

JGR Solid Earth

RESEARCH ARTICLE

10.1029/2021JB021915

Key Points:

- We develop an algorithm for merging gridded multiscale and multidimensional datasets and smoothly embed basin models in two regional models
- One hybrid model produces an ~24% decrease and the other has an ~0.6% increase in median waveform misfit relative to their regional models
- Misfit overlap and variability with stations and events between models show the complexities of model validation

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Ajala,
rajala1@lsu.edu

Citation:

Ajala, R., & Persaud, P. (2021). Effect of merging multiscale models on seismic wavefield predictions near the southern San Andreas fault. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB021915. <https://doi.org/10.1029/2021JB021915>

Received 19 FEB 2021

Accepted 2 SEP 2021

Effect of Merging Multiscale Models on Seismic Wavefield Predictions Near the Southern San Andreas Fault

R. Ajala¹  and P. Persaud¹ 

¹Department of Geology and Geophysics, Louisiana State University, Baton Rouge, LA, USA

Abstract Updating Earth models used by the scientific community in geologic studies and hazard assessment has a significant societal impact but is computationally prohibitive due to the large spatial scale. The advent of urban seismology allowed rapid development of local high-resolution models using short-term dense seismic arrays to become conventional. To incorporate the details in these local models in community models, we developed a technique for constructing window taper functions like the cosine taper in arbitrarily shaped spatial domains on regular grids. We apply our algorithm to the problem of low-frequency ground shaking estimation near the southernmost San Andreas fault by creating two hybrid models. These models consist of basin-scale (top 10 km or less) high-resolution models developed using controlled source data embedded into two popular Southern California Earthquake Center community models. We evaluate the models by computing long period (6–30 s) wavefield energy misfits using 11 earthquakes with moment magnitudes between 3.5 and 5.5 not used in developing any of the models under consideration. One of the hybrid models produces an ~24% decrease while the other has an ~0.6% increase in the overall median misfit relative to their original community models. The overlapping misfit values between the models and variability in waveform fit for different events and stations emphasize the difficulties in model validation. Our approach can merge any type of gridded multiscale and multidimensional datasets, and represents a valuable tool for modeling in the computational sciences.

Plain Language Summary Earth models are helpful to society and play an essential role in exploring for natural resources, geologic hazard assessments, and understanding how our planet works. The models are developed in various scales ranging from the entire Earth to a metropolitan area. It tends to be the case that the bigger models are more expensive to create, especially with cutting-edge methods. We experiment with a simple technique for updating big models by replacing the most critical parts with smaller, more accurate models. When we test our new tool for earthquake ground motion prediction, we achieve some results that show that it can be useful.

1. Introduction

Physics-based probabilistic seismic hazard analysis (PSHA) depends on a good understanding of earthquake processes that are described using rupture models and accurate structural models of the earth, used in 3D deterministic wave propagation simulations to predict ground motions (Milner et al., 2021; B. E. Shaw et al., 2018). The hazard information is utilized in tandem with engineering data to produce risk assessments that form an integral part of building codes and can help preserve the structural integrity of existing infrastructure in the occurrence of a large earthquake (Maechling et al., 2007). In general, areas with low seismic velocities, such as sedimentary basins, are prone to more vigorous and prolonged shaking (Bijelic et al., 2019; Brissaud et al., 2020; Graves et al., 1998; Lovely et al., 2006) and are thus sites where improvements in the shallow parts of structural models can be most beneficial. According to the latest Uniform California Earthquake Rupture Forecast (UCERF3) model of California, the southern San Andreas fault poses the greatest threat in southern California, with the highest probability of a large-magnitude event in the next few decades (Field et al., 2013). Several scenario simulations of the San Andreas fault rupture show significant amplification in the sedimentary basins in the nearby Salton Trough and the densely populated Los Angeles region located further north along the fault (Day et al., 2012; Jones et al., 2008; Porter et al., 2011). Recent modeling also suggests additional complexity through the San Andreas fault's potential simultaneous rupture with the nearby San Jacinto fault (Lozos, 2016) (Figure 1). To be better prepared,

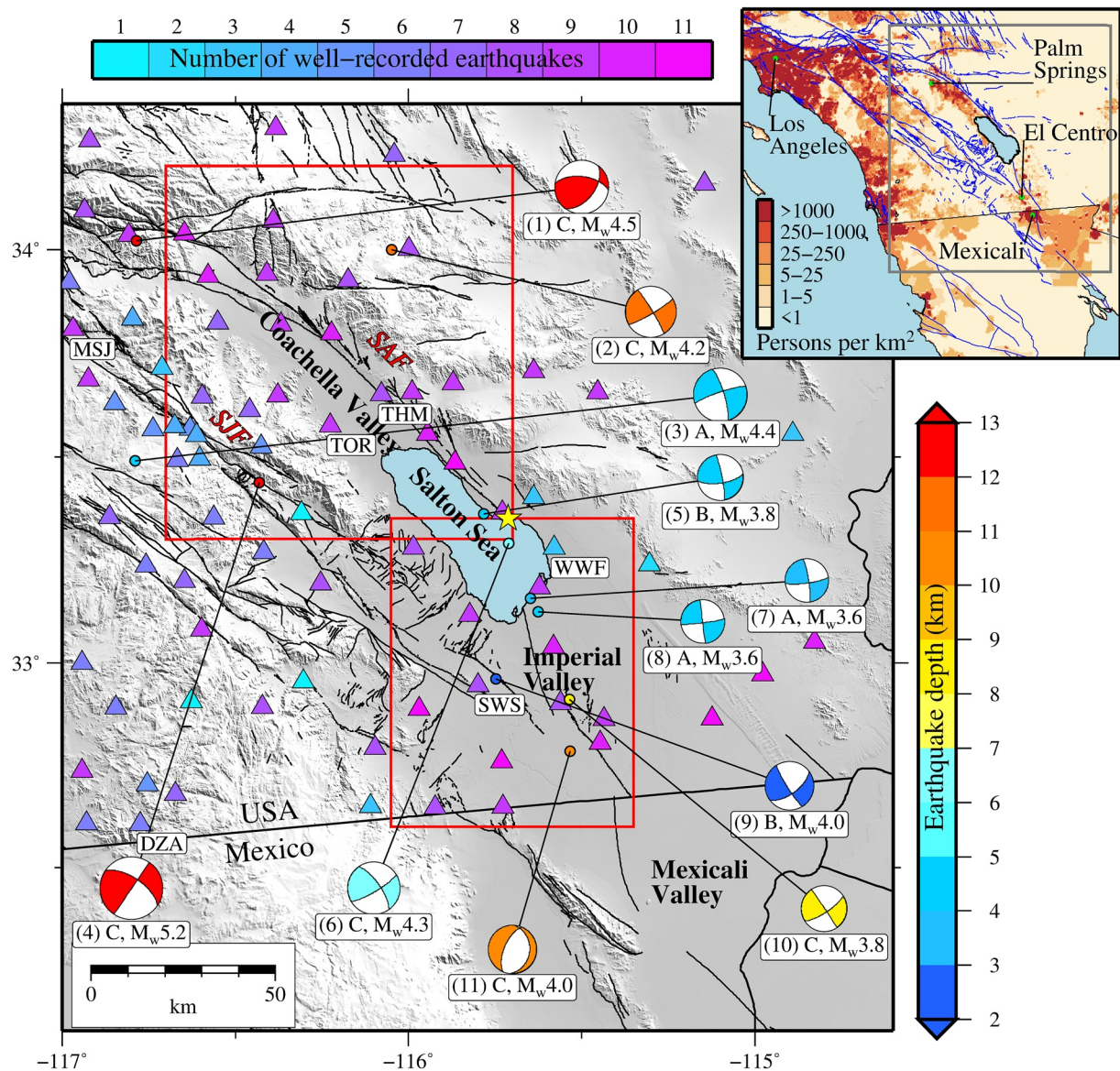


Figure 1. Earthquakes and seismic stations in Salton Trough. Wavefield simulation domain of the validation exercise near the southernmost segment of the San Andreas fault (SAF) and the San Jacinto fault. The circles are locations of the 11 selected earthquakes color-coded by depth and accompanied by focal mechanisms with labels that indicate the focal mechanism quality and the event magnitude. The yellow star indicates the hypothetical rupture site of the SAF (Jones et al., 2008). The triangles are broadband stations color-coded by the number of the selected earthquakes that were well-recorded; stations discussed in the text are labeled. Red polygons are the horizontal bounding areas of local velocity models developed in Coachella and Imperial valleys. Thin black lines are surface traces of mapped faults (Jennings & Bryant, 2010). Inset map shows the population density in the surrounding area (Center for International Earth Science Information Network-CIESIN-Columbia University, 2018). Table S1 shows the earthquake source parameters and Figures S1 and S2 show additional data quality statistics.

Southern California Earthquake Center (SCEC) researchers have developed community Earth models for southern California and computational infrastructure such as the CyberShake platform (Graves et al., 2010; Maechling et al., 2007) to create statewide hazard maps. Descriptions of the Earth models are given in terms of compressional (P) wave velocity, shear (S) wave velocity, and the density of the crust and upper mantle and may also include attenuation and anisotropy. These parameters are essential for determining variations in the intensity of ground shaking. The models are also used in geologic studies, unraveling tectonic history, and fault system modeling. They can help refine the geometry of fault models used in earthquake rupture forecast and seismicity studies to identify new faults and improve earthquake source characterization. Therefore, it is vital to have high-resolution basin-scale models that represent the geology in enough detail

for various scientific studies and engineering applications integrated into these community models. Here, we present a method for seamlessly merging such models.

1.1. Community Earth Models

During the last few years, SCEC has devoted significant effort to improving its community models, particularly the community velocity models (CVM), integral to ground motion prediction studies in southern California. The two popular and latest community velocity models commonly used in producing seismic hazard estimates are CVM-H 15.1 and CVM-S 4.26. These models have a rich historical development and are constructed using different geologic datasets and seismic tomographic techniques. CVM-H 15.1 and CVM-S 4.26 are developed from their immediate predecessors using full-waveform inversion (FWI) techniques. CVM-H 15.1 involved 16 iterations of the adjoint-wavefield FWI method (Tape et al., 2009, 2010), while CVM-S 4.26 was developed using 26 iterations of a combination of the adjoint-wavefield and scattering integral formulations (Lee, Chen, Jordan, et al., 2014). The models primarily provide *P* and *S* wave velocities and density for the crust and upper mantle. The models can also be modified to include geotechnical layers that provide more accurate descriptions of the near-surface, usually the top few hundred meters, that affect ground shaking estimates. Validation studies that interrogate the models' ground motion prediction ability have shown that the community models fit low-frequency seismograms (<0.5 Hz) reasonably well, with CVM-S 4.26 outperforming CVM-H 15.1 in most cases and areas in southern California (Lee, Chen, & Jordan, 2014; Taborda et al., 2016). However, for higher frequencies exceeding 1 Hz that are most important to structural engineers, the community models have weak predictions (Taborda & Bielak, 2014). These results are not surprising since the community velocity models were created using low-frequency seismograms with a maximum of 0.5 Hz for CVM-H 15.1 and 0.2 Hz for CVM-S 4.26. CVM-H 15.1 uses earthquake waveforms filtered in three period-bands: 6–30 s, 3–30 s, and 2–30 s, with about half of the misfit measurements made on 6–30 s windows, while CVM-S 4.26 uses both earthquake seismograms and ambient-noise correlograms filtered in the 5–30 s period. However, due to the computational demands of full-waveform inversion, incorporating higher frequency content into the community models on a regional scale poses a challenge.

The cost of FWI depends on the simulation domain size, the length of the simulated wavefield, and the maximum frequency of the data. The maximum frequency that can be resolved, in turn, depends on the minimum velocity in the model, the minimum spacing between grid points in the model mesh, and the time step of the simulation. For example, the CVM-H 15.1 model has a dimension of 639 km by 503 km in the horizontal plane and a vertical thickness of 60 km. For each iteration in the inversion, three 120–300 s simulations for every one of the 143 earthquake sources are performed, resulting in a total of 0.8 million CPU-hours (Tape et al., 2010). Similar computational resources were required for a recent validation study of the CVM-H 15.1 model in Los Angeles involving a domain size of 180 km by 135 km by 61.875 km and a single 100 s simulation of the Chino Hills earthquake accurate to a maximum frequency of 4 Hz (Taborda & Bielak, 2014). Without regard to the significant storage requirements that can pose data management challenges, this shows how quickly high-frequency FWI on regional scales becomes impractical.

1.2. Salton Trough Active-Source Local Models

In 2011, the Salton Seismic Imaging Project (SSIP) survey comprising seven wide-angle reflection and refraction profiles and a grid of recording stations in the Salton Trough was completed (Rose et al., 2013). The goal of SSIP was to understand the nature of the crust in the region and image the thick sedimentary basins and faults, especially blind faults, that pose a significant earthquake hazard. In particular, Ajala, Persaud, Stock, et al. (2019) and Persaud et al. (2016) used a subset of SSIP data and local earthquakes to develop 3D travel time velocity models of Coachella Valley and Imperial Valley, respectively (red polygons in Figure 1) based on ray theory (Hole, 1992). The 3D models are reasonably resolved to a maximum depth of 10 km in the Coachella Valley and 8 km in the Imperial Valley. Persaud et al. (2016) also use active source data from the 1979 USGS seismic refraction experiment in Imperial Valley (Fuis et al., 1982). Although the models provide only *P* wave velocities, the controlled-source experiments have the advantage of the lack of source location errors and provide better constraints on basin structures than models developed using only

earthquake data. Compared to the regional FWI community models, these models were computationally relatively inexpensive to create, requiring a few hours on a decent workstation.

