chino Hills, CA, earthquake using a finite-fault source model and frequency-dependent anelastic attenuation. Small-scale heterogeneities are modeled as 3-D random fields defined using an elliptically anisotropic von Kármán autocorrelation function with its parameters constrained using Los Angeles basin borehole data. We superimpose the heterogeneity models on a leading deterministic community velocity model (CVM) of southern California. We find that models of velocity and density perturbations can have significant effects on the wavefield at frequencies as low as 0.5 Hz, with ensemble median values of various ground motion metrics varying up to  $\pm 50$  per cent compared to those computed using the deterministic CVM only. In addition, we show that frequency-independent values of the shear-wave quality factor ( $Q_{s_0}$ ) parametrized as  $Q_{s_0} = 150V_s$  ( $V_s$  in  $\text{km s}^{-1}$ ) provides the best agreement with data when assuming the published moment magnitude ( $M_w$ ) of 5.4 ( $M_0 = 1.6 \times 10^{17}$  Nm) for the finite-fault source model. This model for  $Q_{s_0}$  trades off with  $Q_{s_0} = 100V_s$ , assuming  $M_w = 5.5$  ( $M_0 = 2.2 \times 10^{17}$  Nm), which represents an upper bound of the  $M_w$  estimates for this event. We find the addition of small-scale heterogeneities provides limited overall improvement to the misfit between simulations and data for the considered ground motion metrics, because the primary sources of misfit originate from the deterministic CVM and/or the finite-fault source description.

**Key words:** Computational seismology; Earthquake ground motions; Wave scattering and diffraction.

## INTRODUCTION

Gaining an understanding of the expected range of ground motions from future earthquakes remains a principal goal for seismologists and earthquake engineers. Due largely to continual developments in high-performance computing, deterministic earthquake simulations have shown potential for their ability to accurately characterize the broad-band seismic wavefield. For the task of predicting ground motions from future earthquakes, physics-based deterministic earthquake simulations have advantages over purely data-based approaches (e.g. Abrahamson *et al.* 2014; Boore *et al.* 2014; Campbell & Bozorgnia 2014; Chiou & Youngs 2014), because they do not inherently assume that future earthquakes behave like previous earthquakes. With that being said, deterministic simulations are not without their own limitations; namely, the difficulty in characterizing the problem in terms of physical parameters in addition to the computational expense associated with broad-band regional scale simulations. Despite these limitations, deterministic simulations have proven their usefulness for unraveling questions regarding seismic hazard and wave propagation phenomena (e.g. Li *et al.* 1994; Kawase 1996; Olsen 2000); Spudich & Olsen 2001; Aagaard

& Heaton 2004; Day *et al.* 2008; Pitarka *et al.* 2009; Aagaard *et al.* 2010b; Duan & Day 2010; Roten *et al.* 2014).

Deterministic wave propagation simulations require modellers to define the material properties of the earth (i.e. elastic moduli, viscoelastic and potentially plastic parameters) along with a description of the earthquake source. Generally, these parameters are poorly constrained for small scale-lengths (i.e.  $< \sim 100$  m, primarily affecting high-frequency wave propagation. In spite of this assumption, two recent studies (Taborda & Bielak 2013, 2014, hereafter TB13 and TB14, respectively), thoughtfully investigated the limitations of physics-based earthquake simulations with a specific focus on the 29 July 2008 Chino Hills earthquake in Southern California. In both studies, the authors simulate 4 Hz ground motions using Hercules, a parallel finite-element code, used to solve the elastodynamic equations governing seismic wave propagation (Bielak *et al.* 2011). Specifically, the authors validate their simulations against strong motion records using different state-of-the-art community velocity models (CVMs). The validation process quantitatively evaluates whether a physical model accurately represents the system it intends to study. Based on their analysis, the authors find good comparisons below 1 Hz, but note growing discrepancies

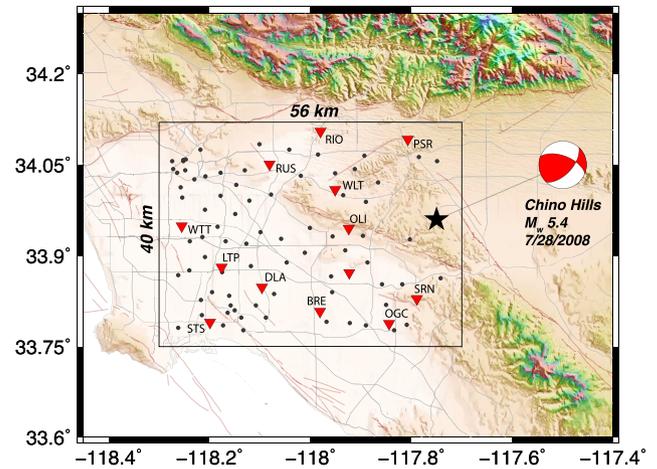
between the wave propagation model and ground motion records as frequencies increase. A contributing reason for this result may be the lack of small-scale complexity in the modeling.

Current state-of-the-art CVMs (e.g. Magistrale *et al.* 2000; Kohler *et al.* 2003; Süß & Shaw 2003; Lee *et al.* 2014) insufficiently resolve the structural complexity known to exist in the earth due predominately to the low cut-off frequency ( $f_{max} \approx 0.5$  Hz) of the data used for 3-D tomographic inversions or the limited data available to describe geologic structures at depth. Due to the cost and complexity of acquiring meter-scale seismic velocity models, a favourable approach to include such media heterogeneity has been to superimpose statistical descriptions of the small-scale velocity and density fluctuations onto a deterministic velocity model (e.g. Frankel & Clayton 1986; Hartzell *et al.* 2010; Imperatori & Mai 2012; Bydlon & Dunham 2015). Assuming these statistical distributions accurately represent the variability present in the earth, we can expect, at least in a statistical sense, the scattered wave field will be comparable with those originating from natural earthquakes. However, a caveat to this approach is that it requires ensembles of ground motions to understand the expected ground motions caused by scattering media (e.g. Imperatori & Mai 2012).

Largely, this manuscript extends the work done by TB13 and TB14 to include effects of small-scale velocity and density fluctuations. We study the effects of small-scale heterogeneity in seismic velocities and densities and their trade-offs with various attenuation models in the context of seismic hazard analysis. Toward this goal, we compare simulated seismograms against strong-motion data recorded for the 2008  $M_w$  5.4 Chino Hills earthquake. Building on TB13 and TB14 we investigate two other factors that influence simulated ground motions; namely, anelastic attenuation and statistical distributions of small-scale velocity and density fluctuations. A recent study by Withers *et al.* (2015) demonstrated the necessity of including frequency-dependent attenuation for simulations of the 2008 Chino Hills earthquake, so we focus our efforts on refining the empirical  $Q$  relationships used to define the attenuation model. Several studies have investigated the effects of random media on scattering attenuation and coda waves (e.g. Wu & Aki 1985; Frankel & Clayton 1986; Korn 1993; Imperatori & Mai 2012; and more recently Bydlon & Dunham 2015; Wang & Shearer 2017). Here, we address the effects of scattering attenuation with a focus on individual ground motion metrics relevant to seismic hazard. With the progress made by physics-based probabilistic seismic hazard techniques, such as the Southern California Earthquake Center (SCEC) CyberShake (Graves *et al.* 2011) or the SCEC Broadband Platform (Maechling *et al.* 2015), understanding the effects of scattering can help inform these large-scale efforts.

### The 29 July 2008 Chino Hills earthquake

At 11:42 am PDT on 29 July 2008, the greater Los Angeles area experienced the largest magnitude earthquake since the 1994  $M_w$  6.7 Northridge event. The preliminary analysis of the event (Hauksson *et al.* 2008) indicated oblique faulting (thrust and strike-slip) shown by the focal mechanism in Fig. 1. Based on their analysis, the epicentre was found to be located between two known structures—the Whitter and Chino Hills faults, and was the fifth earthquake sequence to occur in this region since 1987. Despite its proximity to the Los Angeles basin there were no reported fatalities or significant damage from this earthquake (Hauksson *et al.* 2008). With much fortune to seismologists, this event was also recorded by over 450 stations in the greater Los Angeles area providing an excellent



**Figure 1.** Station locations of the 118 strong-motion recordings used during our validation (black dots). The star represents the epicentre of the 2008 Chino Hills event, and the black box denotes our simulation domain. Stations shown in red triangles are used for qualitative comparisons between the recordings and data. We avoid stations located within 2.5 km from the model boundaries with potential bias by reflected phases damped by the absorbing boundary conditions.

opportunity to validate deterministic ground motion simulations and CVMs.

A subsequent inversion study by Shao *et al.* (2012) finds that a northeast dipping plane (strike =  $289^\circ$  and dip =  $62^\circ$ ) provides the best fits with the recorded data. The inversion results indicate a hypocentral depth of 14.6 km near the base of the seismogenic zone. In addition, they find that approximately 50 per cent of the moment release occurs from a  $1.8 \text{ km}^2$  asperity with a maximum stress drop of 80 MPa and a maximum slip of 1.8 m. The average slip is approximately 0.5 m with an average stress drop between 19 and 38 MPa, much larger than the average stress drop ( $\sim 3$  MPa) observed for southern California earthquakes (Shearer *et al.* 2006).

### Numerical method for simulating ground motions

To simulate 0–2.5 Hz ground motions from the  $M_w$  5.4 Chino Hills event we solve the Navier–Cauchy equations for elastodynamics using the 4th order space, 2nd order time staggered-grid finite difference code AWP-ODC-CPU (Cui *et al.* 2010) including frequency-dependent attenuation (Withers *et al.* 2015). We mitigate reflected energy propagating into the simulation domain using damping boundary conditions described by Cerjan *et al.* (1985) that, using a sufficient number of damping elements (30–50 nodes), prevents any significant spurious energy from contaminating our numerical solution. For an in-depth explanation of the numerical method itself and its implementation for high-performance computers, we refer the reader to Cui *et al.* (2010).

### Velocity model with small-scale velocity and density fluctuations

We discretize a 56 km x 40 km x 24 km region with a regular spacing of 16 m encompassing the majority of Los Angeles basin. Due to the number of simulations required for the ensembles, we use a smaller region than TB13 and TB14 (see Fig. 1).

From the elastic theory of wave propagation, the parameters  $\lambda$ ,  $\mu$  and  $\rho$  are necessary and sufficient to completely describe an isotropic elastic medium. We obtain a model of these elastic parameters from the SCEC CVM-S4.26 (Lee *et al.* 2011) which is an improvement to CVM-S4 (Magistrale *et al.* 2000; Kohler *et al.* 2003) including results from a 3-D tomographic study in Southern California. The inversion was performed using a resolution of 500 m and the results are tri-linearly interpolated to smaller resolutions, if necessary. CVM-S4.26 utilizes a software interface that returns  $V_p$ ,  $V_s$  and  $\rho$  for a queried latitude, longitude, and depth. The CVM-S4.26 model is developed and supported by SCEC and can be acquired through the SCEC UCVM package (Small *et al.* 2017). We used UCVM version 14.3.0 to generate the deterministic velocity model used in this study.

TB14 investigated the effects of different velocity models including models for the shallow sediments (Shaw *et al.* 2015), and they found that ground motions display significant sensitivity to the velocity model considered. In this article, we focus on the effects of small-scale velocity and density fluctuations as opposed to comparing the effectiveness of an individual velocity model. As such, our analysis only includes simulations from a single deterministic velocity model, CVM-S4.26. However, we suspect that our results showing the effects of velocity and density fluctuations are generalizable to other leading 3-D velocity models for the greater Los Angeles region.

