Geophysical Journal International

Geophys. J. Int. (2016) **205,** 1342–1364
Advance Access publication 2016 March 7
GJI Seismology

doi: 10.1093/gji/ggw085

Evaluation of the southern California seismic velocity models through simulation of recorded events

Ricardo Taborda,[1,2] Shima Azizzadeh-Roodpish,[1,2] Naeem Khoshnevis[2] and Keli Cheng[3]

[1]*Department of Civil Engineering, University of Memphis, Memphis, TN* 38152*, USA. E-mail: ricardo.taborda@memphis.edu*
[2]*Center for Earthquake Research and Information, University of Memphis, Memphis, TN* 38152*, USA*
[3]*Department of Computer Science, University of Memphis, TN* 38152*, USA*

Accepted 2016 February 29. Received 2016 February 29; in original form 2015 December 4

SUMMARY

Significant effort has been devoted over the last two decades to the development of various seismic velocity models for the region of southern California, United States. These models are mostly used in forward wave propagation simulation studies, but also as base models for tomographic and source inversions. Two of these models, the community velocity models CVM-S and CVM-H, are among the most commonly used for this region. This includes two alternative variations to the original models, the recently released CVM-S4.26 which incorporates results from a sequence of tomographic inversions into CVM-S, and the user-controlled option of CVM-H to replace the near-surface profiles with a V_{S30}-based geotechnical model. Although either one of these models is regarded as acceptable by the modeling community, it is known that they have differences in their representation of the crustal structure and sedimentary deposits in the region, and thus can lead to different results in forward and inverse problems. In this paper, we evaluate the accuracy of these models when used to predict the ground motion in the greater Los Angeles region by means of an assessment of a collection of simulations of recent events. In total, we consider 30 moderate-magnitude earthquakes ($3.5 < M_{\mathrm{w}} < 5.5$) between 1998 and 2014, and compare synthetics with data recorded by seismic networks during these events. The simulations are done using a finite-element parallel code, with numerical models that satisfy a maximum frequency of 1 Hz and a minimum shear wave velocity of 200 m s^{-1}. The comparisons between data and synthetics are ranked quantitatively by means of a goodness-of-fit (GOF) criteria. We analyse the regional distribution of the GOF results for all events and all models, and draw conclusions from the results and how these correlate to the models. We find that, in light of our comparisons, the model CVM-S4.26 consistently yields better results.

Key words: Time-series analysis; Numerical solutions; Numerical approximations and analysis; Earthquake ground motions; Computational seismology; Wave propagation.

1 INTRODUCTION

Physics-based, deterministic earthquake ground motion simulations are now commonplace in studies of regional and global seismology, earthquake hazard modeling and in some areas of earthquake engineering (e.g. Aagaard *et al.* 2010; Tromp *et al.* 2010; Graves *et al.* 2011; Isbiliroglu *et al.* 2015). This is, in good measure, thanks to the continuous growth in capacity and availability of high-performance computer systems and applications (e.g. Cui *et al.* 2010); but it is especially true for regions where there are available models of the crustal and sedimentary basin structures. These models, often referred to as seismic velocity models, are a necessary ingredient to construct the three-dimensional (3-D) numerical representations used in simulation, thus their importance.

There exist a good number of seismic velocity models developed for different regions around the world such as southern and northern California in the United States (Kohler *et al.* 2003; Brocher *et al.* 2006), the Grenoble and Volvi valleys in Europe (Chaljub *et al.* 2010; Manakou *et al.* 2010) and Japan (Koketsu *et al.* 2008; Fujiwara *et al.* 2009). While some of these and other models of smaller regions have been built for specific research activities (e.g. Graves 2008; Maufroy *et al.* 2015), the last two decades have seen an increasing interest for developing models that can be applied more broadly, and that can be continuously updated by a community of users. These models are known as community velocity models, or

CVMs. The United States Geological Survey, for instance, recently released two such models, the Wasatch Front Community Velocity Model for the region of the Salt Lake basin and the Wasatch fault (Magistrale *et al.* 2006) and the Central United States Velocity Model for the region of the New Madrid seismic zone (Ramírez-Guzmán *et al.* 2012).