1.3. Multiscale Earth Models

1.3.1. Model Merging

Besides the 3D SSIP models and CVM-H 15.1 and CVM-S 4.26 discussed in the previous sections, several regional and local models have been developed by researchers throughout California. Basin-scale models that provide better characterization of sedimentary structures have also been inserted into regional models to improve them. For example, an earlier version of CVM-S 4.26 embedded detailed rule-based basin models in irregular domains (Magistrale et al., 2000) into a regional travel time tomographic model (Hauksson, 2000). The predecessor of CVM-H 15.1 also embedded high-resolution basin models (Komatitsch et al., 2004; J. H. Shaw et al., 2015; Suss & Shaw, 2003) into the same background model of Hauksson (2000). Komatitsch et al. (2004) found that the detailed basin models embedded in the Los Angeles area predict the long-period (2–35 s) ground motion and significant amplification in the region. In northern California, the USGS Bay Area Velocity Model 05.1.0 was developed by embedding a detailed local model into a regional model (Brocher, 2008) and is validated by several studies (Aagaard et al., 2008; Kim et al., 2010; Rodgers et al., 2008) to be a good model. The drawback of these models is that they were developed using complex data structures or commercial packages, lacked flexibility due to the hard-coded process of embedding the models, and often include sharp unrealistic boundaries between the embedded and background models. Outside California, studies like Molinari and Morelli (2011) merge several models of varying spatial scales to develop EPCrust, a crustal model of the European Plate. They use cosine taper weights to smoothen the transition between the different models defined in rectangular domains.

1.3.2. Data Assimilation

More recent advances in the multiscale inversion of geophysical datasets introduced by Fichtner et al. (2018) allow for updating global Earth models on any scale commensurate with the available data. In this case, one can directly use dense seismic datasets to refine the regional models through a Bayesian inversion framework that ensures consistent model updates. The approach has been used to develop the Collaborative Seismic Earth Model Generation 1 using twelve regional-scale FWI refinements.

1.4. Unified Community Velocity Model Software

To easily access multiple models with different resolutions, scales, and coverage in a reproducible fashion, SCEC developed the Unified Community Velocity Model (UCVM) software (Small et al., 2017). UCVM is a collection of tools that allows researchers to interrogate Earth models. Based on a tiling concept, all models registered into the UCVM program can be retrieved on a point-by-point basis using an ordered list to return model parameters in the order of availability. So, statewide Earth models can be readily constructed by using a list of several community models that overlap and span different areas in the order of increasing resolution. UCVM provides additional programs that can modify the community models, such as adding built-in high-resolution topography, shallow geotechnical layer information, and small-scale heterogeneities that mimic high-frequency structural content. UCVM has mesh generation tools that produce finite-difference grids and etree databases for the models, which can be used directly with open-source seismic wave propagation solvers (Olsen et al., 1995; Tu et al., 2006). UCVM easily retrieves depths to specific model parameters such as basement surfaces, Moho surfaces, or Z2.5 and Z1.0 surfaces commonly used by structural engineers in modeling basin amplification effects. As a result, UCVM facilitates comparative validation and PSHA studies.

One shortcoming of UCVM is that different models that are registered can have disagreements at their boundaries. So, when a large-scale hybrid model comprising several kinds of models is constructed, unrealistic sharp boundaries are present in the model (Callaghan et al., 2017). If these transitions are not properly smoothed, they can introduce spurious seismic reflections or refractions that imitate compositional or fault boundaries, especially if the model differences are step-like (Ajala, Persaud, Juarez, et al., 2019). These artifacts can introduce errors into wavefield simulations and can therefore affect the validation and PSHA

results. Another issue not addressed by UCVN is that the model parameters are usually not determined with the same resolution at all grid points. For most tomographic imaging studies, the model resolution deteriorates near the boundaries where the source-receiver path sensitivity density is relatively low. However, since most models are developed and distributed on regular grids, the entire cuboid or rectangular domain includes the low-resolution areas. Researchers often use these models in their entirety without regard to the high-resolution coverage volume, which can lead to less accurate wavefield predictions due to the poorly resolved areas.

1.5. Moving Forward

We develop an algorithm for defining window functions in arbitrary shapes and argue that it offers a solution to the SCEC UCVN program's current limitations and can be used to update regional community models. The properties of window functions (Doerry, 2017) allow our technique to merge multiscale models defined on regular grids smoothly. With the advent of urban seismology, short-term dense seismic deployments, also known as large-N surveys, are becoming increasingly popular (Clayton et al., 2019; Lin et al., 2013; Liu et al., 2018). These surveys can involve thousands of seismometers and are usually deployed for ~1 month, the battery life of nodal seismometers typically used in the surveys, with spacings of a few hundred meters or less between each seismometer. These new surveys allow for the rapid development of localized high-resolution Earth models. Existing fiber optic networks can also complement these surveys by providing additional data with ultradense spatiotemporal resolution (Parker et al., 2018).

One strategy for updating the regional community models is to identify high seismic hazard areas such as, high-density population areas that sit atop sedimentary basins which amplify ground motion, and areas with expensive infrastructure and develop more accurate models using novel seismic arrays. The widespread use of dense seismic arrays has also triggered the development of array-based seismic imaging techniques to produce better subsurface models (Bianco et al., 2019; Castellanos et al., 2020; Zhong & Zhan, 2020). Since local models are small, they can be developed using high-frequency data in a more computationally tractable manner relative to a larger scale model. Once the local models are developed, they can be smoothly embedded into the regional community models to add more detail. To resolve some of the perceived limitations with UCVN, we developed a library based on our algorithm and implemented it to UCVN. Polygons that define the high-resolution coverage of the local models can be specified. Also, any window function can be used to extract the local model volume so that only well-resolved model parameters are returned and help smooth out the transition between the boundaries of different models during the construction of hybrid Earth models. We present a case study in the Salton Trough (Figure 1) where we smoothly embed the local SSIP models in Coachella and Imperial valleys into the regional community models to develop two hybrid models, one of which produces a ~24% lower waveform misfit compared to its original community model. We further examine the misfit between observed and synthetic waveforms in different parts of the model domain and for different events to show the complexities of model validation.

2. Materials and Methods

2.1. Data Selection and Processing

Events are selected from the Southern California Earthquake Data Center (SCEDC) focal mechanism catalog (Hauksson et al., 2012; Yang et al., 2012). Our selection criteria include events occurring after 2015 within our simulation domain, so they postdate the development of the community models and any dataset used in developing the local models in Coachella and Imperial valleys. We use a magnitude range of 3.5–5.5 that can be accurately represented by a moment tensor point source. We then use the event IDs to download the waveforms available at the Southern California Seismic Network stations using the Seismogram Transfer Program (STP) from SCEDC. The final set of events used in our validation exercise were selected with the requirement that they sample different regions of our simulation domain and are recorded by a large number of stations. All records are preprocessed by removing the mean, trend, and instrument response, filtered between 6 and 30 s and converted to displacement seismograms. We then select the high-quality preprocessed records using two definitions of the signal-to-noise ratio (SNR) based on the energy (L^2 norm) and maximum amplitude (L^∞ norm) given by

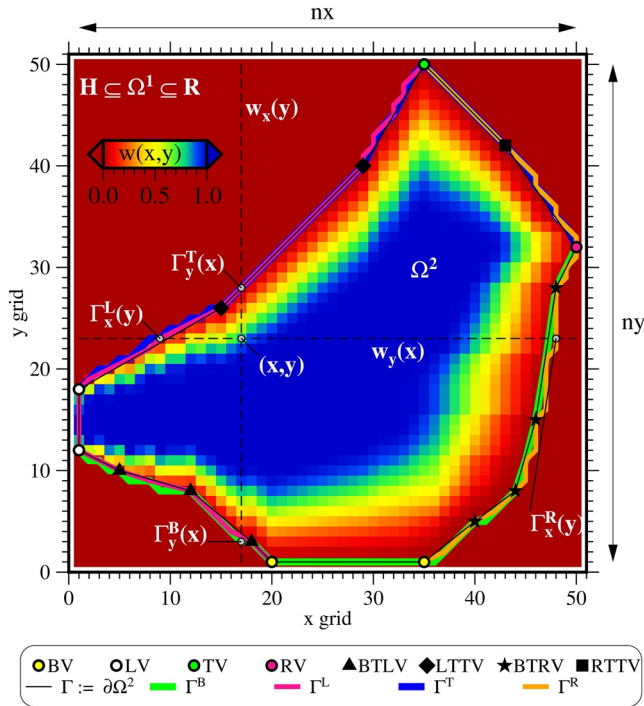


Figure 2. Schematic illustration of the blending algorithm for a 2D case. The required inputs to generate the blending weights $w(x, y)$ include vertices (symbols) that describe the boundary (solid lines) of the local model's high-resolution coverage area Ω^2 and the window functions to be used along each spatial dimension (see Model Merging Using Window Functions in Materials and Methods).

$$\text{SNR}_{L^2} = \frac{\frac{1}{t_e - t_p} \|d(x_r, t) \chi_{[t_p, t_e]}\|_2^2}{\frac{1}{t_p - t_b - 10} \|d(x_r, t) \chi_{[t_b, t_p - 10]}\|_2^2}, \quad (1)$$

$$\text{SNR}_{L^\infty} = \frac{\|d(x_r, t) \chi_{[t_p, t_e]}\|_\infty}{\|d(x_r, t) \chi_{[t_b, t_p - 10]}\|_\infty}, \quad (2)$$

where $d(x_r, t)$ is the preprocessed observed displacement seismogram of an arbitrary component at a station located at x_r . $\chi_{[t_p, t_e]}$ and $\chi_{[t_b, t_p - 10]}$ are indicator functions that divide the record into signal and noise sections. The signal window runs from the first P wave arrival time t_p to the end of the STP triggered window t_e corresponding to the surface wave disappearance. For the noise section, we use the beginning time of the record t_b , usually several seconds before the event's origin time, to 10 s before the first P wave arrival $t_p - 10$. Waveforms recorded at a station are selected if the two SNR measures are greater than or equal to 3 on all components.

2.2. Model Merging Using Window Functions

Given a local model defined over a rectangular domain Ω^1 that is only well-resolved in an irregular domain Ω^2 bounded by Ω^1 , we would like to create a hybrid model in Ω^1 by merging the local model in Ω^2 with a regional model that is well-defined over Ω^1 (Figure 2). The support (supp) of a real-valued function f in a domain Ω refers to the domain subset that is mapped to non-zero values, that is

$$\text{supp}(f) = \{x \in \Omega \mid f(x) \neq 0\}. \quad (3)$$

On the complement of the support, f vanishes. We construct a window function w in Ω^1 such that the support (supp) is approximately the closure

of Ω^2 , i.e., $\text{supp}(w) \cong \overline{\Omega^2} = \Omega^2 \cup \partial\Omega^2$. Then the hybrid model $H(x)$ at any given location x in Ω^1 is developed by simply merging the local model $\Omega^1(x)$ with the regional model $R(x)$ using the blending weight $w(x)$ in a linear interpolation as follows

$$H(x) = \Omega^1(x)w(x) + R(x)(1 - w(x)), \quad (4)$$

with $w(x)$ satisfying

$$w(x) = \begin{cases} 0, & x \in \Omega^1 \setminus \overline{\Omega^2} \\ (0, 1], & x \in \overline{\Omega^2} \end{cases}. \quad (5)$$

Following the architecture of the UCVm program for querying model parameters, we present a simple algorithm for generating these blending weights on a point-by-point basis for a 2D case. Figure 2 shows the blending weight map in grid coordinates, where we utilize projection routines within UCVm for arbitrary projections. The restriction to the technique is that the polygon formed by the boundary Γ must be xy-monotone, so all lines parallel to the x and y axis must intersect Γ at most twice. Extension to higher dimensions is straightforward and is omitted for brevity. Given a set of vertices $V = \{(x_i, y_i) \mid 1 \leq i \leq N\}$ describing the boundary of Ω^2 , the first step of our algorithm is to classify the vertex sets based on their relative location into bottom vertices (BV), left vertices (LV), top vertices (TV), right vertices (RV), bottom-to-left vertices (BTLV), left-to-top vertices (LTTV), bottom-to-right vertices (BTRV), and right-to-top vertices (RTTV) as follows:

$$\text{BV} = \{(x_i, y_i) \in V \mid (\forall i) y_i = V_{\min y}\} \setminus \{1 \leq |BV| \leq 2\}, \quad (6)$$

$$\text{LV} = \{(x_i, y_i) \in V \mid (\forall i) x_i = V_{\min x}\} \setminus \{1 \leq |LV| \leq 2\}, \quad (7)$$

$$TV = \left\{ (x_i, y_i) \in V \mid (\forall i) y_i = V_{\max y} \right\} (1 \leq |TV| \leq 2), \quad (8)$$

$$RV = \left\{ (x_i, y_i) \in V \mid (\forall i) x_i = V_{\max x} \right\} (1 \leq |RV| \leq 2), \quad (9)$$

$$BTLV = \left\{ (x_i, y_i) \in V \mid (\forall i) BV_{\min x} < x_i < LV_{\forall x}, BV_{\forall y} < y_i < LV_{\min y} \right\}, \quad (10)$$

$$LTTV = \left\{ (x_i, y_i) \in V \mid (\forall i) LV_{\forall x} < x_i < TV_{\min x}, LV_{\max y} < y_i < TV_{\forall y} \right\}, \quad (11)$$

$$BTRV = \left\{ (x_i, y_i) \in V \mid (\forall i) BV_{\max x} < x_i < RV_{\forall x}, BV_{\forall y} < y_i < RV_{\min y} \right\}, \quad (12)$$

$$RTTV = \left\{ (x_i, y_i) \in V \mid (\forall i) TV_{\max x} < x_i < RV_{\forall x}, RV_{\max y} < y_i < TV_{\forall y} \right\}, \quad (13)$$

where V_x and V_y represent the set containing only the x coordinates and y coordinates of V , respectively. Of the eight vertex sets, BV, LV, TV, and RV, are required and can share the same elements in the case of a corner vertex. For the remaining vertex sets, $i \in I$ for $I = \{1, \dots, n\}$ where n can vary between the sets and should be sufficiently large to capture the geometry of the high-resolution polygon. Next, we generate sequences that approximate the boundary Γ at the left (Γ^L), right (Γ^R), bottom (Γ^B), and top (Γ^T). These are assembled as