We include small-scale velocity and density fluctuations, known to exist in the crust, by superimposing a statistical model of the fluctuations onto CVM-S4.26. We define velocity and density fluctuations with a normally distributed random variable following a von Kármán autocorrelation function

$$\Phi_{v,a}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{a}\right)^\nu K_\nu\left(\frac{r}{a}\right), \quad (1)$$

which has Fourier transform

$$P(k) = \frac{\sigma^2 (2\sqrt{\pi}a)^E \Gamma(\nu + E/2)}{\Gamma(\nu)(1 + k^2a^2)^{\nu+E/2}} \quad (2)$$

(Klimeš 2002). The parameters of the von Kármán autocorrelation function are: correlation length,  $a$ , Hurst exponent,  $\nu$ , and standard deviation,  $\sigma$ . Also,  $\Gamma(\nu)$  is the gamma function,  $K_\nu$  is the modified Bessel function of the second kind with order  $\nu$ , and  $E$  represents the Euclidean dimension. We generate the velocity and density fluctuations using a spectral approach described in detail by Pardo-Igúzquiza & Chica-Olmo (1993) and Klimeš (2002). Using this approach, we simulate the fluctuation profile in the wavenumber domain using the spectral amplitudes defined in eq. (2) while leaving the phases random. This approach unconditionally simulates a zero-mean stochastic process with the autocorrelation function defined by eq. (1).

We adopt parameters representative of the Los Angeles area from a study by Savran & Olsen (2016) that investigated a suite of 39 borehole sonic logs recorded in Los Angeles basin. The authors found that  $\nu = 0.05$ ,  $a_z = 50 - 150$  m, and  $\sigma = 5$  per cent best characterize the fluctuation profiles observed in the borehole logs. We also incorporate elliptical anisotropy in the fluctuations with a ratio of horizontal-vertical correlation lengths ( $a_x/a_z$ ) = 5. These parameters are comparable with an independent study of the same borehole data (Andreas Plesch 2012, personal communication), as well as spectral analysis of a 3-D velocity model computed from ambient noise data (Nakata *et al.* 2015) generated from a high-resolution seismic dataset deployed in the Long Beach area (Nakata & Beroza 2015). The study area of Nakata & Beroza (2015) spans

a different region of Los Angeles basin than the data locations of Savran & Olsen (2016), but provides similar parameters of the von Kármán autocorrelation function, suggesting that these parameters constitute a representative model for Los Angeles basin.

Fig. 2(a) shows a horizontal slice of CVM-S4.26 superimposed with small-scale velocity and density fluctuations at a depth of 160 m (10 grid points). We show a cross-sectional view of the velocity model (Fig. 2b) illustrating the anisotropy of the media fluctuations, and a vertical profile extracted along this transect (Fig. 2c) in the deeper part of Los Angeles basin. We leave a small region surrounding the finite-fault source unperturbed from small-scale heterogeneities (Fig. 2b) to ensure the moment distribution of the source is consistent with the finite-fault inversion. A further illustration of the composite models for different parameters of the statistical distributions used in this study can be found in Fig. S1 in the electronic supplement.

### Finite fault source model

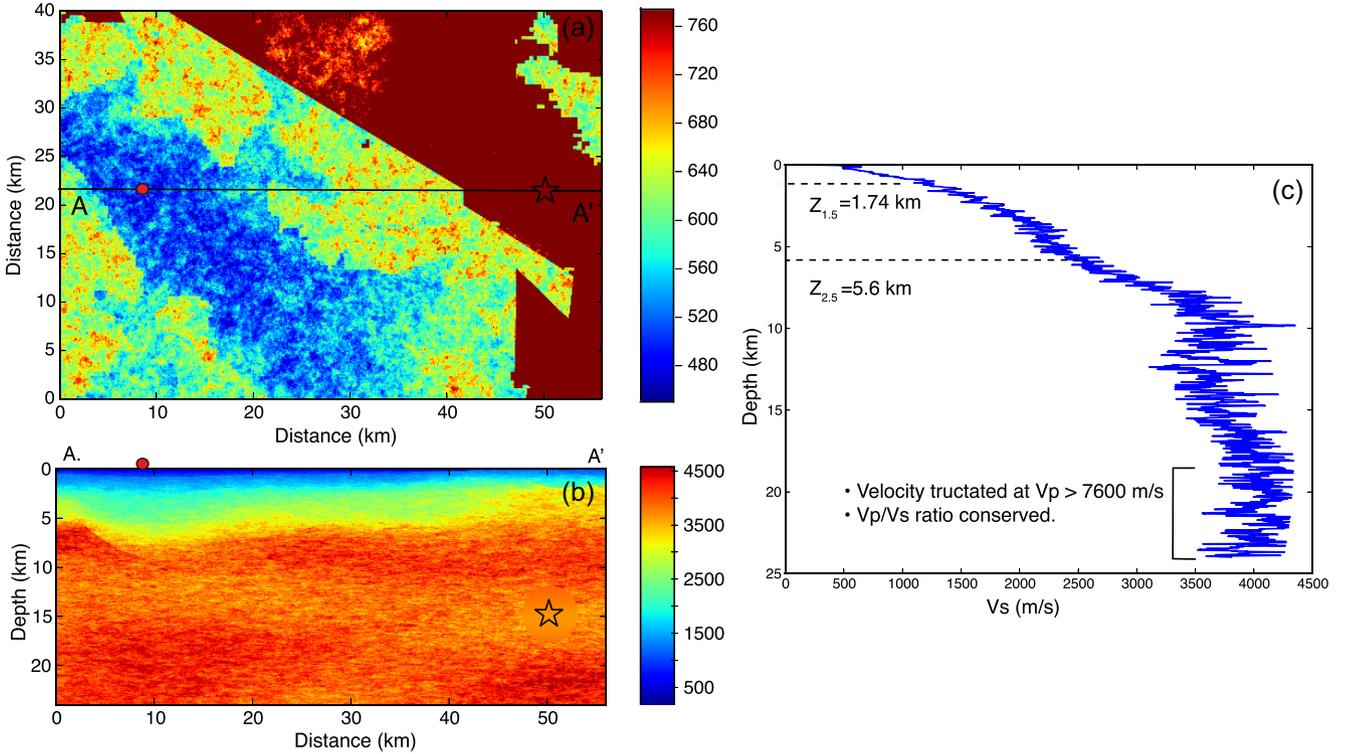
We use a finite-fault source inverted using 26 strong motion stations located at distances less than 50 km from the Chino Hills fault from Shao *et al.* (2012). Before inverting for the finite-fault model, their data were band-pass filtered from 0.16 to 2.5 Hz, and  $P$  and SH waveforms were chosen based on having desirable signal-to-noise ratios. Their inversion used a simulated annealing approach that estimates the slip, rake, rupture-time and parameters of the asymmetric cosine slip-rate function (Ji *et al.* 2003). Fig. S2 in the electronic supplement shows the inverted slip-rate functions and Fourier amplitude spectra for all 140 subfaults.

The top of the fault starts at a depth of 13.2 km and has dimensions 5.6 km along strike and 4.0 km along dip. Each of the 140 subfaults has an area of 0.16 km<sup>2</sup> (400 m x 400 m), with the same strike = 289° and dip = 62° and a mean rake = 133.5°. The rake varies between 115° and 160° across the fault plane. We use the inversion results corresponding to ‘Model 1’ from Shao *et al.* (2012). While TB13 and TB14 used  $f_{\max} = 4.0$  Hz, we choose to consider the band-limited region up to  $f_{\max} = 2.5$  Hz for two main reasons: (1) the data used in the inversion were bandpass filtered between 0.15 and 2.5 Hz and (2) TB13 and TB14 found significant misfits between their simulations and recorded data in addition to uncertainty about the velocity model that we in part attribute to the lack of data resolution above  $f_{\max} = 2.5$  Hz.

An accurate estimate of the size of the earthquake is critical for assessing the effects of the crustal model, including anelastic attenuation and scattering. Shao *et al.* (2012) found a moment of  $1.60 \times 10^{17}$  Nm, corresponding to  $M_w$  5.4, which is used as the reference in our study. However, other source inversion studies have found higher moments, such as USGS moment tensor (TMTS) and Centroid moment tensor ( $M_{wc}$ ) estimates of  $1.83 \times 10^{17}$  and  $1.96 \times 10^{17}$  Nm, respectively, corresponding to  $M_w = 5.5$ . In the ‘Discussion’ section we explore the effects of a possibly larger moment on the fit between data and simulations.

### Frequency-dependent attenuation

Anelastic attenuation is necessary to accurately model the amplitudes of seismic waves that propagate distances much larger than their dominant wavelength, as is the case with regional-scale deterministic ground motion simulations. Typically, numerical studies



**Figure 2.** Velocity fluctuations from the von Kármán correlation function with parameters  $\nu = 0.05$ ,  $a = 150$  m,  $\sigma = 5$  per cent,  $H/V = 5$  superimposed on SCEC CVM-S4.26. (a) Depth slice at 160 m depth, (b) cross-section taken at A-A' showing the deeper area of Los Angeles basin, and (c) profile extracted at the red circle in (a, b). Velocities are listed in ( $\text{m s}^{-1}$ ). The mean of the velocity fluctuation profile represents the CVM background  $V_s$  values. Note in (b) that the perturbations are excluded from the source region following Frankel & Clayton (1986) to preserve the moment released during the earthquake. The star depicts the hypocentre for the Chino Hills event.

(e.g. Olsen *et al.* 2003; Graves & Pitarka 2010; TB13; TB14) incorporate a frequency-independent attenuation model for the quality factor,  $Q$ , meaning that intrinsic energy losses per cycle are equivalent across all frequencies. This model is sufficient for wave propagation with  $f_{\max} < 1.0$  Hz, but tends to disagree with seismic observations at higher frequencies (e.g. Aki 1980; Raoof *et al.* 1999; Erickson *et al.* 2004; Phillips *et al.* 2013; Wang & Shearer 2017), which show that a frequency-dependent attenuation model produces better fits to the recorded data. Typically, empirical relationships for attenuation are based on the local shear-wave velocity due to our understanding that regions with hard rock tends to attenuate the wavefield less than areas with broken and fractured rock or soft-sediments (Hauksson & Shearer 2006).

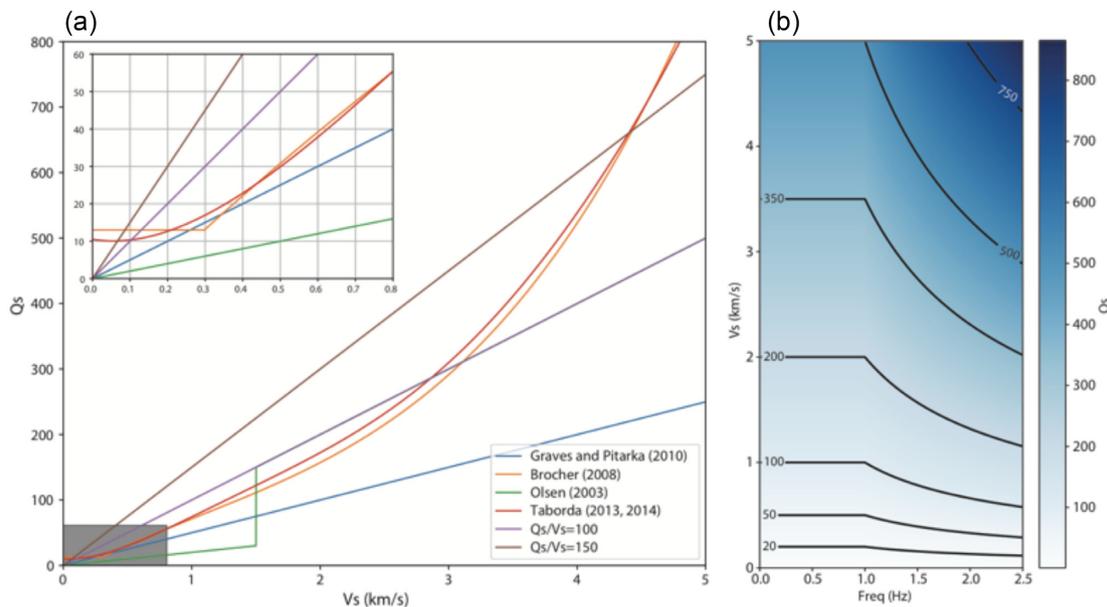
Withers *et al.* (2015) developed a memory-efficient technique to incorporate frequency-dependent attenuation using a coarse-grained memory variable approach following Day (1998) and Day & Bradley (2001). In the frequency-dependent case, the weights of the viscoelastic relaxation times are modified so that the target  $Q$  spectrum follows a power-law (eq. 3) above a reference frequency,  $f_0$ , with power-law exponent  $\gamma$ :

$$Q(f) = \begin{cases} Q_0 \left(\frac{f}{f_0}\right)^\gamma, & f \geq f_0 \\ Q_0, & f < f_0 \end{cases}, \quad (3)$$