Among the various velocity models available in the U.S., the models developed and maintained by the Southern California Earthquake Center (SCEC), CVM-S and CVM-H, stand out as good examples of community models. These CVMs have evolved over time with contributions from a community of researchers who share an interest in studying the earthquake hazards and ground motion characteristics in southern California. CVM-S, also known as CVM-S4, was originally developed by Magistrale *et al.* (1996) and later updated by Magistrale *et al.* (2000) and Kohler *et al.* (2003). Recently, a new version of CVM-S, called CVM-S4.26 was built based on the original model CVM-S4 and the results of a sequence of 3-D full-waveform tomographic inversions done by Chen *et al.* (2007) and Lee *et al.* (2014b). CVM-H, on the other hand, was originally developed by Süss & Shaw (2003) and then periodically improved by Plesch *et al.* (2007, 2008, 2009, 2011). Some of the later additions to CVM-H include results from inversions done by Tape *et al.* (2009, 2010) and the inclusion of an optional geotechnical layer (GTL) model based on Ely *et al.* (2010).

Both the CVM-S and CVM-H families of models are regarded as acceptable representations of the crustal structure in southern California, at least at low frequencies ($f \leq 0.2$ Hz). Nonetheless, it is known that they have distinctions that can lead to significantly different results, even at these very low frequencies (e.g. Lee *et al.* 2014a; Taborda & Bielak 2014). Such differences, however, have never been systematically evaluated in forward simulation problems, or at frequencies beyond the upper limits set by the underlying inversions used to construct the models. Taborda & Bielak (2014), for instance, showed that the models CVM-S and CVM-H (with and without GTL) exhibited meaningful differences in the results of a series of 4-Hz simulations for the 2008 M_w 5.4 Chino Hills, California, earthquake. Using a goodness-of-fit (GOF) criteria, the authors compared synthetics to data and concluded that the choice of the velocity model had a significant effect on the outcome of the simulations. They indicated that, at least for the particular case of Chino Hills, CVM-S4 led to better fits than CVM-H, especially at frequencies below 1 Hz. In turn, Lee *et al.* (2014a) compared results obtained using CVM-S4, CVM-S4.26 and CVM-H for simulations of the 2014 M_w 5.1 La Habra and 2014 M_w 4.4 Encino earthquakes. In both cases, the authors showed that CVM-S4.26 led to better fits with seismograms bandpassed filtered between 0.02 and 0.2 Hz.

In this paper, we evaluate the southern California velocity models when used to predict the ground motion in the greater Los Angeles basin and its surrounding areas. The evaluation is carried out by means of validation of multiple simulations, through quantitative comparisons between synthetics and data, for a collection of recorded events. In total, we consider 30 earthquakes and use four different seismic velocity models. Two of the velocity models come from the CVM-S family. The original model CVM-S4 and the more recent model CVM-S4.26 (with geotechnical data recovered from CVM-S4). This latter model is also known as CVM-S4.26.M01, but we refer to it simply as CVM-S4.26. The other two models correspond to the CVM-H family, one with the GTL option inactive, to which we refer as CVM-H, and the second with the GTL option active, to which we refer to as CVM-H+GTL. All the earthquakes are of moderate size, with moment magnitudes (M_w) between 3.5 and 5.5, and occurred between 1998 and 2014. The epicentres of the events are spread throughout the region and within a simulation domain with a surface projection area of 180 km × 135 km. The simulations are performed using a finite-element application for solving forward wave propagation problems due to kinematic faulting (Tu *et al.* 2006; Taborda *et al.* 2010), with a numerical model built to represent a maximum frequency, $f_{\max} = 1$ Hz and a minimum shear wave velocity, $V_{S_{\min}} = 200$ m s^{-1}.

We based our evaluation of the different simulations on the GOF scores obtained following a modified version of the criteria introduced by Anderson (2004). Our results indicate that the simulations done with the models CVM-S4 and CVM-S4.26 consistently yield better fits with data than the synthetics obtained with CVM-H and CVM-H+GTL. Overall, the best GOF scores are obtained using CVM-S4.26. This indicates that previous observations made by Taborda & Bielak (2014) and Lee *et al.* (2014a) where not exclusive to the cases considered by these authors. More important, it indicates that the improvements done to CVM-S based on the tomographic studies done by Chen *et al.* (2007) and Lee *et al.* (2014a) have positive effects at frequencies above the limit considered for the inversions (0.2 Hz). In the sections ahead, we expand on this and other aspects, and investigate the main characteristics of the models leading to these conclusions.

2 SIMULATION DOMAIN AND VELOCITY MODELS

We consider a simulation volume domain of size 180 km × 135 km × 62 km and four seismic velocity models. The domain covers the entire Los Angeles metropolitan area and most of the significant geological structures in its vicinity. This includes the greater Los Angeles, Chino and San Bernardino basins; the San Fernando and Simi valleys the San Gabriel, Santa Ana and Santa Monica mountains; and part of the Mojave desert, the Santa Clara river valley and the Ventura basin (Fig. 1).