$$\Gamma_x^L = \left[\bigcup_{k=1}^{|L|-1} \left\{ (n-1)\Delta_k + L_{x_k} \right\} \cup \left\{ L_{x_{ny}} \right\} \right] \left[n = 1, \dots, L_{y_{k+1}} - L_{y_k}, \Delta_k = \frac{L_{x_k} - L_{x_{k+1}}}{L_{y_k} - L_{y_{k+1}}} \right], \quad (14)$$

$$\Gamma_x^R = \left[\bigcup_{k=1}^{|R|-1} \left\{ (n-1)\Delta_k + R_{x_k} \right\} \cup \left\{ R_{x_{ny}} \right\} \right] \left[n = 1, \dots, R_{y_{k+1}} - R_{y_k}, \Delta_k = \frac{R_{x_k} - R_{x_{k+1}}}{R_{y_k} - R_{y_{k+1}}} \right], \quad (15)$$

$$\Gamma_y^B = \left[\bigcup_{k=1}^{|B|-1} \left\{ (n-1)\Delta_k + B_{y_k} \right\} \cup \left\{ B_{y_{nx}} \right\} \right] \left[n = 1, \dots, B_{x_{k+1}} - B_{x_k}, \Delta_k = \frac{B_{y_k} - B_{y_{k+1}}}{B_{x_k} - B_{x_{k+1}}} \right], \quad (16)$$

$$\Gamma_y^T = \left[\bigcup_{k=1}^{|T|-1} \left\{ (n-1)\Delta_k + T_{y_k} \right\} \cup \left\{ T_{y_{nx}} \right\} \right] \left[n = 1, \dots, T_{x_{k+1}} - T_{x_k}, \Delta_k = \frac{T_{y_k} - T_{y_{k+1}}}{T_{x_k} - T_{x_{k+1}}} \right], \quad (17)$$

where

$$B = LV_1^{\rightarrow y} \cup BTLV^{\leftarrow y} \cup BV^{\rightarrow x} \cup BTRV^{\rightarrow y} \cup RV_1^{\rightarrow y}, \quad (18)$$

$$L = BV_1^{\rightarrow x} \cup BTLV^{\rightarrow y} \cup LV^{\rightarrow y} \cup LTTV^{\rightarrow y} \cup TV_1^{\rightarrow x}, \quad (19)$$

$$T = LV_1^{\leftarrow y} \cup LTTV^{\rightarrow x} \cup TV^{\rightarrow x} \cup RTTV^{\rightarrow x} \cup RV_1^{\leftarrow y}, \quad (20)$$

$$R = BV_1^{\leftarrow x} \cup BTRV^{\rightarrow y} \cup RV^{\rightarrow y} \cup RTTV^{\rightarrow y} \cup TV_1^{\leftarrow x}. \quad (21)$$

The arrows indicate sorting operations on a vertex set B such that $B^{\rightarrow x} := \left\{ (x_i, y_i) \mid (\forall i) x_i < x_{i+1} \right\}$, $B^{\leftarrow x} := \left\{ (x_i, y_i) \mid (\forall i) x_i > x_{i+1} \right\}$, $B^{\rightarrow y} := \left\{ (x_i, y_i) \mid (\forall i) y_i < y_{i+1} \right\}$, and $B^{\leftarrow y} := \left\{ (x_i, y_i) \mid (\forall i) y_i > y_{i+1} \right\}$. $|B|$ refers to the cardinality of the set B . For any query location (x, y) in Ω^1 , the blending weight is returned as the product of the window functions in the x and y dimensions

$$w(x, y) = w_x(y)w_y(x). \quad (22)$$

Assuming $w(x, y)$ is a cosine taper window function, we can define

$$w_x(y) = \begin{cases} 0, & y < \Gamma_y^B(x) \\ \frac{1}{2} \left(1 - \cos \left(\pi \frac{2\bar{x} - 1}{2n_i} \right) \right), & \Gamma_y^B(x) \leq y \leq \Gamma_y^B(x) + n_i \\ 1, & \Gamma_y^B(x) + n_i < y \leq \Gamma_y^T(x) - n_f \\ \frac{1}{2} \left(1 - \cos \left(\pi \frac{2(\bar{x} - n + 2n_f) - 1}{2n_f} \right) \right), & \Gamma_y^T(x) - n_f < y \leq \Gamma_y^T(x) \\ 0, & y > \Gamma_y^T(x) \end{cases} \quad (23)$$

where $n = \Gamma_y^T(x) - \Gamma_y^B(x) + 1$ and $\bar{x} = x - \Gamma_y^B(x) + 1$,

$$w_y(x) = \begin{cases} 0, & x < \Gamma_x^L(y) \\ \frac{1}{2} \left(1 - \cos \left(\pi \frac{2\bar{y} - 1}{2n_i} \right) \right), & \Gamma_x^L(y) \leq x \leq \Gamma_x^L(y) + n_i \\ 1, & \Gamma_x^L(y) + n_i < x \leq \Gamma_x^R(y) - n_f \\ \frac{1}{2} \left(1 - \cos \left(\pi \frac{2(\bar{y} - n + 2n_f) - 1}{2n_f} \right) \right), & \Gamma_x^R(y) - n_f < x \leq \Gamma_x^R(y) \\ 0, & x > \Gamma_x^R(y) \end{cases} \quad (24)$$

where $n = \Gamma_x^R(y) - \Gamma_x^L(y) + 1$ and $\bar{y} = y - \Gamma_x^L(y) + 1$. n_i and n_f are the lengths of the taper sections for the beginning and end of the window functions and are defined as $n_i = r_1 n$ and $n_f = r_2 n$, for taper ratios r_1 and r_2 in the half-open interval $[0, 0.5)$. The window function can be modified to have restricted tapering as in the westerly boundary segment of Figure 2 or a one-sided taper like the vertical dimension of the profiles shown in Figure 3, with tapering only at the bottom. While we could specify a distance rather than a taper ratio, the size of the tapered section along the support boundary cannot be equal when the support is irregular due to restrictions imposed by the regular grid. Our current implementation to UCVM allows a single taper ratio value to ensure a symmetric taper for the window functions along each dimension and simplicity (Ajala, 2021c). Other window functions like the trapezoid (linear) function in Figure S3 can be constructed and implemented similarly. Doerry (2017) documents 50 window taper functions. We note that the elements of the sets and sequences belong to the integers so that for any $\mathbf{x} \in \mathbb{R} \setminus \mathbb{Z}$, we take advantage of linear interpolation routines in UCVM. For arbitrary nonmonotone boundaries, the new problem becomes finding the closest monotone polygon to utilize the algorithm described herein.

2.3. Spectral-Element Solution to the Wave Equation

The complete seismic wavefield is computed using the spectral-element method via the SPEC-FEM3D package (Komatitsch & Vilotte, 1998). We solve the equation of motion in anelastic isotropic media. The strong formulation S of our problem is

$$(S) \left\{ \begin{array}{ll} \rho(\mathbf{x}) \frac{\partial^2 u(\mathbf{x}, t)}{\partial t^2} = \nabla \cdot \boldsymbol{\sigma}(\mathbf{x}, t) + f(\mathbf{x}, t), & \text{on } \Omega \times I \\ u(\mathbf{x}, t) = g(\mathbf{x}, t), & \text{on } \Gamma_g^{\text{int}} \times I \\ \boldsymbol{\tau}(\mathbf{x}, t) = \boldsymbol{\sigma}(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}), & \text{on } \Gamma_\tau^{\text{int}} \times I, \\ \boldsymbol{\tau}(\mathbf{x}, t) = 0, & \text{on } \Gamma_{\text{free}}^{\text{ext}} \times I \\ \boldsymbol{\tau}(\mathbf{x}, t) = c_p(\mathbf{x}) \rho(\mathbf{x}) v_n(\mathbf{x}, t) + c_x(\mathbf{x}) \rho(\mathbf{x}) v_t(\mathbf{x}, t), & \text{on } \Gamma_{\text{abs}}^{\text{ext}} \times I \\ u(\mathbf{x}, 0) = 0, v(\mathbf{x}, 0) = 0, & \mathbf{x} \in \Omega \end{array} \right. \quad (25)$$

where $f(\mathbf{x}, t) = -M \cdot \nabla \delta(\mathbf{x} - \mathbf{x}_s) S(t)$ is the point moment tensor source initiated at the earthquake hypocenter $\mathbf{x}_s$. The source time function $S(t)$ is a Gaussian function with a width equal to the half duration of the earthquakes obtained by scaling the scalar moment using an empirical relationship (Ekstrom et al., 2012). Our simulation domain Ω is 220 km by 252 km and extends from the free surface down to 50 km below sea

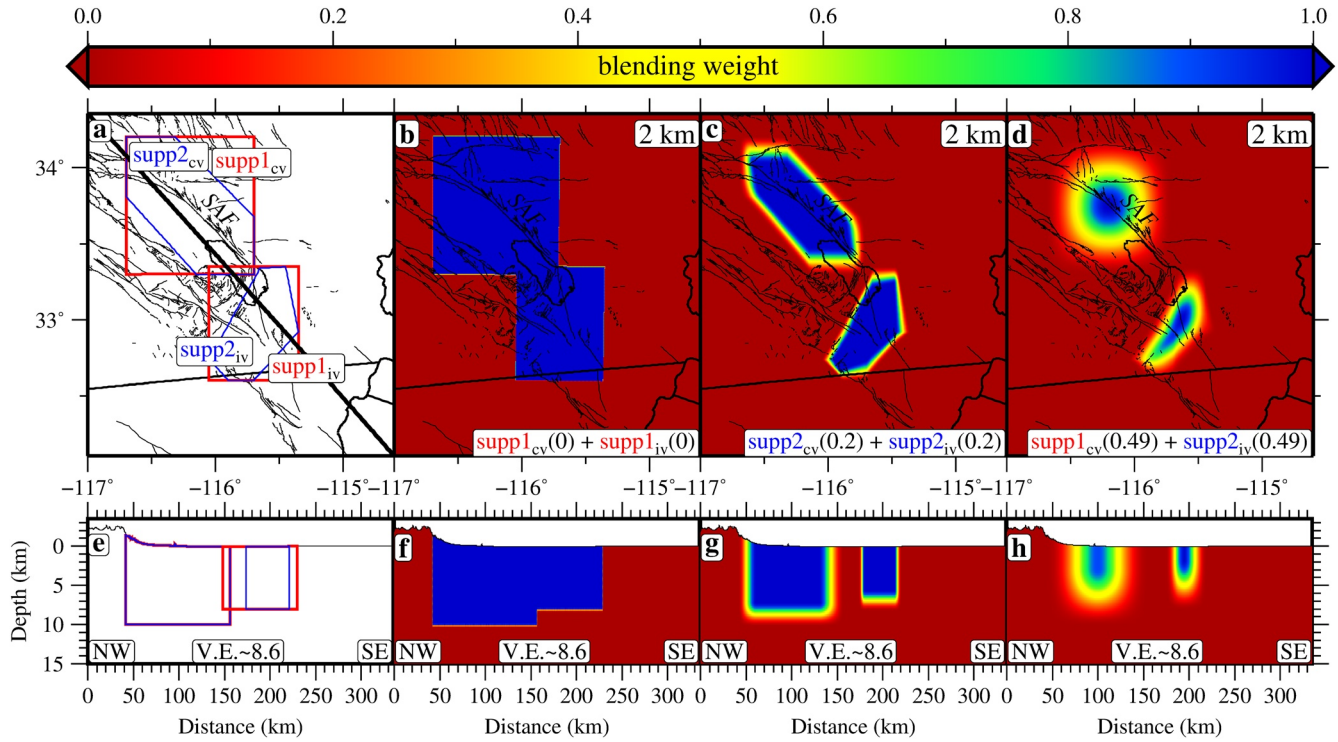


Figure 3. Cosine taper blending weights used to generate hybrid models. (a) Map of Salton Trough showing the areal extents of the local models to be embedded into the regional community models. The polygons indicate the support (supp) of the cosine taper function. Red polygons (supp1) represent the bounding area of the local models, while the blue polygons (supp2) show the subset where the local model parameters are well resolved. The subscripts of the support indicate the domains in Coachella Valley (cv) and Imperial Valley (iv). The black thick line shows the location of the cross-sectional profiles shown in e-h. (b) Map of blending weights from using supp1 with zero taper ratio at 2 km depth below sea level. (c) Map of blending weights from using supp2 with 0.2 taper ratio. (d) Map of blending weights from using supp1 for the cv model and supp2 for the iv model with 0.49 taper ratio. (e) Cross-section view showing the extents of the local models. (f) Cross-section showing blending weights with zero taper ratio. (g) Cross-section showing blending weights with 0.2 taper ratio. (h) Cross-section showing blending weights with 0.49 taper ratio.

level. The length of the simulation time I is 200 s. Displacement $u(\mathbf{x}, t)$ and traction $\tau(\mathbf{x}, t)$ are continuous on internal boundaries Γ_g^{int} and Γ_τ^{int} , respectively. On the free surface $\Gamma_{\text{free}}^{\text{ext}}$, traction vanishes due to the free surface boundary condition. On the artificial boundaries of the model domain $\Gamma_{\text{abs}}^{\text{ext}}$, which includes all external boundaries except the free surface, we enforce the paraxial absorbing boundary conditions (Clayton & Engquist, 1977; Stacey, 1988). For our initial conditions, displacement and velocity are zero at start time ($t = 0$). We use three standard linear solids (Zener models) to simulate a constant quality factor (Q) approximation valid in our frequency band of interest (Savage et al., 2010). The resulting semi-discrete form of S following conversion to the weak formulation and approximation using the Galerkin discretization is advanced forward in time using the Newmark time scheme in a predictor-multicorrector format (Hughes, 1987). The memory variable equations that describe attenuation is separately marched in time using a second order Runge-Kutta method (Komatitsch & Tromp, 2002).