In this study, we focus on empirical models for the reference  $Q_0$ , due largely to our simulation  $f_{\max} = 2.5$  Hz, as  $Q(f)$  has limited effects at 2.5 Hz. We use  $f_0 = 1.0$  Hz and  $\gamma = 0.6$ , with separate

values  $Q_p$  and  $Q_s$  for the  $P$ - and  $S$ -wave quality factors, respectively. Additionally, most models incorporate  $Q_s$  relationships of the form  $Q_s = a_i (V_s)^{b_i}$ , where  $a_i$  and  $b_i$  are scalar coefficients, and  $V_s$  ( $\text{m s}^{-1}$ ) is the shear-wave velocity. In this study, we choose  $a_i = (50, 100, 150)$  and  $b_i = 1$ . Following Brocher (2008), the simulations performed by TB13 and TB14 incorporate higher order terms in their empirical relationship. In this study, we attempt to find the best linear model for the empirical relationship defining  $Q_s$ . We assume that  $Q_p = 2Q_s$ ; while this relation is slightly different than the  $Q_p$  attenuation model chosen by TB13 and TB14, Olsen *et al.* (2003) finds that ground motions are relatively insensitive to  $Q_p$ .

Fig. 3(a) compares several  $Q$  models proposed in the literature (Olsen *et al.* 2003; Brocher 2008; Graves & Pitarka 2010; TB13, TB14) and two other models used in our simulations, namely,  $Q_s = 100V_s$  and  $Q_s = 150V_s$ . We note that the Graves & Pitarka (2010) model is equivalent to the  $Q_s = 50V_s$  model, also investigated here. The inset of Fig. 3(a) shows the attenuation models for  $V_s < 0.8$  km  $\text{s}^{-1}$ . Fig. 3(b) shows the effects of the frequency-dependence of  $Q_s$  for the model where  $Q_0 = 100V_s$  and  $\gamma = 0.6$ . While the different  $Q$  models vary significantly, an important point to mention is that ground motions are influenced by the value of  $Q$  itself and not specifically by the coefficients of the empirical relationship used to define the intrinsic attenuation model. In other words, we suspect that the choice of  $\min(V_s)$  can largely explain the different attenuation models proposed by various simulation studies for the Southern California area, as simulations with larger  $\min(V_s)$  (e.g. Olsen *et al.* 2003; Graves & Pitarka 2010) tend to choose



**Figure 3.** (a) Shear-wave quality factor ( $Q_s$ ) plotted against  $V_s$  ( $\text{km s}^{-1}$ ) for several attenuation models widely used in the literature (Olsen *et al.* 2003; Brocher 2008; Graves & Pitarka 2010; TB13 and TB14). The inset figure in the upper left shows the different attenuation models for  $V_s < 0.8 \text{ km s}^{-1}$ . These plots are valid for the frequency-independent portion ( $Q_0$ ) of our attenuation law ( $< 1 \text{ Hz}$ ), after which the  $Q_s$  values are modified by the equation  $Q(f) = Q_0 f^\gamma$ , where  $\gamma = 0.6$ . This plot is based on TB13. (b) Plot showing frequency-dependent attenuation for  $Q(f) = Q_0 f^\gamma$ , where  $Q_0 = 100 V_s$  and  $\gamma = 0.6$ . Contour lines show the effects of  $Q(f)$ , with the most prominent effects happening at large  $V_s$ . The contour labels and the colours show the value of  $Q_s$ .

lower values of  $a_i$ , but ultimately result in approximately the same  $\min(Q_s) \approx 20$ .

## GROUND MOTION SIMULATIONS

We simulate 100 seconds of wave propagation with  $f_{\max} = 2.5 \text{ Hz}$ , to ensure that the wavefield propagates throughout the entire simulation with the chosen  $\min(V_s) = 200 \text{ m s}^{-1}$ . We discretize the simulation domain with  $dx = 16 \text{ m}$  to resolve the smallest wavelength with 5 gridpoints using the 4th order staggered-grid finite-difference solution (Olsen *et al.* 2003). The resulting mesh has 13.125 billion elements, and each simulation conducted for this study requires approximately 6.5 hours of wall-time using 13 125 processors on NCSA Blue Waters. Table 1 shows the simulation parameters used for the models presented in this study.

### Validating ground motion simulations

Validation is needed to estimate the efficacy in predicting ground motions for future events. The validation of ground motion simulations requires comparing synthesized ground motions against ground motion records to assess the accuracy of the model with respect to the entire physical system under study. Until TB13, TB14 and the SCEC Broadband Platform validation (e.g. Goulet *et al.* 2015), most validation exercises were performed using qualitative comparisons between seismograms at select stations or using simple quantitative comparisons (e.g. Olsen *et al.* 2003; Aagaard *et al.* 2010a; Graves & Pitarka 2010). However, recent developments in model-wide validation techniques using multiple goodness-of-fit metrics (Anderson 2004; Olsen & Mayhew 2010) or strict time-frequency misfit criteria (Kristekova *et al.* 2009) have encouraged

**Table 1.** Simulation parameters used for the deterministic ground motion simulations of the 2008 Chino Hills earthquake.

Domain	
Length	56 km (3500 nodes)
Width	40 km (2500 nodes)
Depth	24 km (1500 nodes)
Southwest corner	33.7500, -118.3000
Spatial resolution	
Maximum frequency	2.5 Hz
Minimum $V_s$	200 $\text{m s}^{-1}$
Points per minimum wavelength	5
Grid discretization	16 m
Number of cells	13.125 billion
Wall-clock time	6.5 hr
Number of processors	13 125
Temporal resolution	
Time discretization	0.001 s
Simulation time	100.0 s
Number of timesteps	100 000

seismologists to focus on more rigorous validation and verification of their models.

In the case of the Chino Hills simulations presented here the physical system encompasses (1) the CVM-S4.26 deterministic velocity model, (2) the stochastic model of velocity and density fluctuations, (3) the empirical model for anelastic attenuation and (4) the finite-fault model used to describe the Chino Hills event. Due to the complexity of nature a perfect validation of the modelled physical system is likely impossible (Oreskes *et al.* 1994), but informative to highlight strengths and weaknesses of different aspects of the modelled physical system. TB13 and TB14 investigated many of these

aspects with respect to the Chino Hills earthquake under study here. As such, we focus our efforts on validating statistical models of velocity and density fluctuations and empirical relationships for intrinsic attenuation rather than the finite-fault source or velocity models.

### Goodness of fit criteria

To perform our quantitative validation, we use a simplified version of the goodness-of-fit (GOF) criteria proposed by Olsen & Mayhew (2010). In this section, we summarize the GOF methodology and explain the motivation for the GOF metrics used in this study. Olsen and Mayhew incorporate ten GOF metrics into their method: peak velocity (PGV); peak acceleration (PGA); peak displacement (PGD); response spectral acceleration averaged between 0.1 and 10 s (RS); spectral acceleration estimated by NGA relationships for 16 periods (SA16); smoothed Fourier amplitude spectrum (FAS); energy duration (DUR); cumulative energy (ENER); cross-correlation (XCOR) and the inelastic to elastic deformation ratio (IE). In general, the method computes a GOF score between 0 and 100 for each station, where 100 indicates a perfect match, and subsequently aggregates them for a global GOF score.

The Olsen and Mayhew method was designed specifically for use with broadband (0–10 + Hz simulations) to assess multiple aspects of the underlying physical model. Here, we select four of their metrics that we believe provide the clearest interpretation of our results, namely, PGV, PGA, ENER, DUR and one additional metric considered by Anderson (2004), TB13 and TB14, the Arias intensity (AI). AI is defined as  $\frac{\pi}{2g} \int_0^T a(t)^2 dt$ , where  $a(t)$  is the acceleration time record,  $T$  is the duration of the signal, and  $g$  is the acceleration due to gravity. This metric describes the cumulative energy-per-weight from a uniformly distributed set of single degree-of-freedom oscillators on  $[0, \infty)$  (Arias 1970). Also, we define DUR for each component as  $DUR = \int_0^t v(\tau)^2 d\tau$ , where  $t$  is the time when 75 per cent of the cumulated energy has arrived at the station. We assume the signal starts when 5 per cent of the total energy has arrived at the station. In effect, DUR gives the time required for a station to obtain 5–75 per cent of the recorded energy from an event.

Following Olsen and Mayhew, we compute the GOF score for each metric ( $G_{met}$ ) using a mapping of the normalized residual to the complementary error function ( $erfc$ ) using eq. (4)

$$G_{met} = 100 \operatorname{erfc} \left( \frac{2|x-y|}{x+y} \right), \quad (4)$$

where  $x$  and  $y$  are two positive scalar metrics, and the result of the mapping is scaled by 100 resulting in GOF scores that lie between 0 and 100. Next, we compute a single GOF score for the time-series ( $G_{ts}$ ) by combining all metrics using an equal-weighted average (eq. 5)

$$G_{ts} = \frac{1}{N} \sum_N G_{met}. \quad (5)$$

We then compute a GOF score for a particular station ( $G_{sta}$ ) by averaging  $G_{ts}$  assuming equal weights for each three ground motion components using  $G_{sta} = \frac{1}{3} \sum_3 G_{ts}$ . Finally, we assign a GOF score to the entire simulation ( $G_{sim}$ ) by averaging the GOF scores across all stations. We favour the  $erfc$  implementation over the exponential

function used by Anderson (2004) due to its ability to provide better discrimination at high GOF scores. In practice, we find that the GOF approach presented here is similar to that used by previous studies.

### Broad-band data processing

From the 450 stations recording the event, we include 118 stations that lie within our domain for validating our simulations. In addition to excluding stations outside of our computational domain, we also remove 24 stations due to having incomplete records, visibly erroneous data, or being located inside or within 2.5 km of the sponge zones. Fig. 1 shows the locations of the 109 stations included in our validation procedure, as well as the locations of 14 stations used to evaluate simulations against ground motion recordings.

We obtained processed strong-motion recordings from R. Taborda (2013, personal communication), thus our records are processed following the approach of TB13 and TB14, summarized below. First, TB13 performed baseline and gain corrections and integrated the acceleration time-series to velocity and displacement with a high-pass cut-off frequency of 0.1 Hz. All recorded seismograms not aligned with the east–west and north–south directions were rotated into this coordinate system. We note that our computational domain aligns with east–west and north–south (+x indicates east and +y indicates north). Next, the recordings were synchronized temporally using the trigger time reported in the header files along with the rupture time of the event based on the USGS records. We apply a systematic time-shift of 0.9 s to all records corresponding to the average rise time in our source description. Finally, in contrast with TB13, we bandpass filter the records to 0.15–2.5 Hz to allow for comparisons with our synthetics. As with TB13, all filters used in the data processing are Chebyshev Type I filters.

## RESULTS OF GROUND MOTION SIMULATIONS

First, we perform a sensitivity test of the von Kármán parameters ( $v$ ,  $a$ ,  $\sigma$ ,  $H/V$ ) to understand their effects on simulated ground motions and the variability introduced by including different statistical models of velocity and density fluctuations. Next, we make qualitative comparisons at select stations to provide some insight into the behaviour of small-scale heterogeneities for metrics relevant to seismic hazard, such as the Fourier velocity spectra, response spectra and cumulated energy. We also show the effects of small-scale heterogeneities on individual metrics aggregated across the simulation domain. In addition to analyzing the effects of the small-scale heterogeneities, we investigate the distance dependence of the five selected metrics (PGA, PGV, DUR, ENER, AI) to showcase the behaviour of different attenuation models.