The elastic properties within the simulation domain are determined based on the models CVM-S4, CVM-S4.26, CVM-H and CVM-H+GTL to be evaluated. At any query point, each of these models provides the values of the particle's *P*- and *S*-wave velocities (V_P and V_S, respectively) and the material's density (ρ). CVM-S4 was initially developed by Magistrale *et al.* (1996), and later improved by Magistrale *et al.* (2000) and Kohler *et al.* (2003). In particular, we use version 4, release 11.11.0. This model integrates available information about the major southern California basins (Los Angeles basin, Ventura basin, San Gabriel Valley, San Fernando Valley, Chino basin, San Bernardino Valley and the Salton trough) using data from boreholes, oil-well samples, gravity observations and seismic refraction surveys. The model in itself is built upon empirical rules that use the depths and ages estimated for a set of geological horizons calibrated for southern California. Below and outside the basins, CVM-S4 relies on the 3-D seismic tomography model proposed by Hauksson (2000) and an upper-mantle model based on teleseismic inversions introduced into the model by Kohler *et al.* (2003).

The second model, CVM-S4.26, is a model recently developed by SCEC based on the results from a full 3-D tomographic (F3DT) inversion done by Lee *et al.* (2014b). In particular, the version of the model used here corresponds to the one which also recovers the geotechnical information in the original model CVM-S4. Such model has been also referred to as CVM-S4.26.M01. Here, we refer to it simply as CVM-S4.26. The raw version of this model is the result of an inversion process that involved a sequence of 26

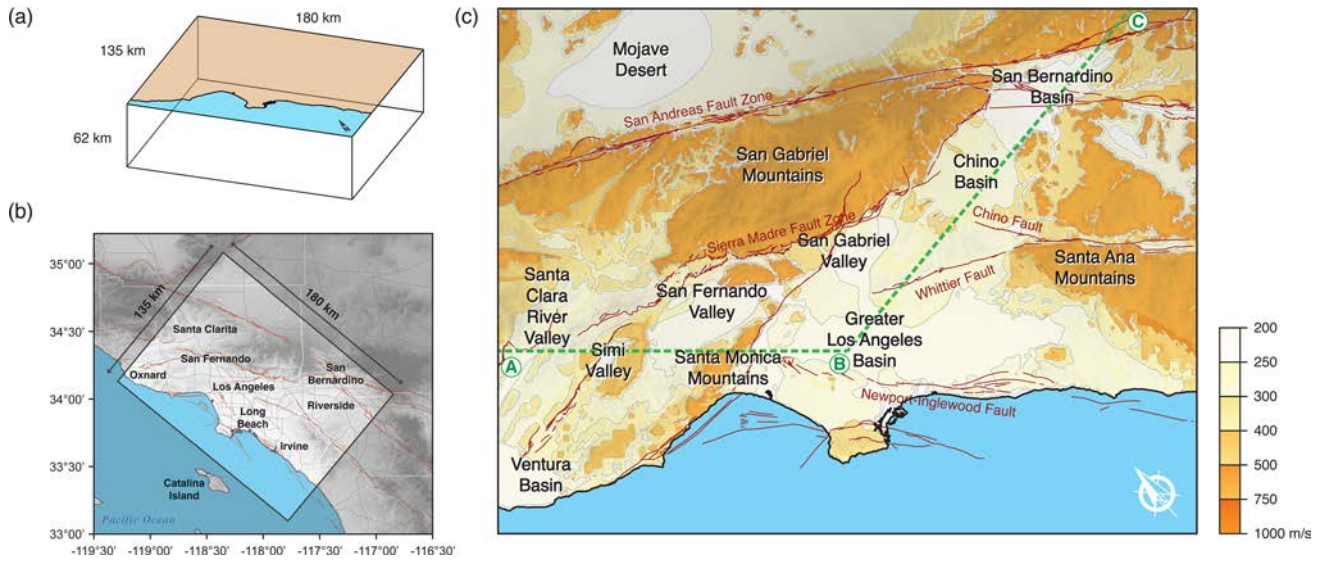

Figure 1. Region of interest and simulation domain. (a) 3-D view of the simulation domain. (b) Geographical location and surface projection of the simulation domain, along with the names of the main cities surrounding the Los Angeles metropolitan area. (c) Major geological structures including basins, valleys and mountains, along with the main quaternary faults in the region. The colour background represents the surface V_{S30} values included in the CVM-H+GTL model, with a topography shading effect. The segments $\overline{AB}$ and $\overline{BC}$ are used as reference for Fig. 3.