2.4. Misfit Assessment

We use the relative waveform misfit (RWM) that measures the distance in L^2 norm between the processed observed and predicted displacement records, which is the energy difference in the waveforms that implicitly takes into account both phase and amplitude misfits. We compute

$$\text{RWM} = \frac{\|d(\mathbf{x}_r, t) - s(\mathbf{x}_r, t; \mathbf{m})\|_{\chi_{[t_p, t_e]}}^2}{\|d(\mathbf{x}_r, t)\|_{\chi_{[t_p, t_e]}}^2 + \|s(\mathbf{x}_r, t; \mathbf{m})\|_{\chi_{[t_p, t_e]}}^2}, \quad (26)$$

from the P arrival time t_p to the end of the surface waves t_e , where $d(\mathbf{x}_r, t)$ is the processed observed seismogram and $s(\mathbf{x}_r, t; \mathbf{m})$ is the processed synthetic seismogram generated by the Earth model $\mathbf{m}$ at the receiver location $\mathbf{x}_r$. RWM is extensively used in validation studies and for assessing model improvements during waveform tomographic inversion updates (Lee, Chen, & Jordan, 2014; Tape et al., 2010).

3. Results

3.1. Validation Data

There is abundant seismicity across the numerous fault zones in southern California. For our validation exercise in Salton Trough, we consider 11 local events and number the events from 1 through 11 with decreasing hypocentral latitude. The events have an excellent spatial distribution in the Imperial and Coachella valleys and the mountain ranges and have a depth range of 2.2–13.5 km. The majority of the 79 broadband stations used in the study are from the Caltech Regional Seismic Network (CI) network. The remaining stations are from the ANZA regional (AZ) network, with all of the stations located in the U.S. For all earthquakes, the maximum and minimum number of stations with recordings are 69 and 74, respectively. However, after using a signal-to-noise ratio (S/N) criterion to filter out stations with noisy records, we use a maximum of 68 stations corresponding to events 6 and 9 and a minimum of 23 stations for event 8 (Figures S1 and S2). Our simulation domain is roughly centered around the hypothetical future rupture site (Jones et al., 2008) of the San Andreas fault indicated by the yellow star in Figure 1, with two of the selected events, 5 and 6, close to the predicted location. All three-component waveforms from the events are oriented in the vertical, north-south, and east-west directions. We use a period range of 6–30 s because it is well-represented in the community models which were developed using waveforms dominant in that band. We do not perform any focal mechanism inversions before validation. CVM-H 15.1 and CVM-S 4.26 models are available everywhere in our selected region. For the hybrid models, we smoothly embed the SSIP local models into the community models.

3.2. Blending Map

Our blending algorithm was implemented as a library using the C programming language (Ajala, 2021a). We then updated the UCVM Application Programming Interface to utilize our library (Ajala, 2021c). Following our modification to UCVM, blending weights and blended model parameters can be returned along with unmodified model parameters at a specified query location. We use UCVM to merge the local models with the community models CVM-H 15.1 and CVM-S 4.26 to make hybrid models for our validation exercise. Since the local model parameters are stored on a regular grid, the entire model volume has a cuboid shape (supp1), which can be modified to an arbitrary shape (supp2), indicating a subset of the local model that is better resolved (Figure 3). The high-resolution polygon (supp2) coincides with the dense ray coverage region in the local models, defined as areas with >50 rays within each grid cell (Persaud et al., 2016). To illustrate the range of possibilities when blending the different models and to also show that a hybrid model can be produced from CVM-H 15.1 that outperforms CVM-S 4.26, we use blending maps derived with different support and window function parameters. Figures 3b and 3f are derived using supp1 in both local model domains and a boxcar function, representing the current operation of the official UCVM with sharp transitions at the model boundaries. Our validation exercise uses the high-resolution polygons (supp2) with a taper ratio of 0.2 when merging the local models with CVM-H 15.1 (Figures 3c and 3g). For the CVM-S 4.26 hybrid model, we use the full model volume (supp1) for the Coachella Valley model, the high-resolution coverage (supp2) for the Imperial Valley model, and a taper ratio of 0.49, resulting in a Hann window (Figures 3d and 3h). We use symmetric taper ratios for the window functions along the horizontal dimensions and a one-sided taper with the same taper ratio in the vertical dimension (Figures 3g and 3h). Therefore, all blending maps are identified by the window function support in the Coachella and Imperial valleys' model domains and the taper ratios. We also use the same polygons at all depths for the local models. However, we note that the models have different ray coverage that changes the shape and deteriorates by shrinking with depth, and the algorithm can accommodate this.

3.3. Smooth Hybrid Earth Models

Figure 4 shows selective maps and vertical cross-sections of P wave velocities for the community models CVM-H 15.1 (Figures 4a and 4d) and CVM-S 4.26 (Figures 4g and 4j) and hybrid models developed using the blending maps shown at the same depth and cross-section in Figure 3. The S wave velocity models are shown in Figure S4. We embed the local models in their full volume (supp1) using boxcar window functions and high-resolution coverage polygons (supp2) using symmetric cosine tapers. For the CVM-H 15.1 hybrid model with no boundary smoothing (Figure 4b), we note the contrast in detail in the local models, including higher velocity heterogeneities and a more defined Coachella Valley basin structure compared to the more homogenous crustal structure present in CVM-H 15.1 (Figure 4a). At the current depth shown, the difference between the CVM-H 15.1 model and the Imperial Valley local model is less significant except at the boundaries of the Imperial Valley model like the northeast and southwest corners where low velocities extend outside the Salton Trough and in the southeast corner of the model where velocities are relatively high. In the cross-section (Figure 4e), there are sharp differences along the entire boundary of the local models, and we note the ~ 40 km northwestward extension of the basin structure provided by the Coachella Valley local model. CVM-S 4.26 has high-velocity structures that define a more heterogeneous crustal structure than CVM-H 15.1 and is thus comparable to the Coachella Valley model. However, discontinuities can still be observed in the hybrid model of CVM-S 4.26 (Figure 4h), particularly at the boundaries of the local models, with the Imperial Valley model having similar areas of disagreements as in the CVM-H 15.1 model but with higher velocities in the case of the CVM-S 4.26. Again, in the cross-section (Figure 4k), there are sharp contrasts around the boundaries of the local models, similar to that in the CVM-H 15.1 hybrid model. These hybrid models generated with no boundary smoothing represent the default model parameters returned by the current version of UCV. The discontinuities do not correspond to known geologic structures, faults, or compositional boundaries and represent artifacts in the hybrid model.

For the CVM-H 15.1 hybrid model with boundary smoothing (Figures 4c and 4f), we use the blending maps of Figures 3c and 3g, which use irregular polygon support that are the high-resolution coverage areas for the Imperial and Coachella valleys' local models with symmetric cosine tapering. The additional detail in crustal heterogeneity and changes in the velocity distribution within the Coachella Valley local model's basin are clear, and overall, there are no sharp boundaries in the entire model. The hybrid model resembles CVM-H 15.1 more in the Imperial Valley, with additional detail from the Imperial Valley model. For the smooth CVM-S 4.26 hybrid model (Figures 4i and 4l), we use the blending maps developed with symmetric cosine taper functions and maximum taper ratios (Figures 3d and 3h). In all hybrid models that use tapering, the local models are assimilated nicely, and there are no sharp discontinuities associated with the local models except those from the background community models due to their historical development of embedding basin models without smoothing. As we shall see, the differences in the basin structure and crustal heterogeneity highlighted in the hybrid and community models play a major role in ground motion prediction for the entire hybrid model domain. For all models considered, we include topography and modify the model parameters in the top 350 m by including a geotechnical layer that provides higher resolution near-surface crustal properties (Ely et al., 2010). We evaluate the smooth hybrid models that utilize boundary smoothing against the community models for low-frequency seismogram prediction in the validation exercise.

3.4. Earthquake Simulations and Wavefield Predictions

The simulation domain's finite element mesh contains ~ 12 million grid points with minimum and maximum grid spacing of ~ 92 and $\sim 1,700$ m, respectively and an average grid spacing of ~ 150 m at the surface. By restricting the minimum velocity in all the models to 600 m/s and using five-grid-points per wavelength sampling for each element in the mesh, our simulations are globally accurate to the ~ 2 s period. We simulate ground motions for the selected earthquakes (Figure 1) to produce 200 s seismograms using a time step of ~ 0.007 s to ensure solution stability. In terms of computational demands, each simulation costs ~ 800 CPU-hours. We determine S wave velocities and densities for the local SSIP models that provide only P wave velocities by using well-known empirical relationships (Brocher, 2005). We determine the quality factor Q used in describing the viscoelastic rheology for all models by using the empirical relationship from Olsen et al. (2003) that scales S wave velocities and use an Olsen attenuation ratio of 0.05. Our constant- Q

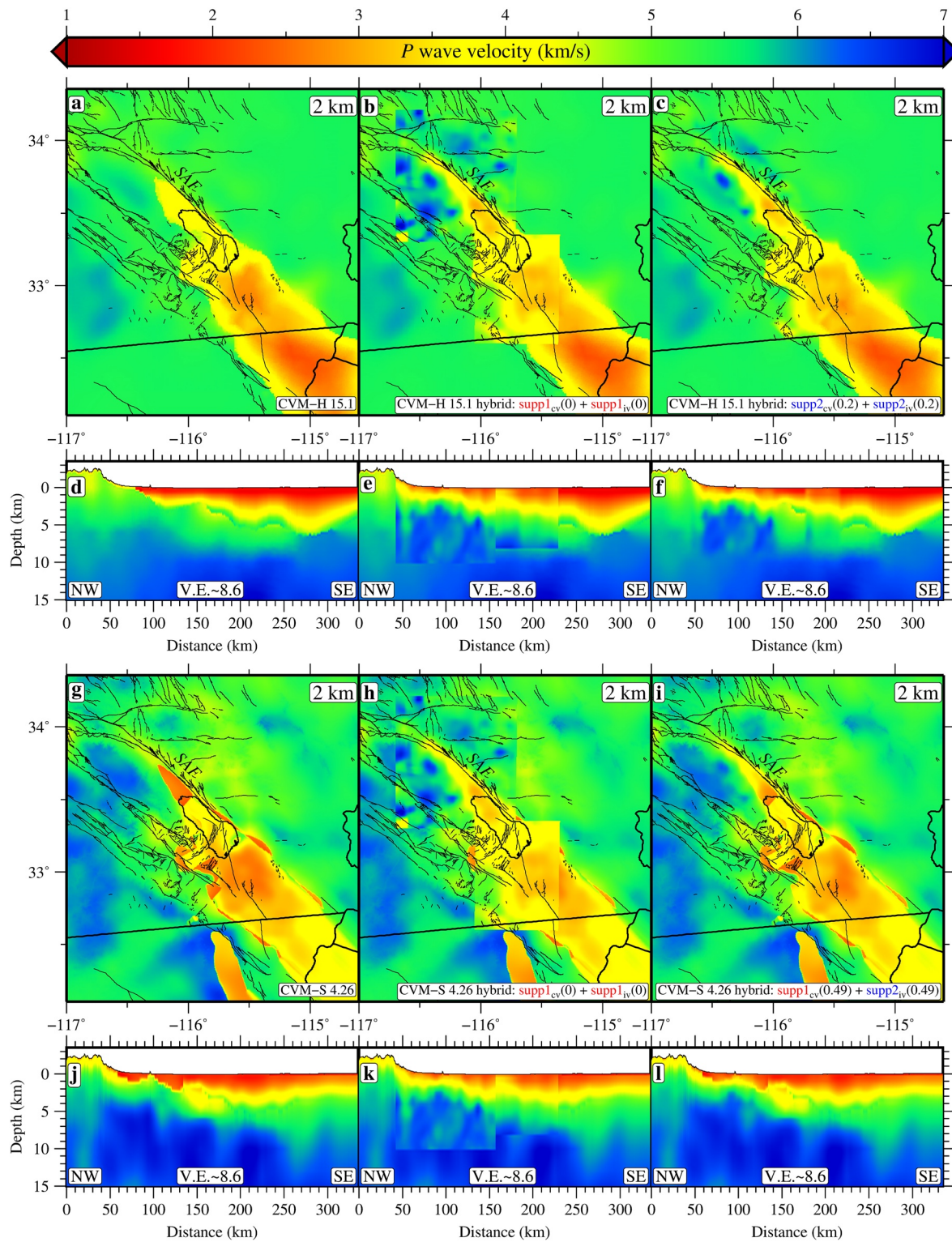


Figure 4.

approximation is valid in a period range of ~ 2 – 108 s, which includes the 6 – 30 s period range that we consider. The thickness of the sedimentary section in Salton Trough ranges from a maximum of ~ 4 km in the Coachella Valley to ~ 7 km in the Imperial Valley. Assuming an average S wave velocity of $1,500$ m/s in the Salton Trough basin (Figure S4), the S wave resonant period is ~ 11 and ~ 19 s in the deepest parts of Coachella and Imperial valley basins, respectively. Therefore, we expect that the differences in the basin structure represented in the models will impact the long-period wavefield predictions. Other factors including the basin shape, edge, and attenuation also play a significant role in the forecast (Brissaud et al., 2020; Rial et al., 1992).

Figure 5 shows an example of the wavefield generated by event 4 in the CVM-H 15.1 and hybrid CVM-H 15.1 models showing the three components of ground velocity at ~ 50 s into the simulation. We can observe the basin amplification effects in the Coachella and Imperial valleys in the models that correspond to low seismic velocities regions in Figure 4. On the horizontal components in the hybrid model (e.g., Figures 5e and 5f), one can observe the seismic waves channeled northward along subbasins in Coachella Valley (Ajala, Persaud, Stock, et al., 2019) and strands of the San Andreas fault. These sites represent high seismic hazard areas and often have high population densities of around 500 persons per square km in some regions (inset map in Figure 1). For the hybrid model with the boxcar window function and full support, we can observe the effects of sharp discontinuities in the wavefield manifested by reflection and refraction artifacts at the Imperial Valley model boundary (arrows in Figure 5b). In the smooth hybrid model (e.g., Figure 5c), these artifacts are subdued, but there are amplifications in the basins particularly around the Salton Sea from the background CVM-H 15.1 model.