### Sensitivity testing of small-scale heterogeneities

Previous studies show correlation length estimates for the distributions of small-scale heterogeneities, ranging from meters to several kilometers. These differences can be explained by inconsistencies during the processing stage of data analysis (e.g. Dolan & Bean 1997) or different data types (i.e. seismic, geological, borehole, simulation studies). Hurst exponent estimates in the literature typically range between 0.0 and 0.3 (e.g. Frankel & Clayton 1986;

**Table 2.** Simulations considered for the sensitivity testing for different parameterizations of velocity and density fluctuations. We consider end-member cases for the Hurst exponent,  $\nu$ , vertical correlation length,  $a_z$ , ratio of horizontal to vertical correlation lengths, H/V, standard deviation expressed as the coefficient of variation,  $\sigma$ , and the spatial locations of fluctuations determined by the seed number. The values inside the parentheses next to the model names indicate the velocity models corresponding to the panels in Fig. S1.

Model	$\nu$	$a_z$ (m)	H/V	$\sigma$ (%)	Seed
C1 (b,c)	0.0, 0.3	150	5	5	1
C2 (a,d)	0.05	150, 5000	5	5	1
C3 (a,f)	0.05	150	2, 5	5	1
C4 (a,e)	0.05	150	5	5, 10	1
C5 (a,g)	0.05	150	5	5	1–5

Levander 1992). We compare horizontal over vertical ratios of correlation length of two and five, to estimate the effect of realistic anisotropy in the heterogeneities, and use standard deviations of  $\sigma = 5$  and  $\sigma = 10$  per cent as constrained by the analysis shown in Nakata & Beroza 2015; Savran & Olsen 2016). Finally, we vary the spatial locations of the media fluctuations while keeping the von Kármán parameters identical. Fig. S1 shows a horizontal slice of  $V_s$  at depth  $z = 160$  m for each velocity model considered in the sensitivity study. Table 2 shows the different parameters for the models considered.

Qualitative comparisons are made at four stations, LTP, RUS, STS, and SRN (Fig. 4), selected at a variety of distances and azimuths, for model ensemble C5 (see Table 2). The effects of varying the locations of the velocity and density fluctuations (via the seed number of the random numbers) cause the largest variability between synthetic ground motions for all the parameters we investigated. Such ensembles of simulations need to be considered to capture the expected systematic effects of scattering, since the exact locations of the velocity and density fluctuations are not known.

The effects on the time-series due to the correlation length, Hurst number, horizontal–vertical anisotropy, and strength of the small-scale heterogeneities are shown in Figs S3–S6, respectively, of the electronic supplement. We find a relatively large variation between cumulated energy from variation in the correlation length (150–5000 m, C2, see Fig. S3). For example, synthetics with  $a = 5000$  m show approximately twice the cumulative energy as compared to those with  $a = 150$  m for station STS. This effect can be explained by the relatively larger size of fluctuations interacting with a larger portion of the wavefield. The nondimensional parameter  $ak$  where  $a$  is the correlation length and  $k$  is the wave number largely controls the behaviour of scattering. In the case where  $ak \gg 1$  (strong scattering regime, Przybilla *et al.* 2009), most of the energy is scattered forward, with large peak amplitudes generated from geometric focusing effects in early arrivals (e.g. at STS). In contrast, when  $ak \ll 1$ , the energy is predominantly scattered in the backward direction, typically distributing energy from the initial arrivals to the coda. With  $a = 5000$  m the modeled wavefield results in  $ak \gg 1$  for most of the domain. We find that stations exhibit only minor differences between cumulated energy (generally  $< 10$  per cent) for models with Hurst exponent of  $\nu = 0.0$  and  $\nu = 0.3$  (C1, see Fig. S4). Fig. S5 tests differences in ground motions between horizontal-vertical anisotropy ratios (Model C3). In these tests, the vertical correlations lengths are fixed at 160 m. In general, the results are similar to those for the correlation length described above. We find no direct correlation between the anisotropy and the cumulated energy. Increasing the strength of the fluctuations from 5 to 10 per cent can cause up to 50 per cent larger ground motions (e.g.

at LTP, see Fig. S6, Model C4). The increased peak motions occur at both early and late arrivals.

### Qualitative validation at select stations

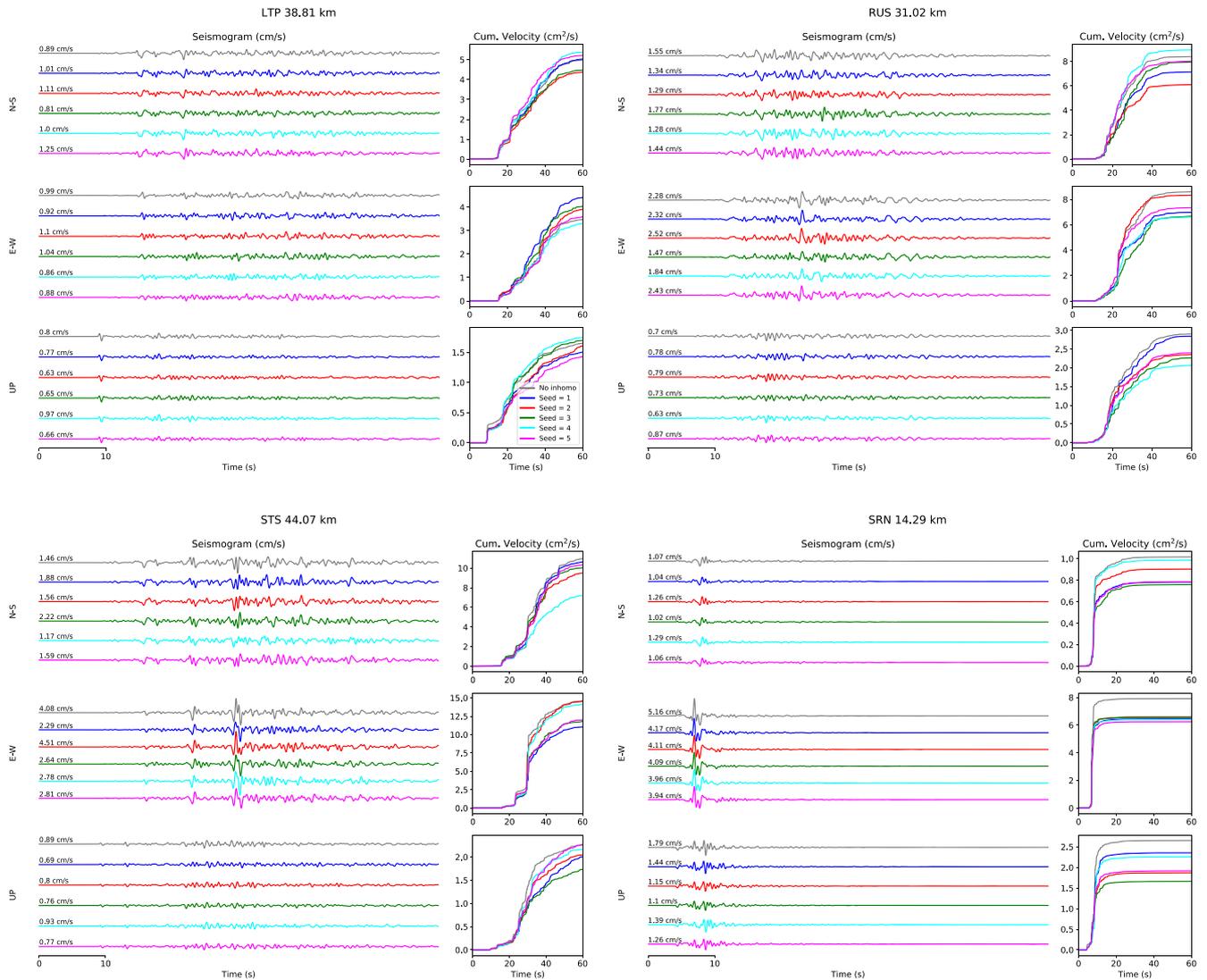
We begin our validation by showing plots of ground motion metrics at stations inside our simulation domain; namely, the waveform envelope (one-sided Hilbert transform), the cumulated energy, Fourier velocity spectrum, and acceleration response spectra (Fig. 5, and see Fig. S7 for additional support). Moreover, we also show  $G_{sta}$  for models with and without media fluctuations. These stations are relatively evenly distributed across our simulation domain, belong to the same seismic network, represent a fair distribution of the 109 stations included in our analysis and include some of the best and worst performing stations.

The selected progress-in-time metric, the cumulative energy (ENER), provides the most agreeable fits with the observed data (e.g. at OLI and PSR). At some stations, for example DLA, LTP, FUL, all sets of synthetics are underpredicting the amplitude of the  $S$ -wave arrivals causing large discrepancies in the resulting metrics. We attribute this mainly to the velocity model, as the distributions of velocity and density fluctuations appear to only slightly alter the peak arrivals.

A noteworthy feature of the comparisons is that the ensemble means of models including small-scale heterogeneities are not equal to models that do not consider velocity and density fluctuations. This is especially apparent in the plots of acceleration response spectrum, kinetic energy, and Fourier velocity. This result indicates that media fluctuations affect expected ground motion metrics at a single station and do not simply average out when considering an ensemble of realizations. Like the comparisons made during the sensitivity test, there is no systematic way in which the media fluctuations affect these metrics. However, the results without media perturbations generally fall within the extrema of the ensemble.

Seismic scattering disturbs the rupture front and redistributes energy that is recorded later in the seismogram. This mechanism provides an explanation for differences between ensemble means and models not considering small-scale heterogeneities. Even though there is seemingly more energy arriving at the stations, velocity and density fluctuations do not increase the overall energy present in the wavefield; they simply redistribute energy toward or away from the receivers. Additionally, scattering from the small-scale heterogeneities seems to play an appreciable role above 0.5 Hz at most stations, consistent with Hartzell *et al.* (2010).

We find particularly good comparisons at station PSR, which is relatively close to the source location and situated over a shallow basin (depth to the isosurface of  $V_s = 1 \text{ km s}^{-1}$ ,  $Z_{1.0} = 150$  m). In contrast, we find that stations in the deep basin (e.g. LTP) tend to have poorer GOF scores indicating there are discrepancies between the underlying velocity model and the real earth. This could also be related to local site effects that are not correctly captured by the CVM-S4.26 used for this study. TB14 found that including geotechnical models cannot explain all of the observed differences between the data and synthetics. In addition, most stations with poor GOF scores (e.g. RUS, SRN) tend to have much larger durations than obtained by our simulations. This could be due to strong scattering in the upper crust as noted by Wang & Shearer (2017). Thus, the first-order 0–2.5 Hz misfits between data and synthetics appear to be caused by inadequacies in the CVM, the attenuation model, or the source model.



**Figure 4.** Qualitative comparisons shown for stations LTP, RUS, STS and SRN (see Fig. 1 for location). For each station, the left panels show velocity seismograms ( $\text{cm s}^{-1}$ ) and the right panel shows cumulated energy ( $\text{cm}^2 \text{s}^{-1}$ ). Synthetics are shown for model C5 in Table 2.