We compare the predicted seismograms from our wavefield simulations at the available broadband stations with displacement records with a high signal-to-noise (S/N) ratio to validate the models. In evaluating the models for their ground motion prediction ability, we use both a qualitative method by visualizing the seismograms and a quantitative assessment by measuring the misfit between the predicted and observed wavefield. Figure 6 shows selected examples of 6 – 30 s seismograms that highlight different waveform fit levels for the models. For event 9 recorded at station TOR in the Coachella Valley and event 1 recorded at station SWS in the Imperial Valley, the waveforms produced by all the models except CVM-H 15.1 decently fit the observations. CVM-H 15.1 overpredicts the horizontal component of the ground motion amplitudes at the stations, while the hybrid CVM-H 15.1 model appears to fit the observations best. We have a similar situation for event 9 recorded at station WWF located in the Imperial Valley, where CVM-H 15.1 significantly overpredicts the ground motion amplitudes compared to the rest of the models. At this station, CVM-S 4.26 appears to provide the best fit for the observations. For stations like DZA located outside the local model domains, the hybrid model matches the observations better than the community models that have sustained amplifications not required by data. We also find unique cases like event 6 recorded at station THM in the Coachella Valley, where the CVM-H 15.1 model best predicts the amplification in the basins. For all the models, we find cases where the model predictions poorly match the observations. An example is event 6 recorded at station MSJ in the mountain ranges outside the local models' domain, where none of the models reproduce the surface waves and reverberations present on the horizontal components (Figure 6). For a more rigorous assessment, we compute the relative waveform misfit at the stations with high-quality data.

Figure 7 shows the spatial distribution of the average RWM for all components and events recorded at each station in each model. Comparing the misfit maps of the CVM-H 15.1 model (Figure 7a) with that of the CVM-H hybrid model (Figure 7b), there is significant improvement in the wavefield predictions within the basins, which represents the most hazardous sites with dense population. Embedding the local Imperial and Coachella valleys' models into the CVM-H 15.1 model improves the wavefield prediction ability. In contrast,

Figure 4. P wave velocity models of hybrid and community models. (a) Map of P wave velocities of the community model CVM-H 15.1 at 2 km depth below sea level. (b) The resulting CVM-H 15.1 hybrid model upon using the blending weights in Figure 3b. (c) The resulting CVM-H 15.1 hybrid model upon using the blending weights in Figure 3c. (d) Cross-section of the community model CVM-H 15.1 along the profile indicated by the thick black line in Figure 3a. (e) The CVM-H 15.1 hybrid model generated from the blending weights in Figure 3f. (f) The CVM-H 15.1 hybrid model developed from the blending weights in Figure 3g. (g) Map of P wave velocities of the community model CVM-S 4.26 at 2 km depth below sea level. (h) The resulting CVM-S 4.26 hybrid model from using the blending weights in Figure 3b. (i) The resulting CVM-S 4.26 hybrid model upon using the blending weights in Figure 3d. (j) Cross-section of the community model CVM-S 4.26 along the profile indicated by the thick black line in Figure 3a. (k) The CVM-S 4.26 hybrid model generated from the blending weights in Figure 3f. (l) The CVM-S 4.26 hybrid model developed from the blending weights in Figure 3h. Figure S4 shows the S wave velocity model for the same maps and profiles.

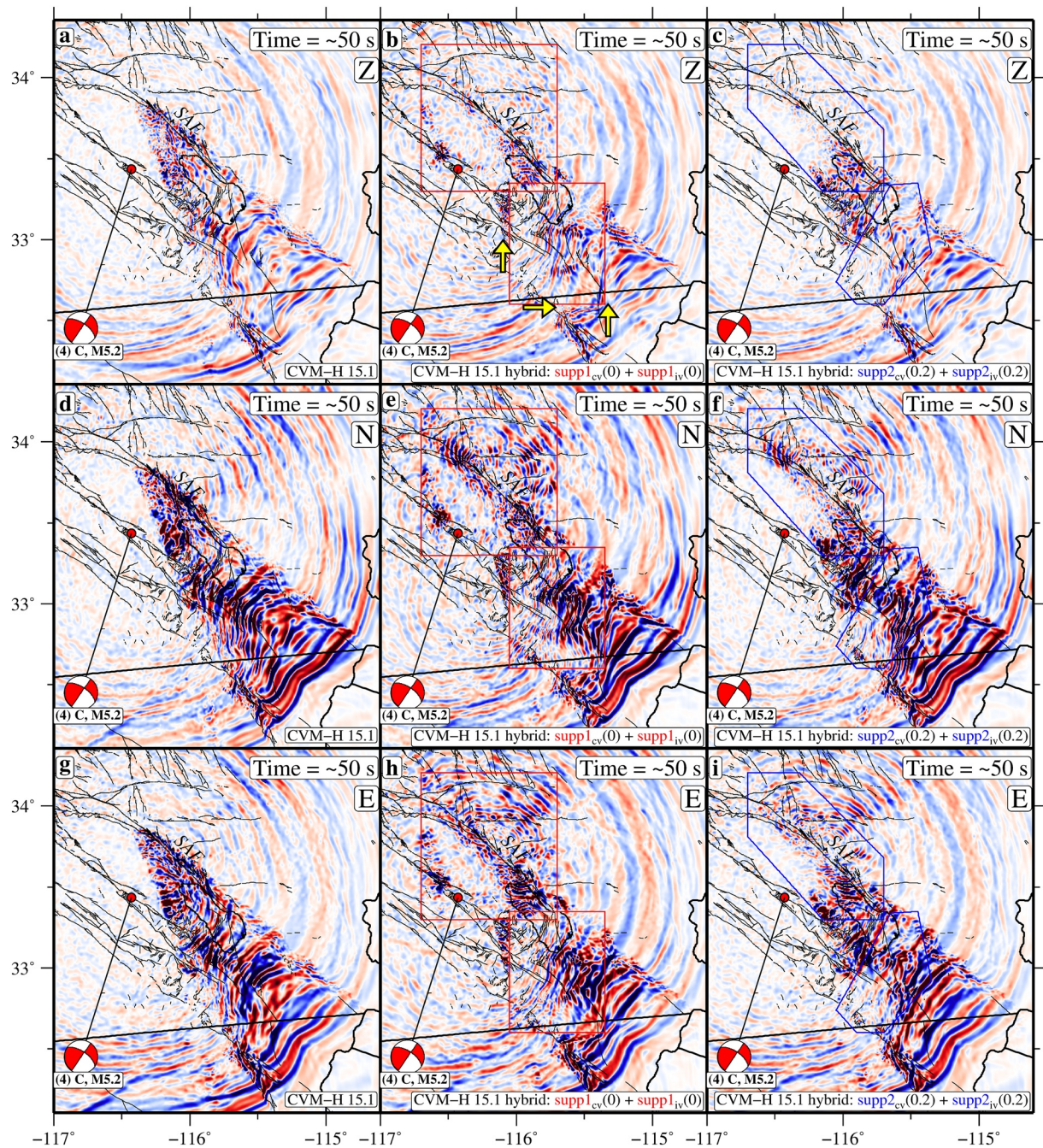


Figure 5. Wavefield snapshots in the CVM-H 15.1 hybrid and community models. Maps showing three component (Z-vertical; N-north-south; E-east-west) ground velocities for earthquake 4 at ~50 s. (a, d, g) Community model CVM-H 15.1 (Figures 4a and 4d) highlighting seismic amplification in the triangular region representing the basin. (b, e, h) Hybrid model of Figures 4b and 4e showing wavefield artifacts that correspond to the boundary of the local model in the Imperial Valley that is marked with yellow arrows. (c, f, i) Hybrid model of Figures 4c and 4f.

CVM-S 4.26 (Figure 7c) and CVM-S 4.26 hybrid (Figure 7d) misfit maps look similar, with notably increased misfit at selective stations in the Coachella Valley such as station THM. In the Imperial Valley, the CVM-S 4.26 hybrid model generally has more stations with improved average misfit values. The similarity in the waveform misfit for the CVM-S 4.26 models can be attributed to the fact that the CVM-S 4.26 hybrid model uses a blending map with the maximum taper ratio (Figure 3d) providing little contributions from the local models in the Imperial and Coachella valleys. These results indicate how sensitive ground motion estimates can be to model parameter changes. By visual inspection of all misfit maps, the CVM-H 15.1 model has

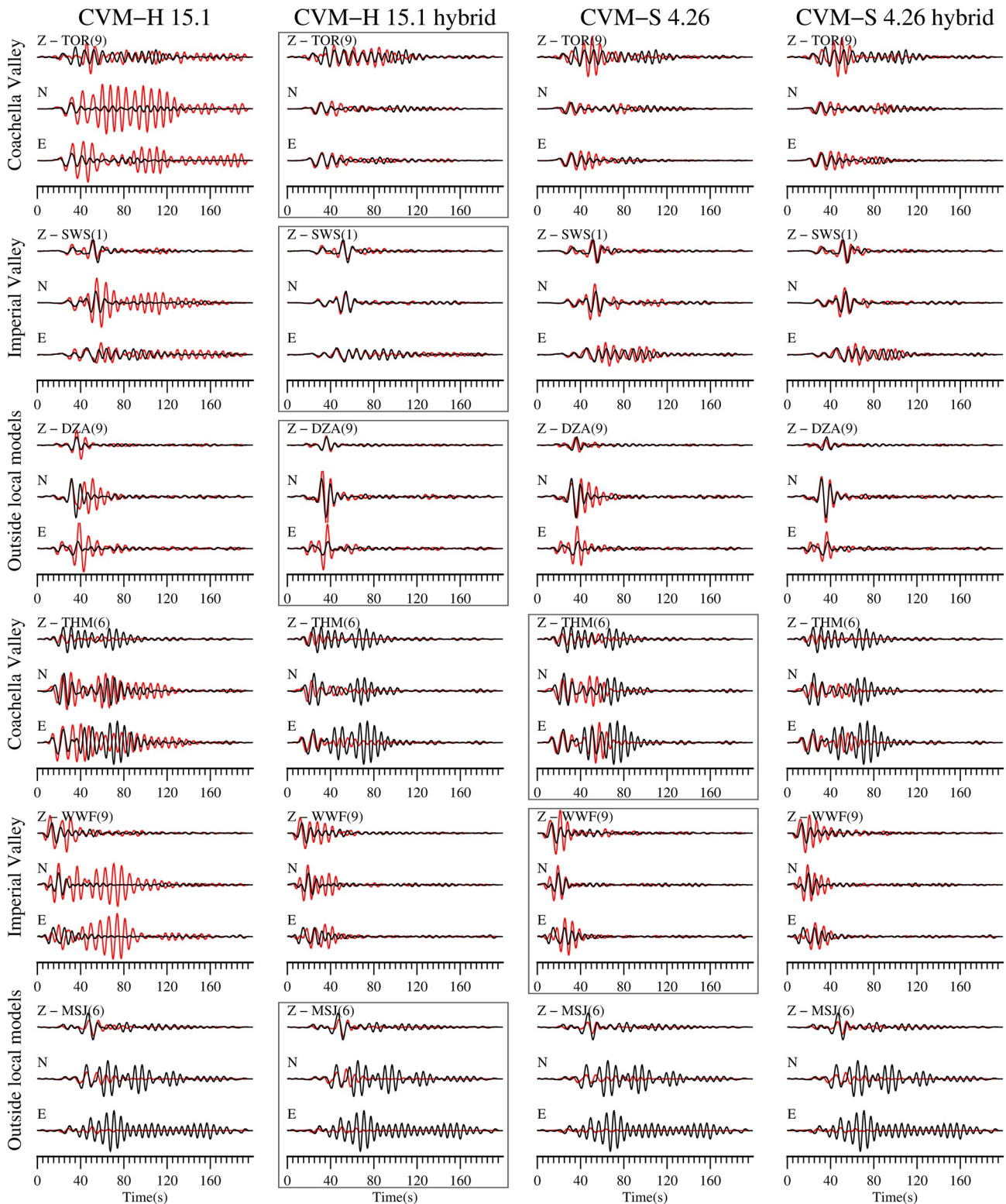


Figure 6. Examples showing comparisons of 6–30 s three-component (Z-vertical; N-north-south; E-east-west) ground displacement records between observed seismograms (black) and predicted seismograms (red) generated from the community models and hybrid models labeled in the top row, at selective stations labeled in Figure 1. Station locations are labeled on the left and station names are indicated next to the Z component name with the number in the parenthesis corresponding to the events in Figure 1. Waveforms with the lowest average misfit are highlighted with a gray box.

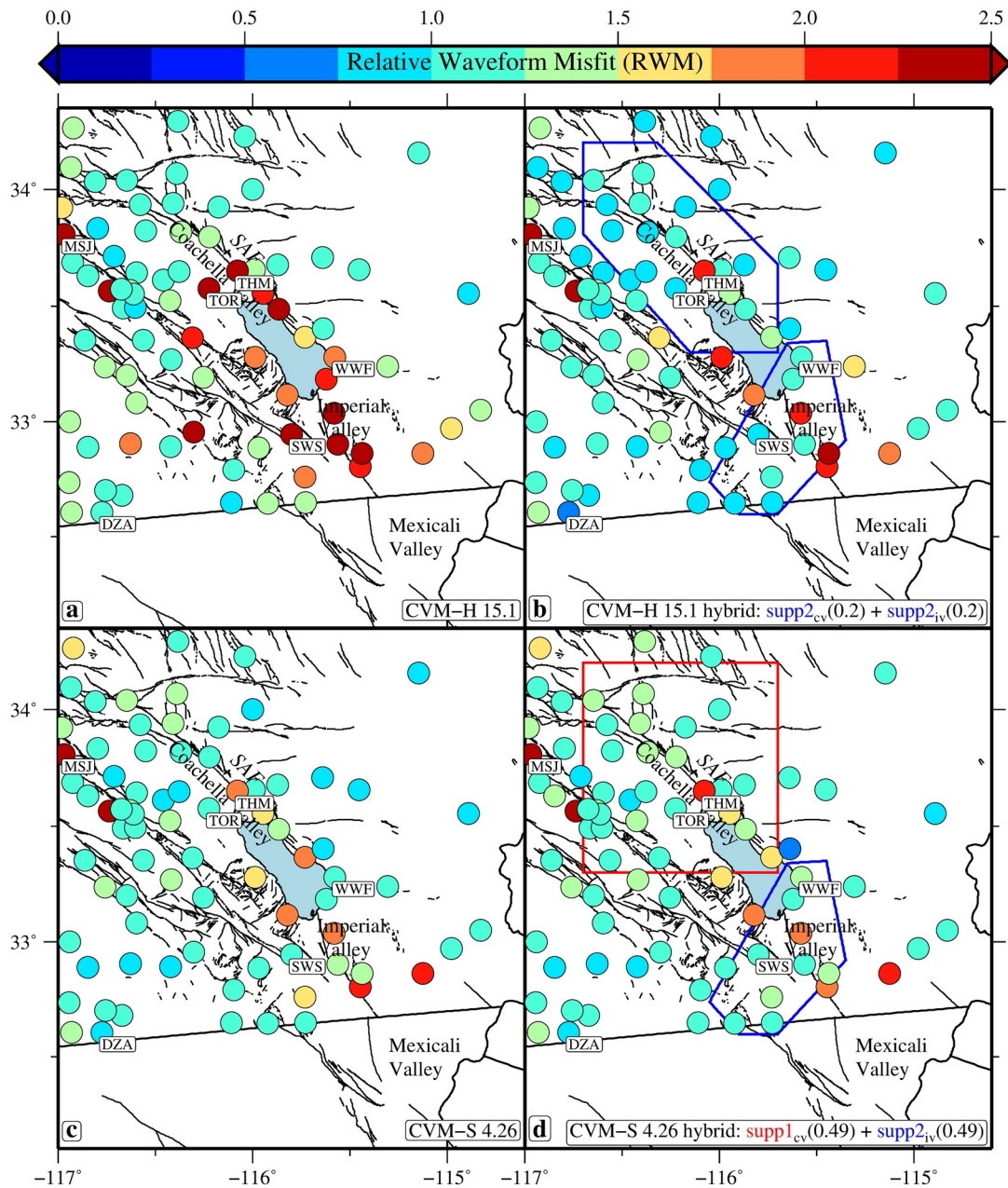


Figure 7. Maps of the average relative waveform misfits for all events and components color-coded at individual broadband stations. (a) Misfit map for the CVM-H 15.1 model. (b) Misfit map for the CVM-H 15.1 hybrid model shown in Figures 4c and 4f. (c) Misfit map for the CVM-S 4.26 model. (d) Misfit map for the CVM-S 4.26 hybrid model shown in Figures 4i and 4l.

more stations with high RWM values. Finally, to rank the models based on the RWM metric, we consider the misfit of all predicted and observed seismogram pairs generated by the 11 events for each station with a high S/N ratio resulting in 1719 pairs and plotted the histograms (Figure 8). Due to the misfit distribution's asymmetry, we use the median and median absolute deviation as robust central tendency measures. There is a ~24% reduction in misfit from CVM-H 15.1 (Figure 8a) to the CVM-H hybrid model (Figure 8b). The misfit histograms for the CVM-S 4.26 (Figure 8c) and CVM-S 4.26 hybrid models (Figure 8d) have a similar distribution with a slight ~0.6% increase in misfit in the CVM-S 4.26 hybrid model, highlighting another key result showing that embedding local models into regional models may not always lead to an improvement

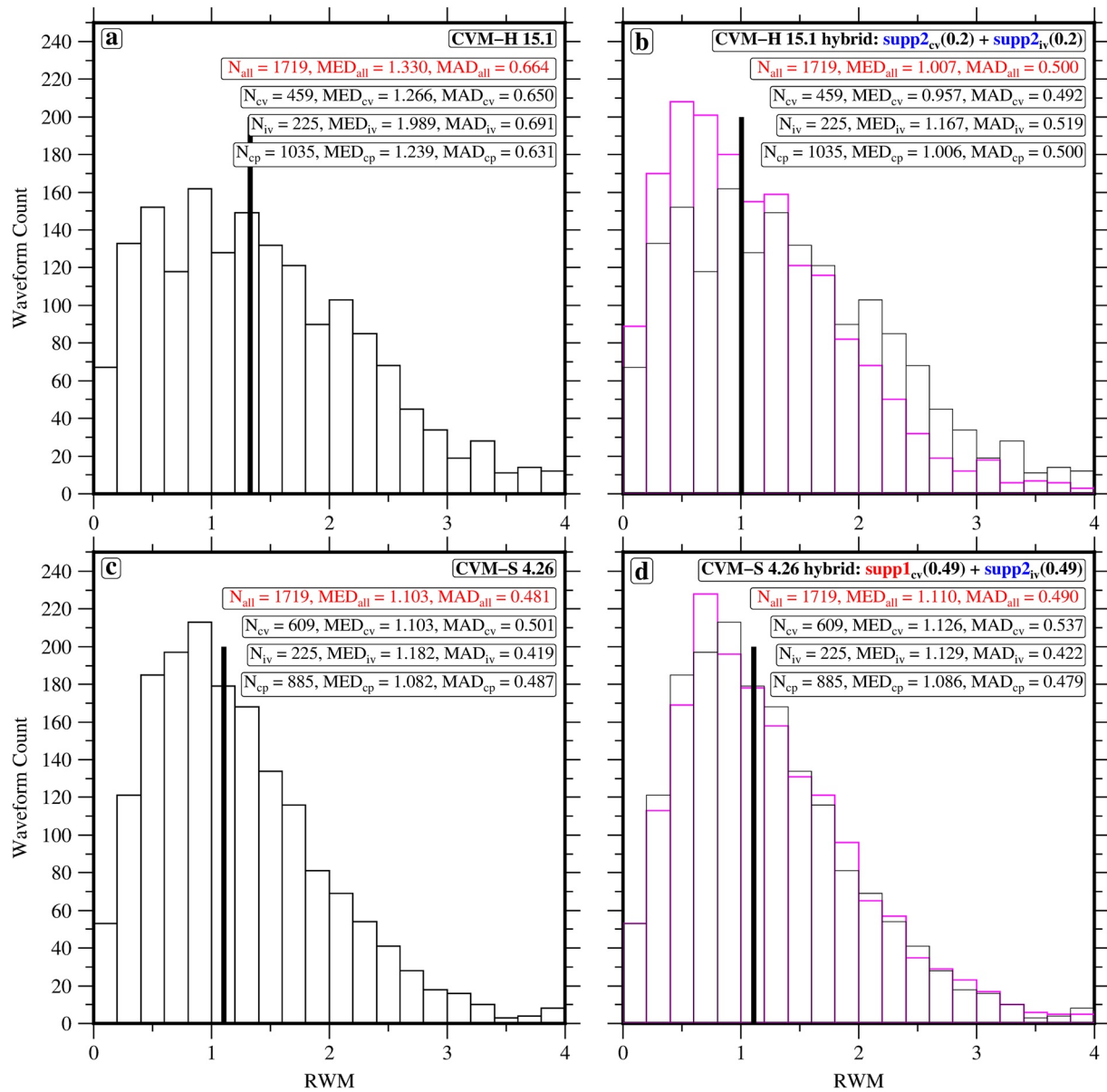


Figure 8. Histograms of the relative waveform misfits for high S/N waveforms in the entire model domain corresponding to the models and misfits shown in Figure 7. (a) Misfit histogram for the CVM-H 15.1 model. (b) Misfit histogram for the CVM-H 15.1 hybrid model (pink) overlain with the CVM-H 15.1 model misfit histogram in gray. (c) Misfit histogram for the CVM-S 4.26 model. (d) Misfit histogram for the CVM-S 4.26 hybrid model (pink) overlain with the CVM-S 4.26 model misfit histogram in gray. N—number of seismograms; MED—median (vertical black line); MAD—median absolute deviation. The subscript of the statistics given in the legend indicates the entire model domain (all), the Coachella Valley model domain (cv), the Imperial Valley model domain (iv), and the complement (cp), which refers to the entire model domain without the local model domains. A summary of the misfit statistics is provided in Table S2.

in the waveform prediction ability of the hybrid model. Of all the models, the CVM-H 15.1 hybrid model has the lowest median RWM.

4. Discussion

4.1. Window Taper Function Parameters

Our choice in the selection of the window function, support, and taper ratio is somewhat subjective. We chose the cosine taper function because it is often used for smoothing overlapping gridded models in the

spatial domain (Molinari & Morelli, 2011) and filtering geophysical signals (Park et al., 1987). In this manner, we view the cosine taper as a canonical window function in scientific applications. As previously noted, any of the many available window functions, such as the trapezoid function (Figure S3), can be used (Dorothy, 2017). The support of the window function should correspond to the volume of the detailed models that is well constrained by data. Since we use local models derived from travel time tomography based on ray theory in this study, the support (supp2) of the window function roughly correlates with regions of the models with high ray coverage. For adjoint tomographic models, the support would be related to the misfit gradient. Finally, the taper ratio should be selected such that there are no artificial interfaces created along the boundaries of the embedded models. One can achieve this by visually inspecting the hybrid models, considering the geology and qualitatively gauging the appropriate amount of required smoothing.

4.2. Simulation Limitations

We note several simplifications used in the simulation that could potentially represent sources of errors. These include uncertainty in the earthquake source parameters, empirical model parameters, and the limited inclusion of low near-surface velocities. We use the earthquake locations and source parameters from SCEDC without performing any inversion. The focal mechanism quality (Figure 1) is based on nodal plane uncertainties determined using the *P* arrival polarities and *S/P* wave amplitude ratios (Yang et al., 2012). Six out of the 11 earthquakes we use have a C quality, which represents an uncertainty of 35°–45°. Three events have a quality of A ($\leq 25^\circ$), and two events have a quality of B (25°–35°). These can introduce energy directivity errors into the simulation (Aagaard et al., 2008). Similar validation studies like Lee, Chen, and Jordan (2014) that refine the source parameters of two medium moment magnitude earthquakes (M_w 4.4 and 5.1) through waveform inversion report significantly better predictions in low-frequency (0.02–0.2 Hz) seismograms. For the SSIP models, we empirically determine the *S* wave velocities and densities, and for all the models, we use empirical attenuation models. Having independent constraints on each model parameter may lead to better results, especially *S* wave velocities that have a dominant effect on low-frequency waveforms and attenuation in the deep sedimentary basins. To reduce the computational cost of the simulations, we set the minimum velocities to 600 m/s. However, with the Ely geotechnical layer, the actual near-surface velocities are lower than 200 m/s. Even for long periods, the shallow low velocity layer is known to influence the seismic wavefield and, as such, represents a source of error (Juarez & Ben-Zion, 2020). Also, the Salton Sea, which should be an acoustic domain, is not incorporated in the simulations. The error contribution from these factors will become more critical at higher frequencies and should not be ignored whenever possible.

4.3. Spatial Misfit Distribution

Upon closer inspection of the average waveform misfit maps (Figure 7), we can observe that the stations with changes in average misfit between the community and hybrid models are not limited to the domain of the local models. For example, several stations like DZA in the mountain ranges show an average improvement in ground motion prediction in the CVM-H 15.1 hybrid model (Figure 7b) compared to the CVM-H 15.1 model (Figure 7a). The reverse situation can be observed by considering the average station misfits of the CVM-S 4.26 (Figure 7c) and CVM-S 4.26 hybrid (Figure 7d) models. These results indicate that embedding local models can improve or deteriorate the wavefield prediction ability of the community model over the entire model volume. In the case of improvements, we hypothesize that the local models provide a more accurate crustal structure in the wave path segment between the source and receiver that samples the local models. Nevertheless, the majority of the improvements happen in the domain of the local models. To quantify the misfit reduction in the different areas of our simulation domain, we group the stations based on their location in the domain of the embedded local models into the Coachella Valley (cv), the Imperial Valley (iv), and everywhere else excluding cv and iv is called the complement (cp). We consider these three regions in our misfit histograms (Figures 8 and S5–S7).