### Area-wide effects of statistical descriptions of small-scale heterogeneities

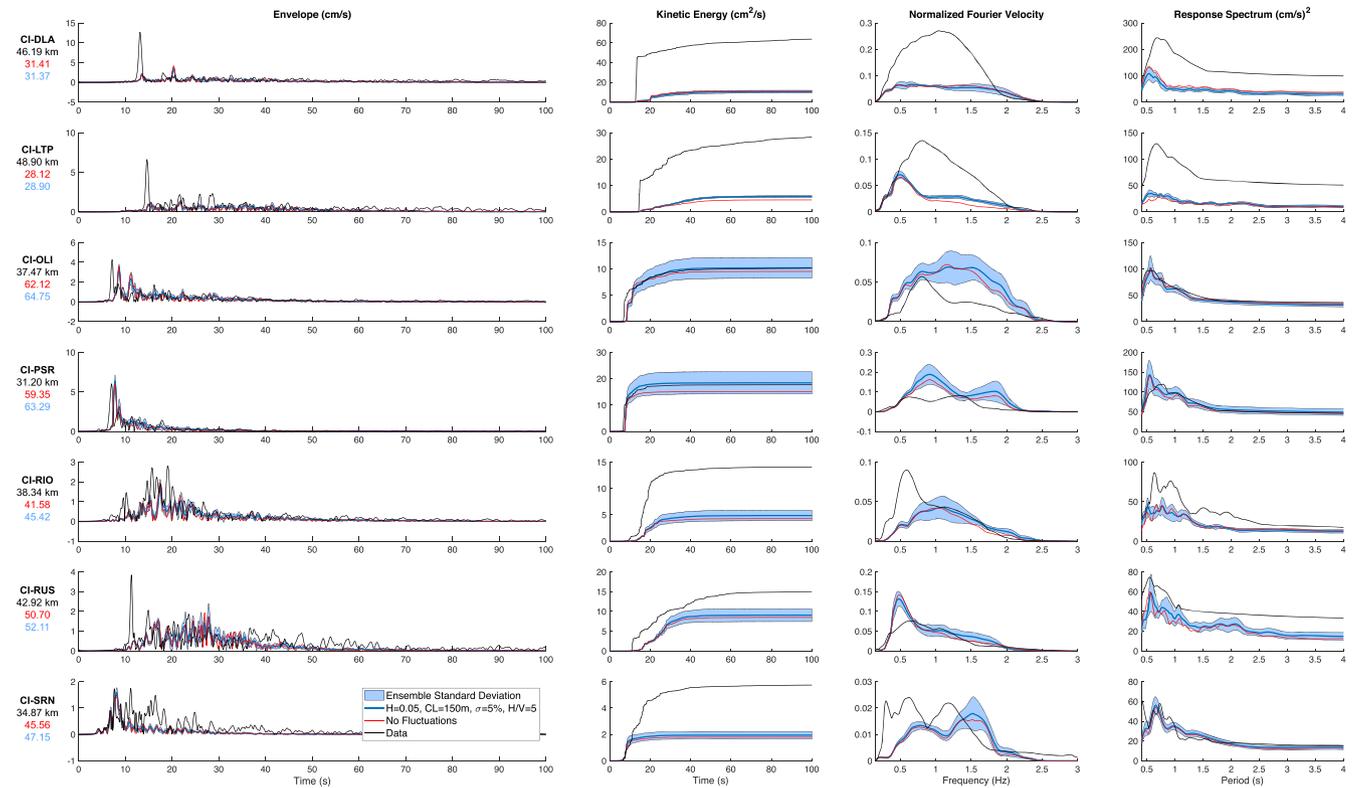
In this section, we investigate area-wide effects of small-scale heterogeneities with a focus on differences in the amplitudes of ground motion metrics between ensembles of models with and without heterogeneities. The simulations use  $Q_{s_0}/V_s = 100$  and we compute ensemble averages over five realizations of ground motions with small-scale velocity and density fluctuations.

Fig. 6 shows interpolated PGV from both the data and synthetics at each station in the simulation domain, treating the simulations and data identically. The ensemble-averaged amplitude values for PGV are changed by less than 1 per cent, on average, due to the small-scale heterogeneities. However, at individual stations, up to a  $\pm 25$  per cent difference in PGV can occur when compared with models not including small-scale heterogeneities. The magnitude of this effect increases as a function of distance from the source, which indicates that small-scale heterogeneities will play an increasingly important role at larger propagation distances. The results for PGA (see Fig. S9 in the electronic supplement) are within 1 per cent of those for

PGV. We note that in this manuscript differences are reported as per cent fluctuations (e.g.  $100 (PGV_{\text{het}} - PGV_{\text{hom}})/PGV_{\text{hom}}$ ).

It is apparent that the simulations cannot reproduce substantial basin-amplification effects or local site effects, including the larger PGV or PGA observed in the data over the deeper portion of Los Angeles basin. We find that the discrepancies between synthetics and data are larger than the effects of including velocity and density fluctuations, so the misfit between simulations and data are likely related to inadequacies in the underlying velocity model or source description.

The ENER, AI and DUR metrics involve the entire waveform ('progress-in-time') where, as expected, the effects of considering velocity and density fluctuations become much more apparent. AI (Fig. 6) displays the largest effect from ensembles of velocity and density fluctuations with a 13 per cent increase on average inside the model, while ENER (Fig. S9) increases 7 per cent as shown by the median value of the histogram. Also, the majority of stations in the simulation domain experience increased ENER, in contrast with both peak metrics. Like ENER, the increase is observed at the



**Figure 5.** Comparisons between synthetics including media fluctuations (blue), synthetics without media fluctuations (red) and data (black). We plot one-sided envelope functions, cumulative energy, Fourier velocity, and acceleration response spectra. Fourier velocity is normalized by the variance of the signal. The shaded blue region represents extrema from the five-simulation ensemble. The labels to the left of each figure provide the station name,  $r_{rup}$  distance, and the value of the average GOF score for the station including fluctuations (blue) and without fluctuations (red). The GOF score for the ensemble (blue) represents the GOF averaged across all five simulations. The station locations of this analysis are shown in Fig. 1.

majority of stations in the domain. Finally, DUR (Fig. S9) shows that stations on average experience 4 per cent longer shaking.

Figs 7(a)–(e) shows the selected metrics plotted as a function of distance ( $r_{jb}$ ) for CVM-S4.26 + media fluctuations, CVM-S4.26, and recorded data. We observe good fits at close stations with increasing underprediction in amplitude as a function of distance. The results indicate that the attenuation model  $Q_{s0}/V_s = 100$  provides too much damping. The overly attenuating model is apparent by the difference in slopes between the data (black line) and the simulations (red and blue lines). However, this difference could potentially be explained by underestimation of the moment, which we consider in the Discussion section. TB13 observed a similar trend when considering their non-linear  $Q_s/V_s$  relationship (Fig. 3). The under-prediction around 30–40 km from the source occurs above the deeper portion of Los Angeles basin, suggesting a refinement of the parameters for the underlying CVM-S4.26.

## DISCUSSION

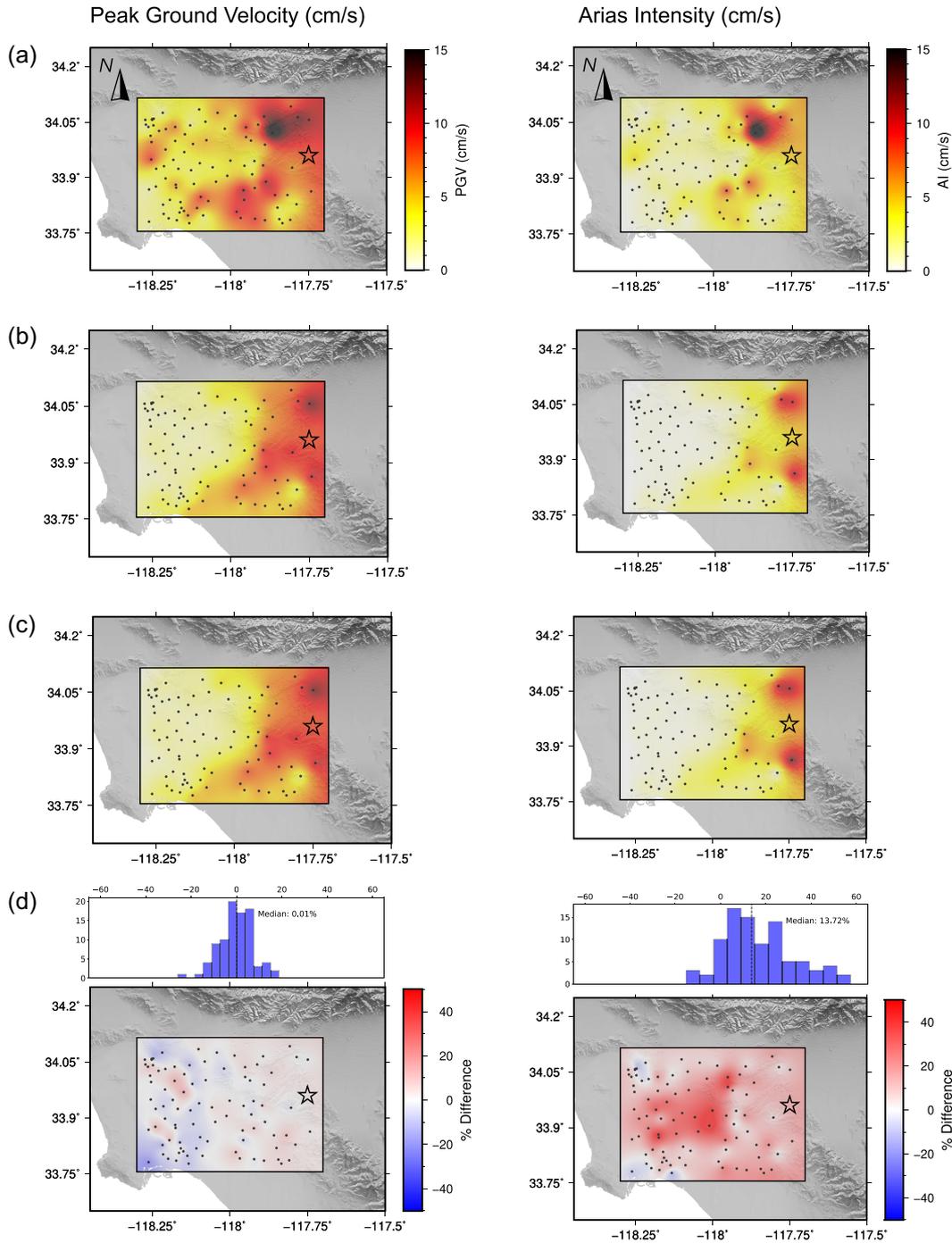
Our simulations show promising results toward accurately modeling deterministic wave propagation to higher frequencies. However, we find persistent trends in the misfit between simulations and data, including at stations located over deeper portions of the basin and for the distance decay of the metrics. While TB13 pointed out regions in the velocity model that likely contribute to these misfits, we investigate the role of the attenuation model in our simulations. Moreover, we explore the effects of uncertainty in the moment

estimates on the GOF estimates. According to the several moment tensor inversions provided by the USGS, the magnitude of the 2008 Chino Hills Earthquake was reported between  $M_w$  5.3 and  $M_w$  5.5.

Fig. 7(d) shows comparisons between our simulations and data for the attenuation model  $Q_{s0}/V_s = 100$  when considering  $M_w$  5.5 ( $M_0 = 2.2 \times 10^{17}$  Nm). We find that increasing the moment provides better distance decay and comparisons for all metrics considered across the simulated bandwidth of 0.15–2.5 Hz. However, even with an increased  $M_w$ , we find similar discrepancies observed with PGV as seen with the  $M_w$  5.4 source, namely the underprediction of peak amplitudes at certain stations with large distances from the fault. We suspect that these discrepancies are related to either basin effects or local site effects that are not properly accounted for in CVM-S4.26. Based on these comparisons, we believe that the moment magnitude of the 2008 Chino Hills earthquake is between 5.4 and 5.5.

## Anelastic attenuation

First, we show the simulation results considering  $M_w$  5.4 and  $Q_s/V_s = 50$  (Figs. 7, S8) to compare our results against previously published relationships between  $Q_{s0}$  and  $V_s$  (e.g. Graves & Pitarka 2010). The simulation results show a much steeper slope in predicted amplitudes as a function of distance for  $Q_{s0}/V_s = 50$  as compared with data, indicating that the former is likely too attenuating for southern California when considering simulations with  $\min(V_s) = 200 \text{ m s}^{-1}$ . The simulation results using  $Q_{s0}/V_s = 100$

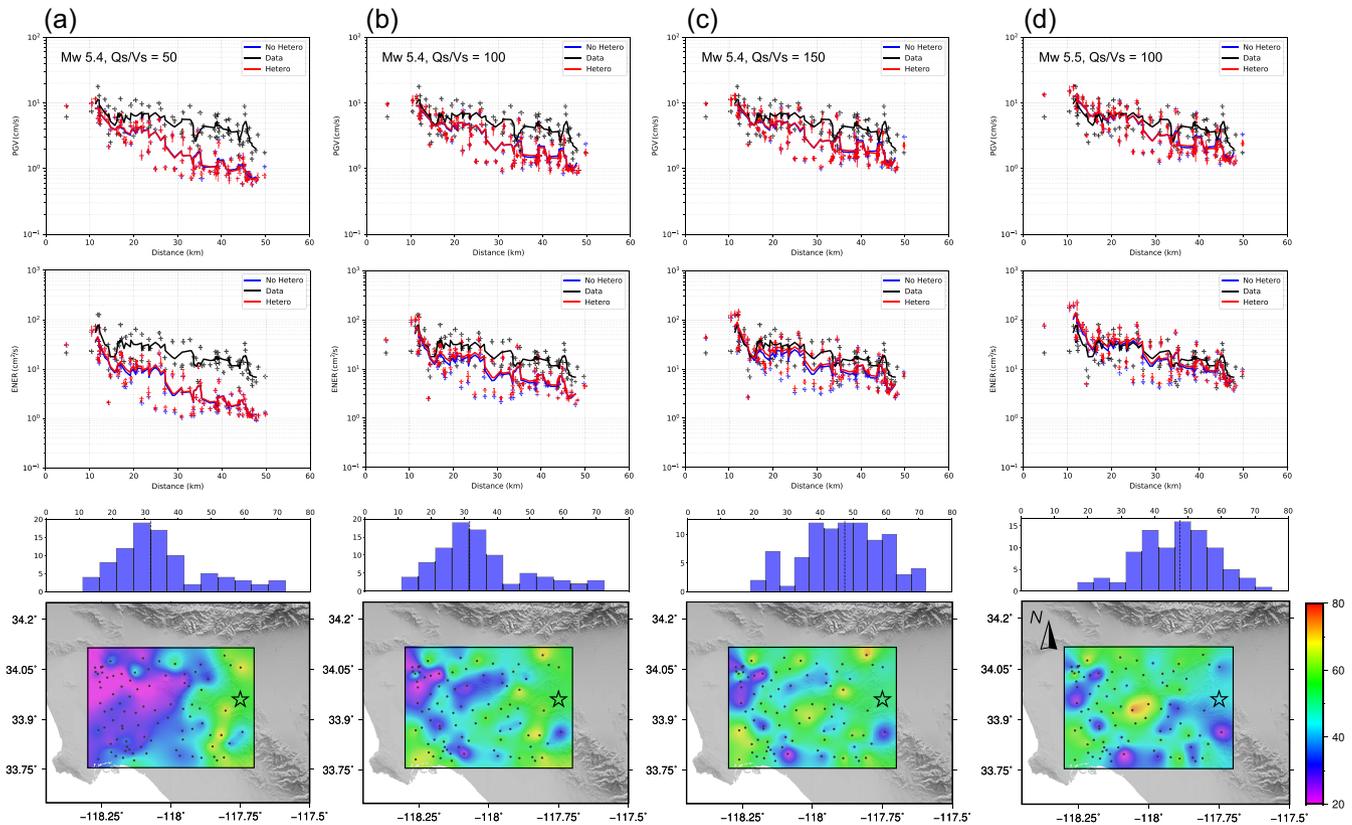


**Figure 6.** Model-wide comparison of (left-hand panel) peak-ground velocity (PGV) and (right-hand panel) Arias Intensity (AI) between data and synthetics computed with and without models of media fluctuations. (a) Interpolated metrics recorded from the 2008 Chino Hills earthquake. (b) Interpolated metrics from simulation without models of media fluctuations. (c) Interpolated metrics of the ensemble average of five simulations considering models of media fluctuations with seeds 1–5. (d) Percent difference in metric amplitude between ensemble of models considering media fluctuations against simulation that does not consider media fluctuations. The histogram shows the values represented in (d).

show a much-improved fit to data while  $Q_{S_0}/V_s = 150$  produces the smallest misfit considering  $M_w 5.4$  (although still including poor GOF scores in the deeper portion of Los Angeles basin).

Fig. 7(d) shows the results for a model with  $M_w 5.5$  and  $Q_{S_0}/V_s = 100$  which provides even smaller misfits than with  $M_w 5.4$  and  $Q_{S_0}/V_s = 150$  (most clearly for PGV and ENER, but not for

DUR, see Fig. S8), especially at farther distances from the source. Both models reproduce the general distance decay observed in the data equally well (see also the distance-dependent bias in Fig. S10). In addition to the  $M_w 5.5$  simulation using  $Q_{S_0}/V_s = 100$ , we also simulated a  $M_w 5.5$  model with  $Q_{S_0}/V_s = 150$  (not shown here), producing systematic over-predictions of the data.



**Figure 7.** Plots of peak ground velocity (PGV,  $\text{cm s}^{-1}$ ), cumulative energy (ENER,  $\text{cm}^2 \text{s}^{-1}$ ) and simulation-wide GOF scores with their histograms for models assuming (a)  $M_w$  5.4 and  $Q_{s0}/V_s = 50$ , (b)  $M_w$  5.4 and  $Q_{s0}/V_s = 100$ , (c)  $M_w$  5.4,  $Q_{s0}/V_s = 150$  and (d)  $M_w$  5.5,  $Q_{s0}/V_s = 100$ . The red lines indicate simulations that include small-scale heterogeneities with  $\nu = 0.05$ ,  $a_x = 150 \text{ m}$ ,  $a_x/a_z = 5$  and  $\sigma = 5$  per cent, and the error bars indicate the range of solutions at each station. Blue lines indicate models that do not include small-scale heterogeneities, and the black lines represent observations at the black circles in the spatial plots of GOF scores.

### Source topography

Several 3-D simulation studies (Rodgers *et al.* 2010; Takemura *et al.* 2015; Imperatori & Mai 2015) have shown that topography near the source region can significantly influence ground motions. However, these topographic relief in these studies is much larger than in our model of the Los Angeles basin (see Fig. 1). In addition, we expect stronger topographic scattering from the relatively shallower seismic sources used in those studies (0.6–5 km), as compared to the Chino Hills fault. Finally, these studies included maximum frequencies up to 5–10 Hz, where topographic scattering is expected to be strong. On the other hand, Rodgers *et al.* (2010) found that for frequencies between 0.5 and 2 Hz, similar to our modeling bandwidth, the wavefield did not experience significant influence from the topography. These results suggest that topographic effects would play a minor role in our ground motion estimates.

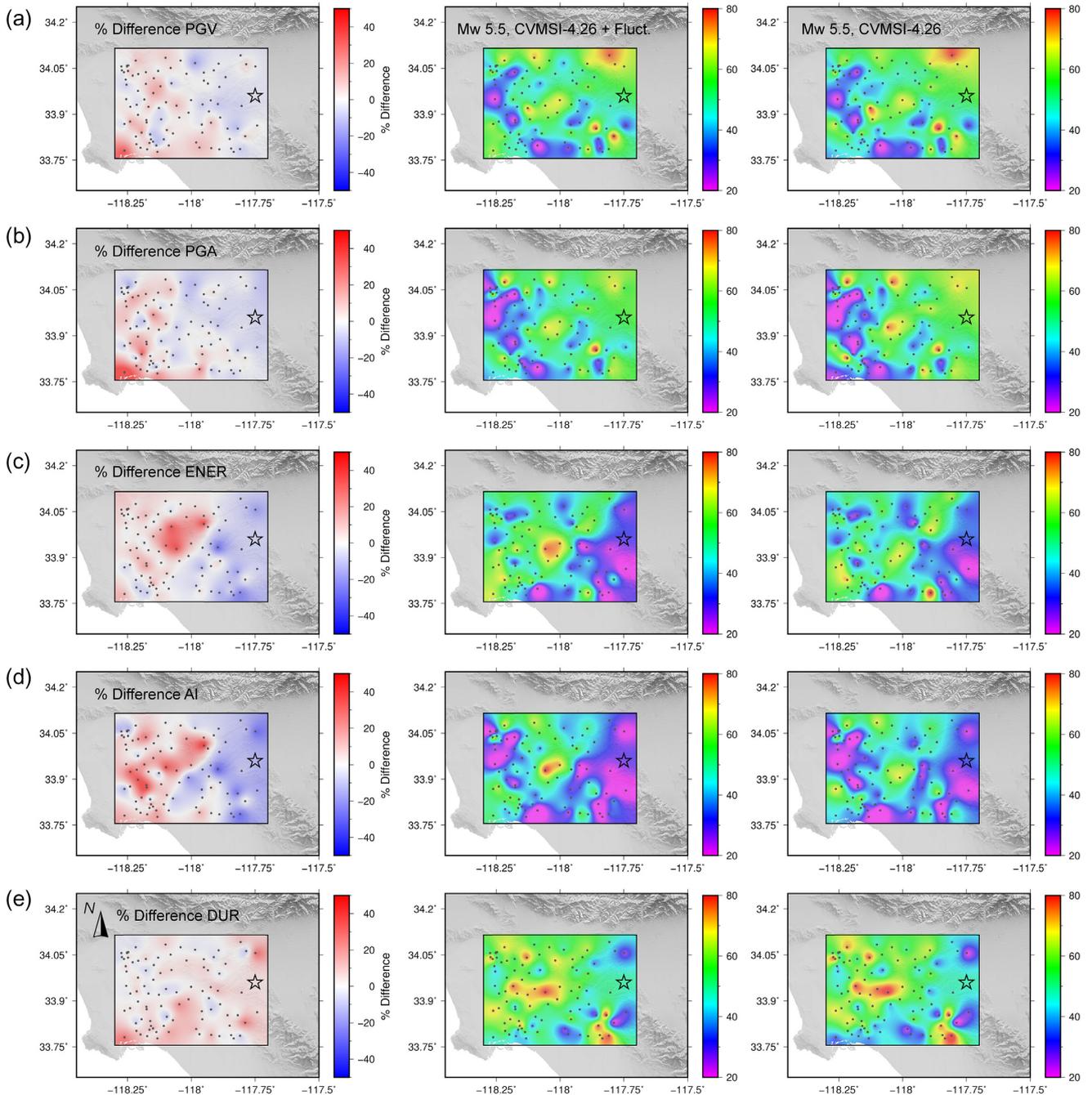
### Size of modelled domain

Metrics considering the entire time-series such as ENER, AI and DUR are influenced by later arriving phases. The deep basin below Los Angeles can trap multiply-scattered waves that are back- (or wide angle) scattered due to the small-scale heterogeneities. Assuming this occurs in our simulations, the resulting metrics ENER, AI and DUR would be somewhat underestimated as compared to simulations carried out in a larger domain. We attempt to mitigate this potential source of bias by removing stations located close to

the model boundary. This effect would not systematically improve our GOF scores. For instance, DUR comparisons would be worse as our simulations already over predict the DUR values for each station. On the other hand, ENER and AI fits would improve as we tend to underpredict these metrics at the same stations (Fig. S8).