The number of waveform pairs in each subdomain varies and depends on the support polygon size for the local models. For the CVM-H models that use the high-resolution coverage areas, the number of waveform pairs in the Coachella Valley is lower than for the CVM-S models that use the full coverage. The number of waveform pairs for the Imperial Valley domain is the same for all models since the hybrid models uti-

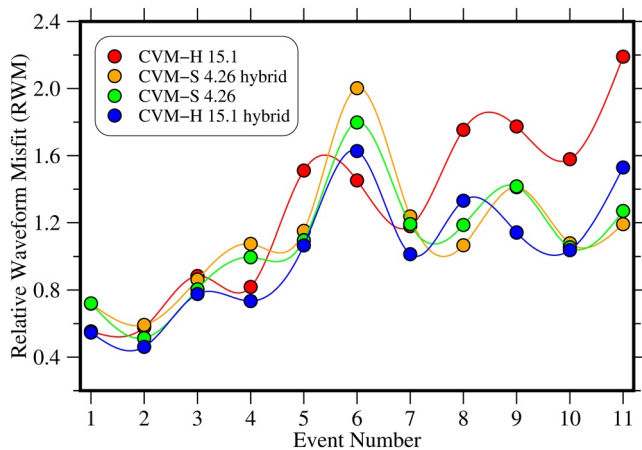


Figure 9. Comparison of waveform misfits by event. Median value of the relative waveform misfits for each model and event.

lize the Imperial Valley model's high-resolution coverage. However, the number of waveform pairs considered in the complement region is higher for the CVM-H models than the CVM-S models. So, only the results between each community model and the resulting hybrid model are directly comparable in this case. In the CVM-H model, the area with the highest RWM value is the Imperial Valley, followed by the Coachella Valley and the complement region. By comparing the misfit changes of CVM-H 15.1 with the CVM-H 15.1 hybrid model, the region with the most significant improvement in the waveform misfit is the Imperial Valley domain with a $\sim 41.1\%$ decrease (Figures S6a and S6b). In the Coachella Valley domain, we see a $\sim 24.3\%$ misfit reduction (Figures S5a and S5b), and in the complement, there is a misfit reduction of $\sim 18.7\%$ (Figures S7a and S7b). These results show that CVM-H 15.1 has a poor representation of the Imperial Valley basin structure and is most benefitted by embedding the Imperial Valley local model. The CVM-S 4.26 model has relatively low misfit values in the three regions compared to the CVM-H 15.1 model. In the CVM-S 4.26 hybrid model, there is a 2% increase in the waveform misfit in the Coachella Valley (Figures S5c and S5d) and a 0.4% increase in cp (Figures S7c and S7d) compared to CVM-S 4.26 alone. However,

there is an improvement in the waveform fit in the Imperial Valley domain, with a $\sim 4.6\%$ decrease in misfit (Figures S6c and S6d). These results indicate how wavefield simulations can be used to determine which local models to embed. For the CVM-S 4.26 model, embedding the Imperial Valley model alone might improve the overall waveform prediction ability of the CVM-S 4.26 model. Results from the CVM-S 4.26 community and hybrid models emphasize that little contributions from local models can globally affect the waveform prediction ability of the resulting hybrid model. We note that the misfit results depend on the discrete locations that we sample in the entire wavefield. We utilize 79 broadband stations, and though widely distributed, each of the 11 selected events is not always well-recorded at all of the stations (Figure 1).

4.4. Misfit Variability With Events

Figure 9 shows that a single event is not sufficient for validating Earth models since the results can vary across different events (Taborda et al., 2016). Although the CVM-H 15.1 hybrid model has the lowest overall waveform misfit of all the models that we evaluate, this is not the case for all individual events. For example, considering events 6, 8, and 11, the CVM-H hybrid model does not produce the lowest misfit. For event 6, the CVM-H 15.1 model achieves the best wavefield predictions. CVM-S 4.26 hybrid outperforms others for events 8 and 11. If only these three events were used in our misfit assessment, we would have reached a different conclusion. However, it is not uncommon for fewer than three events to be used in validation exercises (Fang et al., 2016; Lee, Chen, & Jordan, 2014; Taborda & Bielak, 2014). There is substantial variability in the misfit at the recording station for each model depending on the earthquake's location (Figures S8–S11). The general trend and an important result is that the misfit values for the events are higher for events located near the southern end of the San Andreas fault and farther south. This is especially apparent in the average misfit maps for each event for the CVM-H model (Figure S8). Almost all of the stations for event 11 have high misfit values (Figure S8k). By comparing the average misfit maps for the CVM-H 15.1 and CVM-H 15.1 hybrid models (Figure S9), it is immediately apparent that there is an improvement in misfit for almost all the events, except for event 6. CVM-S 4.26 has good misfits for almost all events (Figure S10), except for events 6 and 9, where there are considerably more stations with high misfit values. Comparing these results with the CVM-S 4.26 hybrid model (Figure S11), we can observe an increase in misfit in the Coachella Valley domain for most of the events and a decreased misfit in the Imperial Valley domain. Of all the events, 5 and 6 are particularly curious since their epicenters are located close to the hypothetical rupture site of the San Andreas fault, and they can represent a lower magnitude scenario on nearby faults. For event 5, the CVM-H 15.1 hybrid model produces the best predictions while CVM-H 15.1 has the worst misfit values. In contrast, for event 6, CVM-H 15.1 outperforms all the other models. An excellent example of the CVM-H 15.1 model performance for event 6 is shown in the waveform comparisons between the predicted and observed seismograms at station THM in the Coachella Valley (Figure 6). CVM-H 15.1

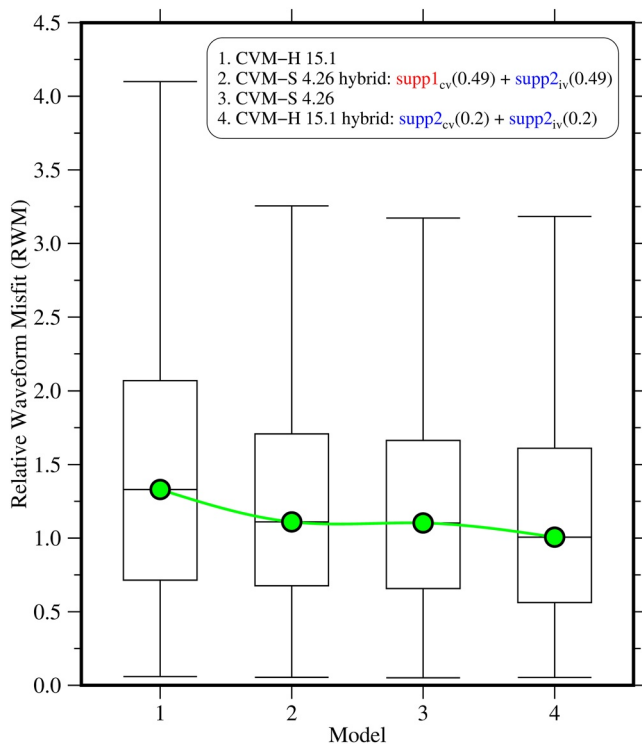


Figure 10. Summary of waveform misfits by model. Boxplot of the relative waveform misfit for the four models ordered by decreasing median relative waveform misfit.

does the best in matching the surface wave amplitudes on the horizontal components. For this event, the CVM-S 4.26 hybrid model produces the worst misfit. We note that while the event is at the boundary of the local models, the boundary smoothing ensures only minimal changes in the model parameters initially present in the community models. These results emphasize the importance of having accurate models in the close vicinity of the scenario events. Based on the waveform misfit results for the CVM-H 15.1 model for the other events, we hypothesize that the CVM-H 15.1 model is the best near-source model for event 6 and thus produces better wavefield predictions. In the different subdomains, we also observe significant variability in the waveform fit (Figure S12). CVM-H 15.1 produces significantly higher misfit values in the Imperial Valley domain for almost all the events, especially events 5 and 9 (Figure S12c). Again, the general trend of increasing misfit with decreasing latitude of the events, particularly for CVM-H 15.1, is strong and emphasizes our earlier point of CVM-H 15.1 having low-resolution model parameters in the Imperial Valley domain. Based on these results, we can state that a validation exercise is not well defined for a single event since the model with the lowest RWM is not the same for every event. Thus, as many events as possible that sample different regions of the entire model domain should be considered during the evaluation of different models. In theory, we want to utilize many events that sample almost everywhere in the simulation domain.

4.5. Model Validation

Figure 10 shows the boxplot summary of RWM for our validation exercise. The largest reduction in the median RWM of $\sim 24\%$ is between the CVM-H 15.1 and CVM-H 15.1 hybrid models (models 1 and 4). On the

other hand, the CVM-S 4.26 hybrid model has a $\sim 0.6\%$ modest increase in median RWM relative to CVM-S 4.26 (models 2 and 3). However, the CVM-H 15.1 hybrid model also has a $\sim 9\%$ reduction in median RWM relative to CVM-S 4.26 (models 3 and 4). So, the CVM-H 15.1 hybrid model has the lowest overall median RWM of the four models we validate. The scatter in the boxplot indicates that there are still some problematic regions in the models that need refinement (e.g., station MSJ in Figure 7). The overlap in the interquartile range is not surprising since the community models are validated in a similar frequency band as the seismograms used in their development. We, therefore, expect that the community models will produce a decent fit to the low-frequency waveforms.

Lastly, we note that there is no strict definition of what constitutes an improvement in a validation exercise. While we use the RWM metric, several other misfit measures can be utilized to quantify different aspects of the wavefield predictions. As an example, Fang et al. (2016) evaluate their new southern California model developed by jointly inverting both body and surface waves against CVM-H 15.1 using a single event and the vertical components of 10 velocity waveforms. They use the correlation coefficient as a misfit metric, and the misfits are scattered, significantly overlap, and have a $\sim 2\%$ overall reduction in median value. The authors state that their model improves the ability to simulate earthquake waveforms as a major conclusion of the study. Taborda et al. (2016) also present improved wavefield predictions for CVM-S 4.26 over three other community models based on $\sim 1\%$ – 6% in overall mean misfit reduction in the goodness-of-fit (GOF). The GOF values show scatter and significant overlap among the four models. Structural engineers focus more on response spectra and peak ground motions and are less interested in the full-waveform fit or the vertical component of the seismograms (Maechling et al., 2007). To summarize, a different choice of misfit measure and changes in any of the modeling parameters and assumptions or using a different data selection and processing scheme can produce a different result than we present.

5. Conclusion

We developed a flexible technique to define window functions such as the cosine taper using arbitrary support on regular grids. This tool can smoothly embed detailed local models developed using dense seismic arrays and novel imaging methods into regional models in irregular volumes. Our algorithm is implemented in the open-source Unified Community Velocity Model software to facilitate the reproducibility of complex hybrid models. We test our method by embedding high-resolution local models developed using active source data into two popular community models - CVM-H 15.1 and CVM-S 4.26 - used for simulation-based seismic hazard analysis in southern California. Using the relative waveform misfit, the derived hybrid models are evaluated against the community models in terms of low-frequency wavefield prediction for 11 local earthquakes that were not used in developing the models. The hybrid CVM-H 15.1 model produces the lowest median relative waveform misfit, suggesting that our method offers a computationally tractable approach for updating large-scale community Earth models used for different geologic and hazard studies. However, merging multiscale models may not produce a better model as we observe with the CVM-S 4.26 hybrid model being degraded relative to the CVM-S 4.26 model. We also show that the process of model validation can be complicated and gives spatially variable results. Hybrid models therefore need to be analyzed for spatial misfit distribution to ensure that any improvements are global and not locally restricted to the domain of the embedded models, and one should use as many validation events as possible in different parts of the model.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The blending library is available at <https://doi.org/10.5281/zenodo.4552676> (Ajala, 2021a). The non-official updated version of UCVL that implements our blending library is available at <https://doi.org/10.5281/zenodo.4533337> (Ajala, 2021c). All waveform and simulation data including scripts needed to reproduce the entire results and figures in the paper can be found here <https://doi.org/10.5281/zenodo.4533394> (Ajala, 2021b).

Acknowledgments

We thank the Editor Yehuda Ben-Zion, Associate Editor Andreas Fichtner, and reviewers John Vidale and Arthur Rodgers for comments and suggestions that helped improve the manuscript. R. Ajala thanks Alan Juarez for helpful discussions about seismic wave propagation simulations. Computational resources were provided by the LSU high performance computing center. All figures are plotted using the Generic Mapping Tools (Wessel et al., 2019). The work was supported by the National Science Foundation award 2105320 and SCEC research awards 18074, 19014, 20023, and 21059. The SCEC contribution number for this paper is 10948. R. Ajala was also supported by the merit-based scholarship from the Society of Exploration Geophysicists Foundation. P. Persaud was supported as a 2020–21 fellow of the Radcliffe Institute for Advanced Study at Harvard University.

References

- Aagaard, B. T., Brocher, T. M., Dolenc, D., Dreger, D., Graves, R. W., Harmsen, S., et al. (2008). Ground-motion modeling of the 1906 San Francisco earthquake, part I: Validation using the 1989 Loma Prieta Earthquake. *Bulletin of the Seismological of America*, 98(2), 989–1011. <https://doi.org/10.1785/0120060409>
- Ajala, R. (2021a). *Algorithm for merging gridded multiscale and multidimensional datasets using window functions*: Zenodo. <https://doi.org/10.5281/zenodo.4552676>
- Ajala, R. (2021b). *Data and scripts to reproduce research on seismological application of multiscale hybrid Earth models*: Zenodo. <https://doi.org/10.5281/zenodo.4548149>
- Ajala, R. (2021c). *Modified UCVL software with blending functionality*: Zenodo. <https://doi.org/10.5281/zenodo.4533337>
- Ajala, R., Persaud, P., Juarez, A., & Ayeni, G. (2019). *Evaluating seismic velocity models in Salton Trough, southern California using spectral-element wave simulation of validation events*. Paper presented at the Southern California Earthquake Center.
- Ajala, R., Persaud, P., Stock, J. M., Fuis, G. S., Hole, J. A., Goldman, M., & Scheirer, D. (2019). Three-dimensional basin and fault structure from a detailed seismic velocity model of Coachella Valley, Southern California. *Journal of Geophysical Research: Solid Earth*, 124, 4728–4750. <https://doi.org/10.1029/2018JB016260>
- Bianco, M. J., Gerstoft, P., Olsen, K. B., & Lin, F. (2019). High-resolution seismic tomography of Long Beach, CA using machine learning. *Nature*, 1–11. <https://doi.org/10.1038/s41598-019-50381-z>
- Bijelic, N., Lin, T., & Deierlein, G. G. (2019). Quantification of the influence of deep basin effects on structural collapse using SCEC cyber-shake earthquake ground motion simulations. *Earthquake Spectra*, 35(4), 1845–1864.
- Brissaud, Q., Bowden, C. D., & Tsai, V. (2020). Extension of the basin rayleigh-wave amplification theory to include basin-edge effects. *Bulletin of the Seismological of America*, 110(3), 1305–1322. <https://doi.org/10.1785/0120190161>
- Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the earth's crust. *Bulletin of the Seismological of America*, 95(6), 2081–2092. <https://doi.org/10.1785/0120050077>
- Brocher, T. M. (2008). Compressional and shear-wave velocity versus depth relations for common rock types in Northern California. *Bulletin of the Seismological Society of America*, 98(2), 950–968. <https://doi.org/10.1785/0120060403>
- Callaghan, S., Maechling, P. J., Goulet, C., Milner, K. R., Graves, R. W., Olsen, K. B., & Jordan, H. P. (2017). *Cybershake: Bringing physics-based PSHA to central California*. Paper presented at the Southern California Earthquake Center.