### Effects of velocity heterogeneities on goodness of fit

Fig. 8 shows GOF scores for individual metrics and differences between the GOF scores computed for CVM-S4.26 and CVM-S4.26 + media fluctuations, using the simulation ensemble with  $M_w$  5.5 and attenuation model  $Q_{s0}/V_s = 100$ , as this model provides near optimal GOF values. The median of model ensembles of velocity and density fluctuations can affect the GOF scores up to 50 per cent for some metrics (ENER, AI and DUR) with others such as PGV and PGA much less influenced. Media fluctuations likely play an increasingly important role in high-frequency ground motions to larger distances than considered here. Recent work by Wang & Shearer (2017) shows the interdependence of anelastic attenuation and scattering attenuation, which implies that scattering should be included for high-frequency simulations when determining the parameters for frequency-dependent attenuation, such as the power-law exponent  $\gamma$ . Future work should focus on properly addressing the trade-off between intrinsic attenuation and scattering attenuation in high-frequency simulations including  $Q(f)$  and media fluctuations.



**Figure 8.** Percent difference of GOF (left-hand column) computed between GOF scores for CVM-S4.26 (centre column) and CVM-S4.26 + media fluctuations (right-hand column). We show PGV (a), PGA (b), ENER (c), AI (d) and DUR (e) as individual metrics to be consistent with previous analyses. Percent difference plots are coloured such that red areas indicate the CVM-S4.26 + media fluctuations provide better GOF to data as compared with CVM-S4.26, and blue areas indicate the CVM-S4.26 provides better GOF. Synthetics and data are lowpass filtered to 2.5 Hz before computing GOF scores and per cent difference.

We find the ground motion metrics are correlated amongst one another in both the simulations and the data (Figs 6–8, S8–S9) with exception of the DUR metric showing anticorrelations with PGV, PGA, ENER, and AI. For example, stations experiencing larger than average PGV also experience larger than average PGA, ENER, and AI, but smaller DUR. This agrees with recent work by Bradley (2015) investigating correlations in similar metrics from earthquake records present in the NGA-West2 database (Ancheta *et al.* 2014).

There appears to be a distance dependency in the GOF scores in Fig. 8, in particular for ENER and AI, showing better general

agreement in the left half of the model. We tested different parameters for the scattering model before choosing the parameters used in our study, which generated the overall best fit to data. The trend may be an effect of stations at farther distances primarily located in the deep Los Angeles basin, as opposed to the stations on shallow sediments and rock in the eastern half of the model. Thus, the apparent distance dependency might indicate that different scattering models should be considered for basins and rock/shallow sediment, which we leave for future work.

In addition to predicting realistic amplitudes for ground motions, understanding the intra-event variability and spatial coherence of ground motions is important for seismic hazard analysis. In this paper, we have not addressed the spatial correlations of ground motions or correlations amongst metrics (e.g. Loth & Baker 2013; Bradley 2015); however, we suspect that media fluctuations will play an increasingly important role in the analysis of ground motion variability as small-scale heterogeneities effectively reduce the intra-event standard deviations in ground motions through a spatial redistribution of energy (Withers *et al.* 2018a,b). Future work should include running a large number of simulations to quantify the spatial variability of these ground motion models including topography and small-scale velocity fluctuations. Comparisons against data are important to ensure that our models provide useful results and the added complexity is justified. For example, recent work from Loth & Baker (2013) can be used to validate these proposed models against observations.

## CONCLUSIONS

We have simulated 0–2.5 Hz ground motions for the 2008  $M_w$  5.4 Chino Hills earthquake using frequency-dependent attenuation and ensembles of ground motions computed in realizations of stochastic models of small-scale crustal velocity and density fluctuations described by a von Kármán autocorrelation function. We find that the parameter of the small-scale perturbations most influential on the ground motions is the seed number of the random distribution used to define the locations of the fluctuations. We find that five simulation ensembles start to divert from simulations that do not consider small-scale heterogeneities, with the trend increasing for larger frequencies with effects as low as 0.5 Hz at certain stations. The ensemble median values for ground motion metrics (PGV, PGA, ENER, AI and DUR) show amplification and de-amplification of up to  $\pm 50$  per cent compared to results without small-scale heterogeneities at various locations in the model. The progress-in-time metrics (ENER, AI and DUR) display systematic increases at nearly all stations. The largest effect is observed with AI showing a difference of 13 per cent, on average, between models with and without velocity and density fluctuations.

We find that our simulations drastically under predict ground motions with an increase in misfit as a function of distance using  $Q_{s0}/V_s = 50$ , as used in other studies. Our preferred model is obtained using a moment magnitude of  $M_w$  5.5 and  $Q_{s0}/V_s = 100$  using the ensemble of small-scale heterogeneities. This model trades off with smaller magnitudes and less anelastic attenuation (e.g.  $M_w$  5.4 and  $Q_{s0}/V_s = 150$  produces almost equally good GOF scores).

Our results suggest that the first-order sources of misfit associated with the Chino Hills simulation are related to the underlying velocity model and/or the source model, and cannot be remedied only using median values of ensemble metrics computed from wave propagations through media containing small-scale heterogeneities for the models involved here. However, we suspect that small-scale heterogeneities will play a much more important role when investigating the intraevent variability, spatial correlation structure of observed ground motions, and when considering higher-frequency waves propagated to regional distances.

## REFERENCES

Levander, A.R., 1992. A stochastic view of lower crustal fabric based on evidence from the ivrea zone, *Geophys. Res. Lett.*, **19**(11), 1153–1156.

- Aagaard, B.T. & Heaton, T.H., 2004. Near-source ground motions from simulations of sustained interseismic and supersonic fault ruptures, *Bull. seism. Soc. Am.*, **94**(6), 2064–2078.
- Aagaard, B.T. *et al.*, 2010a. Ground-motion modeling of hayward fault scenario earthquakes, part II: simulation of long-period and broadband ground motions, *Bull. seism. Soc. Am.*, **100**(6), 2945–2977.
- Aagaard, B.T. *et al.*, 2010b. Ground-motion modeling of hayward fault scenario earthquakes, part I: construction of the suite of scenarios, *Bull. seism. Soc. Am.*, **100**(6), 2927–2944.
- Abrahamson, N.A., Silva, W.J. & Kamai, R., 2014. Summary of the ASK14 ground motion relation for active crustal regions, *Earthq. Spectra*, **30**(3), 1025–1055.
- Aki, K., 1980. Attenuation of shear-waves in the lithosphere for frequencies from 0.05 to 25 Hz, *Phys. Earth planet. Inter.*, **21**(1), 50–60.
- Ancheta, T.D. *et al.*, 2014. NGA-West2 Database, *Earthq. Spectra*, **30**(3), 989–1005.
- Anderson, J., 2004. Quantitative measure of the goodness-of-fit of synthetic seismograms, in *13th World Conference on Earthquake Engineering Conference Proceedings*, Vancouver, Canada, Paper (Vol. **243**).
- Arias, A., 1970. Measure of earthquake intensity, in *Seismic Design for Nuclear Power Plants*, pp. 438–483, Massachusetts Inst. of Tech., Cambridge. Univ. of Chile, Santiago de Chile
- Bielak, J., Karaoglu, H. & Tabor, R., 2011. Memory-efficient displacement-based internal friction for wave propagation simulation, *Geophysics*, **76**(6), T131–T145.
- Boore, D.M., Stewart, J.P., Seyhan, E. & Atkinson, G.M., 2014. *NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes*, Earthquake Engineering Research Institute.
- Bradley, B., 2015. Correlation of Arias intensity with amplitude, duration and cumulative intensity measures, *Soil Dyn. Earthq. Eng.*, **78**, 89–98.
- Brocher, T.M., 2008. Compressional and shear-wave velocity versus depth relations for common rock types in northern California, *Bull. seism. Soc. Am.*, **98**(2), 950–968.
- Bydlon, S.A. & Dunham, E.M., 2015. Rupture dynamics and ground motions from earthquakes in 2-D heterogeneous media, *Geophys. Res. Lett.*, **42**(6), 1701–1709.
- Campbell, K.W. & Bozorgnia, Y., 2014. *NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra*, Earthquake Engineering Research Institute.
- Cerjan, C. *et al.*, 1985. A nonreflecting boundary condition for discrete acoustic and elastic wave equations, *Geophysics*, **50**(4), 705–708.
- Chiou, B.S.J. & Youngs, R.R., 2014. Update of the chiou and youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra*, **30**(3), 1117–1153.
- Cui, Y. *et al.*, 2010. *Scalable Earthquake Simulation on Petascale Supercomputers*, pp. 1–20, IEEE.
- Day, S.M., 1998. Efficient simulation of constant Q using coarse-grained memory variables, *Bull. seism. Soc. Am.*, **88**(4), 1051–1062.
- Day, S.M. & Bradley, C.R., 2001. Memory-efficient simulation of anelastic wave propagation, *Bull. seism. Soc. Am.*, **91**(3), 520–531.
- Day, S.M. *et al.*, 2008. *Model for Basin Effects on Long-Period Response Spectra in Southern California*, *Earthq. Spectra*, **24**(1), 257–277.
- Dolan, S.S. & Bean, C.J., 1997. Some remarks on the estimation of fractal scaling parameters from borehole wire-line logs, *Geophys. Res. Lett.*, **24**(10), 1271–1274.
- Duan, B. & Day, S.M., 2010. Sensitivity study of physical limits on ground motion at yucca mountain, *Bull. seism. Soc. Am.*, **100**(6), 2996–3019.
- Erickson, D., McNamara, D.E. & Benz, H.M., 2004. Frequency-dependent  $\lg Q$  within the continental united states, *Bull. seism. Soc. Am.*, **94**(5), 1630–1643.
- Frankel, A. & Clayton, R., 1986. Finite difference simulations of seismic scattering: implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity, *J. geophys. Res.: Solid Earth*, **91**(B6), 6465–6489.
- Goulet, C.A. *et al.*, 2015. The SCEC broadband platform validation exercise: methodology for code validation in the context of seismic-hazard analyses, *Seismol. Res. Lett.*, **86**(1), 17–26.