- Castellanos, J. C., Clayton, R. W., & Juarez, A. (2020). Using a time-based subarray method to extract and invert noise-derived body waves at Long Beach, California. *Journal of Geophysical Research: Solid Earth*, 125, 1–16. <https://doi.org/10.1029/2019jb018855>
- Center for International Earth Science Information Network - CIESIN - Columbia University. (2018). *Gridded population of the World, version 4 (GPWv4): Population density, revision 11*. <https://doi.org/10.7927/H49C6VHW>
- Clayton, R., & Engquist, B. (1977). Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the Seismological of America*, 67(6), 1529–1540.
- Clayton, R., Persaud, P., Denolle, M. A., & Polet, J. (2019). *Exposing Los Angeles's Shaky Geologic Underbelly* (Vol. 100). EOS. <https://doi.org/10.1029/2019eo135099>
- Day, S. M., Roten, D., & Olsen, K. B. (2012). Adjoint analysis of the source and path sensitivities of basin-guided waves. *Geophysical Journal International*, 189(2), 1103–1124. <https://doi.org/10.1111/j.1365-246x.2012.05416.x>
- Doerry, A. W. (2017). Catalog of window taper functions for sidelobe control. *Sandia Report SAN2017-4042*, 1–207.
- Ekstrom, G., Nettles, M., & Dziewonski, A. M. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201, 1–9. <https://doi.org/10.1016/j.pepi.2012.04.002>
- Ely, G. P., Small, P., Jordan, T. H., Maechling, P. J., & Wang, F. (2010). *A Vs30-derived near-surface seismic velocity model*. Paper presented at the AGU Fall Meeting.
- Fang, H., Zhang, H., Yao, H., Allam, A. A., Zigone, D., Ben-Zion, Y., et al. (2016). A new algorithm for three-dimensional joint inversion of body wave and surface wave data and its application to the Southern California plate boundary region. *Journal of Geophysical Research: Solid Earth*, 121, 3557–3569. <https://doi.org/10.1002/2015jb012702>
- Fichtner, A., van Herwaarden, D., Afanasiev, M., Simute, S., Krischer, L., Cubuk-Sabuncu, Y., et al. (2018). The Collaborative Seismic Earth Model: Generation 1. *Geophysical Research Letters*, 45, 4007–4016. <https://doi.org/10.1029/2018gl077338>
- Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., et al. (2013). Uniform California earthquake rupture forecast, version 3 (UCERF3)-the time independent model. *USGS Open File Report 2013-1165; CGS Special Report 228; Southern California Earthquake Center Publication 1792*.
- Fuis, G. S., Mooney, W., Healy, J., McMechan, G., & Lutter, W. (1982). Crustal structure of the Imperial Valley region. *U.S.G.S Professional Paper*, 1254, 25–49.
- Graves, R. W., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., et al. (2010). Cybershake: A physics-based seismic hazard model for southern California. *Pure and Applied Geophysics*, 168, 1–15. <https://doi.org/10.1007/s00024-010-0161-6>
- Graves, R. W., Pitarka, A., & Somerville, P. G. (1998). Ground-motion amplification in the santa monica area: Effects of shallow basin-edge structure. *Bulletin of the Seismological of America*, 88(5), 1224–1242.
- Hauksson, E. (2000). Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California. *Journal of Geophysical Research*, 105(B6), 13875–813903. <https://doi.org/10.1029/2000jb900016>
- Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2011). *Bulletin of the Seismological Society of America*, 102(5), 2239–2244. <https://doi.org/10.1785/0120120010>
- Hole, J. A. (1992). Nonlinear high-resolution three-dimensional seismic travel time tomography. *Journal of Geophysical Research: Solid Earth*, 97(B5), 6553–6562. <https://doi.org/10.1029/92JB00235>
- Hughes, T. J. R. (1987). *The finite element method, linear static and dynamic finite element analysis*: Prentice-Hall.
- Jennings, C. W., & Bryant, W. A. (2010). *Fault activity map of California: California geological survey geologic data map 2*.
- Jones, L. M., Bernknopf, R., Cox, D., Goltz, J., Hudnut, K., Mileti, D., et al. (2008). The ShakeOut scenario. *USGS Open File. Report 2008-1150*.
- Juarez, A., & Ben-Zion, Y. (2020). Effects of Shallow-Velocity Reductions on 3D Propagation of Seismic Waves. *Seismological Research Letters*, 91(6), 1–10. <https://doi.org/10.1785/0220200183>
- Kim, A., Dreger, D. S., & Larsen, S. (2010). Moderate earthquake ground-motion validation in the San Francisco bay Area. *Bulletin of the Seismological Society of America*, 100(2), 819–825. <https://doi.org/10.1785/0120090076>
- Komatitsch, D., Liu, Q., Tromp, J., Suss, P., Stidham, C., & Shaw, J. H. (2004). Simulations of ground motion in the Los Angeles basin based upon the spectral-element method. *Bulletin of the Seismological of America*, 94(1), 187–206. <https://doi.org/10.1785/0120030077>
- Komatitsch, D., & Tromp, J. (2002). Spectral-element simulations of global seismic wave propagation—I. Validation. *Geophysical Journal International*, 149, 390–412. <https://doi.org/10.1046/j.1365-246x.2002.01653.x>
- Komatitsch, D., & Vilotte, J. (1998). The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the Seismological of America*, 88(2), 368–392.
- Lee, E., Chen, P., & Jordan, T. H. (2014). Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes. *Seismological Research Letters*, 85(6), 1275–1284. <https://doi.org/10.1785/0220140093>
- Lee, E., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A., & Beroza, G. C. (2014). Full-3-D tomography for crustal structure in Southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research: Solid Earth*, 119(8), 6421–6451. <https://doi.org/10.1002/2014jb011346>
- Lin, F., Li, D., Clayton, R., & Hollis, D. (2013). High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array. *Geophysics*, 78(4), Q45–Q56. <https://doi.org/10.1190/geo2012-0453.1>
- Liu, G., Persaud, P., & Clayton, R. W. (2018). Structure of the northern Los Angeles basins revealed in teleseismic receiver functions from short-term nodal seismic arrays. *Seismological Research Letters*, 89(5), 1680–1689. <https://doi.org/10.1785/0220180071>
- Lovely, P., Shaw, J. H., Liu, Q., & Tromp, J. (2006). A structural Vp model of the Salton Trough, California, and its implications for seismic hazard. *Bulletin of the Seismological of America*, 96(5), 1882–1896. <https://doi.org/10.1785/0120050166>
- Lozos, J. C. (2016). A case for historic joint rupture of the San Andreas and San Jacinto faults. *Science Advances*, 2(3), e1500621. <https://doi.org/10.1126/sciadv.1500621>
- Maechling, P., Deelman, E., Zhao, L., Graves, R. W., Mehta, G., Gupta, N., et al. (2007). SCEC CyberShake workflows-Automating probabilistic seismic hazard analysis calculations. In I. J. Taylor, E. Deelman, D. B. Gannon, & M. Shields (Eds.), *Workflows for e-Science*: Springer.
- Magistrale, H., Day, S., Clayton, R. W., & Graves, R. W. (2000). The SCEC Southern California reference three-dimensional seismic velocity model version 2. *Bulletin of the Seismological Society of America*, 90(6B). <https://doi.org/10.1785/0120000510>
- Milner, K., Shaw, B. E., Goulet, C., Richards-Dinger, K., Callaghan, S., Jordan, H., et al. (2021). Toward physics-based nonergodic PSHA: A prototype fully deterministic seismic hazard model for southern California. *Bulletin of the Seismological of America*, 111(2), 1–18. <https://doi.org/10.1785/0120200216>
- Molinari, I., & Morelli, A. (2011). EPcrust: A reference crustal model for the European Plate. *Geophysical Journal International*, 185, 352–364. <https://doi.org/10.1111/j.1365-246x.2011.04940.x>

- Olsen, K. B., Day, S. M., & Bradley, C. R. (2003). Estimation of Q for long-period (>2 sec) waves in the Los Angeles basin. *Bulletin of the Seismological of America*, 93(2), 627–638. <https://doi.org/10.1785/0120020135>
- Olsen, K. B., Pechmann, J. C., & Schuster, G. (1995). Simulation of 3D elastic wave propagation in the Salt Lake basin. *Bulletin of the Seismological of America*, 85(6), 1688–1710.
- Park, J., Lindberg, C. R., & Vernon, F. L. (1987). Multitaper Spectral analysis of high-frequency seismograms. *Journal of Geophysical Research*, 92(B12), 12675–12684. <https://doi.org/10.1029/jb092ib12p12675>
- Parker, L. M., Thurber, C. H., Zeng, X., Li, P., Lord, N. E., Fratta, D., et al. (2018). Active-source seismic tomography at the Brady geothermal field, Nevada, with dense nodal and fiber-optic seismic arrays. *Seismological Research Letters*, 89(5), 1629–1640. <https://doi.org/10.1785/0220180085>
- Persaud, P., Ma, Y., Stock, J. M., Hole, J. A., Fuis, G. S., & Han, L. (2016). Fault zone characteristics and basin complexity in the southern Salton Trough, California. *Geology*, 44(9), 747–750. <https://doi.org/10.1130/g38033.1>
- Porter, K., Jones, L., Cox, D., Goltz, J., Hudnut, K., Mileti, D., et al. (2011). The ShakeOut Scenario: A hypothetical Mw7.8 earthquake on the Southern San Andreas fault. *Earthquake Spectra*, 27(2), 239–261. <https://doi.org/10.1193/1.3563624>
- Rial, J. A., Saltzman, N. G., & Ling, H. (1992). Earthquake-Induced resonance in sedimentary basins. *American Scientist*, 80(6), 566–578.
- Rodgers, A., Petersson, N. A., Nilsson, S., Sjogreen, B., & McCandless, K. (2008). Broadband waveform modeling of moderate earthquakes in the San Francisco Bay Area and preliminary assessment of the USGS 3D seismic velocity model. *Bulletin of the Seismological Society of America*, 98(2), 969–988. <https://doi.org/10.1785/0120060407>
- Rose, E. J., Fuis, G. S., Stock, J. M., Hole, J. A., Kell, A. M., Kent, G. M., et al. (2013). Borehole-explosion and air-gun data acquired in the 2011 Salton seismic imaging Project (SSIP), southern California: Description of the survey. *U.S.G.S. Open-file Report 2013-1172*.
- Savage, B., Komatitsch, D., & Tromp, J. (2010). Effects of 3D attenuation on seismic wave amplitude and phase measurements. *Bulletin of the Seismological Society of America*, 100(3), 1241–1251. <https://doi.org/10.1785/0120090263>
- Shaw, B. E., Milner, K. R., Field, E. H., Richards-Dinger, K., Gilchrist, J. J., Dieterich, J. H., & Jordan, T. H. (2018). A physics-based earthquake simulator replicates seismic hazard statistics across California. *Science Advances*, 4(8), 1–9. <https://doi.org/10.1126/sciadv.aau0688>
- Shaw, J. H., Plesch, A., Tape, C., Suess, M., Jordan, H., Ely, G., et al. (2015). Unified structural representation of the southern California crust and upper mantle. *Earth and Planetary Science Letters*, 415, 1–15. <https://doi.org/10.1016/j.epsl.2015.01.016>
- Small, P., Gill, D., Maechling, P., Taborda, R., Callaghan, S., Jordan, T., et al. (2017). The SCEC unified community velocity model software framework. *Seismological Research Letters*, 88(6), 1539–1552. <https://doi.org/10.1785/0220170082>
- Stacey, R. (1988). Improved transparent boundary formulations for the elastic-wave equation. *Bulletin of the Seismological of America*, 78(6), 2089–2097. <https://doi.org/10.1785/bssa0780062089>
- Suss, P., & Shaw, J. H. (2003). P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California. *Journal of Geophysical Research*, 108(B3), 1–18. <https://doi.org/10.1029/2001jb001628>
- Taborda, R., Azizzadeh-Roodpish, S., Khoshnevis, N., & Cheng, K. (2016). Evaluation of the southern California seismic velocity models through simulation of recorded events. *Geophysical Journal International*, 205, 1342–1364. <https://doi.org/10.1093/gji/ggw085>
- Taborda, R., & Bielak, J. (2014). Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake using different velocity models. *Bulletin of the Seismological of America*, 104(4), 1876–1898. <https://doi.org/10.1785/0120130266>
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2009). Adjoint tomography of the Southern California Crust. *Science*, 325(5943), 988–992. <https://doi.org/10.1126/science.1175298>
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180(1), 433–462. <https://doi.org/10.1111/j.1365-246X.2009.04429.x>
- Tu, T., Yu, H., Ramirez-Guzman, L., Bielak, J., Ghattas, O., Ma, K., & O'Hallaron, D. R. (2006). *From mesh generation to scientific visualization: An end-to-end approach to parallel supercomputing*. Paper presented at the IEEE conference on supercomputing. <https://doi.org/10.1109/sc.2006.32>
- Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, 20, 1–20. <https://doi.org/10.1029/2019gc008515>
- Yang, W., Hauksson, E., & Shearer, P. M. (2012). Computing a Large Refined Catalog of Focal Mechanisms for Southern California (1981–2010): Temporal Stability of the Style of Faulting. *Bulletin of the Seismological Society of America*, 102(3), 1179–1194. <https://doi.org/10.1785/0120110311>
- Zhong, M., & Zhan, Z. (2020). An array-based receiver function deconvolution method: Methodology and application. *Geophysical Journal International*, 222, 1–14. <https://doi.org/10.1093/gji/ggaa113>