- Graves, R.W. & Pitarka, A., 2010. Broadband ground-motion simulation using a hybrid approach, *Bull. seism. Soc. Am.*, **100**(5A), 2095–2123.
- Graves, R.W. *et al.*, 2011. CyberShake: a physics-based seismic hazard model for southern California, *Pageoph*, **168**(3-4), 367–381.
- Hartzell, S., Harmsen, S. & Frankel, F., 2010. Effects of 3D random correlated velocity perturbations on predicted ground motions, *Bull. seism. Soc. Am.*, **100**(4), 1415–1426.
- Hauksson, E. & Shearer, P.M., 2006. Attenuation models (QP and QS) in three dimensions of the southern California crust: inferred fluid saturation at seismogenic depths, *J. geophys. Res.*, **111**(B5), B05302.
- Hauksson, E. *et al.*, 2008. Preliminary report on the 29 July 2008 Mw 5.4 Chino Hills, Eastern Los Angeles Basin, California, earthquake sequence, *Seismol. Res. Lett.*, **79**(6), 855–866.
- Imperatori, W. & Mai, P.M., 2012. Broad-band near-field ground motion simulations in 3-dimensional scattering media, *Geophys. J. Int.*, **192**(2), 725–744.
- Imperatori, W. & Mai, P., 2015. The role of topography and lateral velocity heterogeneities on near-source scattering and ground-motion variability, *Geophys. J. Int.*, **202**(3), 2163–2181.
- Ji, C., Helmberger, D.V. & Wald, D.J., 2003. Slip history and dynamic implications of the 1999 Chi-Chi, Taiwan, earthquake, *J. geophys. Res.*, **108**(B9), 2412.
- Kawase, H., 1996. The Cause of the damage belt in kobe: “the basin-edge effect,” constructive interference of the direct s-wave with the basin-induced diffracted/rayleigh waves, *Seismol. Res. Lett.*, **67**(5), 25–34.
- Klimeš, L., 2002. Correlation functions of random media, *Pageoph*, **159**(7-8), 1811–1831.
- Kohler, M.D., Magistrale, H. & Clayton, R., 2003. Mantle heterogeneities and the scec reference three-dimensional seismic velocity model version 3, *Bull. seism. Soc. Am.*, **93**(2), 757–774.
- Korn, M., 1993. Determination of site-dependent scattering-q from p-wave coda analysis with an energy-flux model, *Geophys. J. Int.*, **113**(1), 54–72.
- Kristekova, M., Kristek, J. & Moczo, P., 2009. Time-frequency misfit and goodness-of-fit criteria for quantitative comparison of time signals, *Geophys. J. Int.*, **178**(2), 813–825.
- Lee, E.J. *et al.*, 2011. Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California, *Geophys. J. Int.*, **186**(1), 311–330.
- Lee, E.J., Chen, P. & Jordan, T.H., 2014. Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes, *Seismol. Res. Lett.*, **85**(6), 1275–1284.
- Li, Y-G. *et al.*, 1994. Seismic guided-waves trapped in the fault zone of the Landers, California, earthquake of 1992, *J. geophys. Res.: Solid Earth*, **99**(B6), 11705–11722.
- Loth, C. & Baker, J., 2013. A spatial cross-correlation model of spectral accelerations at multiple periods, *Earthq. Eng. Struct. Dyn.*, **42**(3), 397–417.
- Maechling, P.J. *et al.*, 2015. SCEC broadband platform: system architecture and software implementation, *Seismol. Res. Lett.*, **86**(1), 27–38.
- Magistrale, H. *et al.*, 2000. The SCEC southern california reference three-dimensional seismic velocity model version 2, *Bull. seism. Soc. Am.*, **90**(6B), S65–S76.
- Nakata, N. & Beroza, G.C., 2015. Stochastic characterization of mesoscale seismic velocity heterogeneity in Long Beach, California, *Geophys. J. Int.*, **203**(3), 2049–2054.
- Nakata, N. *et al.*, 2015. Body wave extraction and tomography at Long Beach, California, with ambient-noise interferometry, *J. geophys. Res.: Solid Earth*, **120**(2), 1159–1173.
- Olsen, K.B., 2000. Site amplification in the los angeles basin from three-dimensional modeling of ground motion, *Bull. seism. Soc. Am.*, **90**(6B), S77–S94.
- Olsen, K.B. & Mayhew, J.E., 2010. Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 Mw 5.4 Chino Hills, California, earthquake, *Seismol. Res. Lett.*, **81**(5), 715–723.
- Olsen, K.B., Day, S.M. & Bradley, C.R., 2003. Estimation of Q for long-period (>2 sec) waves in the los angeles basin, *Bull. seism. Soc. Am.*, **93**(2), 627–638.
- Oreskes, N., Shraderfrechette, K. & Belitz, K., 1994. Verification, validation, and confirmation of numerical-models in the earth-sciences, *Science*, **263**(5147), 641–646.
- Pardo-Igúzquiza, E. & Chica-Olmo, M., 1993. The Fourier integral method: an efficient spectral method for simulation of random fields, *Math Geol.*, **25**(2), 177–217.
- Phillips, W.S., Mayeda, K.M. & Malagnini, L., 2013. How to invert multi-band, regional phase amplitudes for 2-D attenuation and source parameters: tests using the USarray, *Pageoph*, **171**(3-5), 469–484.
- Pitarka, A. *et al.*, 2009. Numerical study of ground-motion differences between buried-rupturing and surface-rupturing earthquakes, *Bull. seism. Soc. Am.*, **99**(3), 1521–1537.
- Przybilla, J., Wegler, U. & Korn, M., 2009. Estimation of crustal scattering parameters with elastic radiative transfer theory, *Geophys. J. Int.*, **178**(2), 1105–1111.
- Raoof, M., Herrmann, R.B. & Malagnini, L., 1999. Attenuation and excitation of three-component ground motion in southern California, *Bull. seism. Soc. Am.*, **89**, 1–15.
- Rodgers, A.J., Petersson, N.A. & Sjogreen, B., 2010. Simulation of topographic effects on seismic waves from shallow explosions near the North Korean nuclear test site with emphasis on shear wave generation, *J. geophys. Res.: Solid Earth*, **115**(B11).
- Roten, D. *et al.*, 2014. Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity, *Geophys. Res. Lett.*, **41**(8), 2769–2777.
- Savran, W.H. & Olsen, K.B., 2016. Model for small-scale crustal heterogeneity in Los Angeles basin based on inversion of sonic log data, *Geophys. J. Int.*, **205**(2), 856–863.
- Shao, G., Ji, C. & Hauksson, E., 2012. Rupture process and energy budget of the 29 July 2008 Mw 5.4 Chino Hills, California, earthquake, *J. geophys. Res.: Solid Earth*, **117**(B7), doi:10.1029/2011JB008856.
- Shaw, J.H. *et al.*, 2015. Unified structural representation of the southern California crust and upper mantle, *Earth planet. Sci. Lett.*, **415**(C), 1–15.
- Shearer, P.M., Prieto, G.A. & Hauksson, E., 2006. Comprehensive analysis of earthquake source spectra in southern California, *J. geophys. Res.*, **111**(B6), B06303–n/a.
- Small, P. *et al.*, 2017. The SCEC unified community velocity model software framework, *Seismol. Res. Lett.*, **86**(6), 1539–1552.
- Spudich, P. & Olsen, K.B., 2001. Fault zone amplified waves as a possible seismic hazard along the Calaveras fault in central California, *Geophys. Res. Lett.*, **28**(13), 2533–2536.
- Süss, M.P. & Shaw, J.H., 2003. P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *J. geophys. Res.*, **108**(B3), 2170.
- Taborda, R. & Bielak, J., 2013. Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake, *Bull. seism. Soc. Am.*, **103**(1), 131–156.
- Taborda, R. & Bielak, J., 2014. Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake using different velocity models, *Bull. seism. Soc. Am.*, **104**(4), 1876–1898.
- Takemura, S., Furumura, T. & Maeda, T., 2015. Scattering of high-frequency seismic waves caused by irregular surface topography and small-scale velocity inhomogeneity, *Geophys. J. Int.*, **201**(1), 459–474.
- Wang, W. & Shearer, P.M., 2017. Using direct and coda wave envelopes to resolve the scattering and intrinsic attenuation structure of Southern California, *J. geophys. Res.: Solid Earth*, **100**(B12), 24 015.
- Withers, K.B., Olsen, K.B. & Day, S.M., 2015. Memory-efficient simulation of frequency-dependent Q, *Bull. seism. Soc. Am.*, **105**(6), 1–14.
- Withers, K.B., Olsen, K.B., Day, S.M. & Shi, Z., 2018a. Ground motion and intra-event variability from 3-D deterministic broadband (0–7.5 Hz) simulations along a non-planar strike-slip fault, *Bull. seism. Soc. Am.*, **109**(1), 229–250.
- Withers, K.B., Olsen, K.B., Day, S.M. & Shi, Z., 2018b. Validation of deterministic broadband ground motion and variability from dynamic rupture simulations of buried thrust earthquakes, *Bull. seism. Soc. Am.*, **109**(1), 212–228.
- Wu, R.S. & Aki, K., 1985. The fractal nature of the inhomogeneities in the lithosphere evidenced from seismic-wave scattering, *Pageoph*, **123**(6), 805–818.

## SUPPORTING INFORMATION

Supplementary data are available at [GJI](https://doi.org/10.1093/gji/ggz111) online.

**Figure S1.** Horizontal slices shown at  $z = 160$  m for different models of  $V_s$  velocity fluctuations ( $\text{m s}^{-1}$ ) superimposed onto the background CVM, that we considered during sensitivity testing. The contours indicate depth in km to the  $V_s = 2.5 \text{ km s}^{-1}$  iso surface ( $Z_{2.5}$ ). We consider a base model (a) with parameters  $\nu = 0.05$ ,  $a = 150$  m,  $H/V = 5$ ,  $\sigma = 5$  per cent and random seed = 1. Panels (b)–(g) show the effects of different parameterizations of the von Karman autocorrelation function. The parameters for each model are shown directly above the corresponding model. Table 2 shows the parameters of the velocity fluctuation models compared during the sensitivity test.

**Figure S2.** (a) Slip-rate functions from Model 1 of the Shao *et al.* (2012) finite-fault inversion. (b) Fourier amplitude spectrum of the slip-rate functions. Despite showing frequency content above 2.5 Hz for some subfaults, we limit our simulation to  $f_{\text{max}} = 2.5$  Hz corresponding to the  $f_{\text{max}}$  of the data used for the finite-fault slip inversion.

**Figure S3.** Qualitative comparisons shown for stations LTP, RUS, STS and SRN for correlation length  $a_x = 150$  m and  $a_x = 5000$  m. The left-hand panels show velocity seismograms ( $\text{cm s}^{-1}$ ) and the right-hand panel shows cumulated velocity for each component.

**Figure S4.** Qualitative comparisons shown for stations LTP, RUS, STS and SRN for Hurst exponent  $\nu = 0.0$  and  $\nu = 0.3$ . The left-hand panels show velocity seismograms ( $\text{cm s}^{-1}$ ) and the right-hand panel shows cumulated velocity for each component.

**Figure S5.** Qualitative comparisons shown for stations LTP, RUS, STS and SRN for anisotropy ratios  $a_x/a_z = 2$  and  $a_x/a_z = 5$ . The left-hand panels show velocity seismograms ( $\text{cm s}^{-1}$ ) and the right-hand panel shows cumulated velocity for each component.

**Figure S6.** Qualitative comparisons shown for stations LTP, RUS, STS and SRN for standard deviation  $\sigma = 5$  per cent and  $\sigma = 10$  per cent. The left-hand panels show velocity seismograms ( $\text{cm s}^{-1}$ ) and the right-hand panel shows cumulated velocity for each component.

**Figure S7.** Comparisons between synthetics including velocity fluctuations (blue), synthetics without velocity fluctuations (red) and data (black). We plot one-side envelop functions, cumulative energy, Fourier velocity and acceleration response spectra. Fourier

velocity is normalized by the variance of the signal. The shaded blue region represents extrema from the five-simulation ensemble. The labels to the left of each figure provide the station name,  $r_{\text{top}}$  distance, and the value of the average GOF score for the station including fluctuations (blue) and without fluctuations (red). The GOF score for the ensemble (blue) represents the GOF averaged across all five simulations. The station locations of this analysis are shown in Fig. 5.

**Figure S8.** Distance plots showing the comparison between models using different attenuation models; namely,  $Q_{s0}/V_s = 50$ ,  $Q_{s0}/V_s = 100$ , and  $Q_{s0}/V_s = 150$ . This figure shows all metrics computed for Fig. 7 in the main text. The red line indicates models that include small-scale heterogeneities, the blue line indicates models that do not, and the black line represents data. (a) PGV ( $\text{cm s}^{-1}$ ), (b) PGA ( $\text{cm s}^{-2}$ ), (c) ENER ( $\text{cm}^2 \text{ s}^{-1}$ ), (d) AI ( $\text{cm s}^{-1}$ ), (e) DUR (s), (f) shows a histogram of GOF scores for this simulation and the spatial distribution of GOF scores in the simulation domain.

**Figure S9.** Model-wide comparison of (left) peak-ground acceleration (PGA), (centre) cumulative energy (ENER), and (right) duration (DUR) between data and synthetics computed with and without models of velocity fluctuations. (a) Interpolated metrics recorded from the 2008 Chino Hills earthquake. (b) Interpolated metrics from simulation without models of velocity fluctuations. (c) Interpolated metrics of the ensemble average of five simulations considering models of velocity fluctuations with seeds 1–5. (d) Percent difference in metric amplitude between ensemble of models considering velocity fluctuations against simulation that does not consider velocity fluctuations. The histogram shows the values represented in (d).

**Figure S10.** Plots showing the natural logarithm of the ratio between simulations and recorded data for metrics PGV (a,b), ENER (c,d) and DUR (e,f). We show biases for both the  $M_w$  5.5 simulation using  $Q_s/V_s = 100$  (left-hand column) and the  $M_w$  5.4 simulation using  $Q_s/V_s = 150$  (right-hand column). Bias plots are made using bandpass filtered data from 0.15 to 1.0 Hz.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.