iterations over a reference model extracted from CVM-S4, thus its name. This effort followed the procedure first applied to the Los Angeles region by Chen *et al.* (2007). In Lee *et al.* (2014b), the reference model corresponded to a regular grid of 500-m spacing in which the material properties extracted from CVM-S4 were truncated to minimum values of $V_P = 2000$ m s^{-1}, $V_S = 1000$ m s^{-1} and $\rho = 2000$ kg m^{-3}. To compute the perturbations to the initial model, Lee *et al.* (2014b) used about 38 000 earthquake records and 12 000 ambient noise Green's functions, and combined two inversion methods, the adjoint-wavefield method (AW-F3DT; Tromp *et al.* 2005) and the scattering-integral method (SI-F3DT; Zhao *et al.* 2006). Each iteration in the procedure involved the computation of a forward simulation done using a staggered-grid finite-differences approach (Olsen 1994). The forward simulations were done for a maximum frequency, $f_{\max} = 0.2$ Hz; and the misfits were computed using seismograms bandpass filtered at 0.02–0.2 Hz. The inversion results were later merged with the original model CVM-S4 using an interpolation scheme designed to recover the geotechnical data (soft deposit profiles) present in the original model (see UCVM Developers 2013).

CVM-H was originally developed by Süss & Shaw (2003) and has since undergone multiple periodic updates (Plesch *et al.* 2007, 2008, 2009, 2011). Similarly to CVM-S4, CVM-H is built using the regional seismic traveltime tomography model of Hauksson (2000) in the background, but with differences in the structure of the basins. The major basins in CVM-H were defined using seismic reflection profiles and tens of thousands of borehole measurements (Süss & Shaw 2003). A particular aspect of CVM-H is that its structural representation is compatible with the geometry of major faults in southern California, as represented in the SCEC Community Fault Model (Plesch *et al.* 2007; Shaw *et al.* 2015). Besides the background tomography, the model also incorporates information from upper-mantle teleseismic and surface wave models extending to depths of 300 km (Prindle & Tanimoto 2006); and as done in the case of CVM-S4.26, CVM-H was improved using results from an AW-F3DT inversion process with 16 iterations (Tape *et al.* 2009, 2010). Here, we use CVM-H version 11.9.1. Although there is a more recent version 15.1.0, the changes in the latter pertain to structures out of the simulation domain. CVM-H also includes a GTL model. We, however, refer to CVM-H alone as the model with the GTL option inactive.

Last, we use the model CVM-H+GTL, which is the same as CVM-H but with the GTL option active. This model incorporates an algorithm proposed by Ely *et al.* (2010) to soften the material properties in the near-surface layers. The procedure uses the characteristic value of V_S in the top 30 m (V_{S30}) in order to modify the free-surface V_S and then interpolate the values of V_S, V_P and ρ with depth, until they match those of the original model at depth, $z = 350$ m. The values of V_{S30} are obtained from a geology-based map developed by Wills & Clahan (2006) for California, and the slope-dependent (topography-based) estimation proposed by Wald & Allen (2007) for points outside California. By default, the GTL model is active in CVM-H as distributed by SCEC. However, for evaluation purposes, we refer to it as if it were a different model (CVM-H+GTL). This is especially relevant because previous related work suggests that the approach used to implement the GTL model in CVM-H introduces changes in the shape of the basins and their velocity contrasts with the bedrock (Taborda & Bielak 2014).

In all cases, we constructed rasterized versions of the models for the volume simulation domain shown in Fig. 1a. This was done using the Unified Community Velocity Model (UCVM) software framework developed by SCEC (Small *et al.* 2011; Gill *et al.* 2015) and the UCVM implementation of the etree library (Tu *et al.* 2003), which follows a similar procedure to that described in Taborda *et al.* (2007) and Schlosser *et al.* (2008).

Fig. 2 shows a comparison between the four models for the free-surface V_S and the depths to the isosurfaces at which V_S values reach 1.0 and 2.5 km s^{-1}. Although the models have some general similarities, especially within each of the two families, the images in the figure clearly illustrate some of the more relevant contrasts. From the isosurfaces, for instance, it is seen that CVM-S4 and CVM-S4.26 are as similar near the surface ($z < 1$ km), as CVM-H and CVM-H+GTL are similar at depth ($z > 0.5$ km). This is consistent, on the one hand, with the construction of CVM-S4.26 as

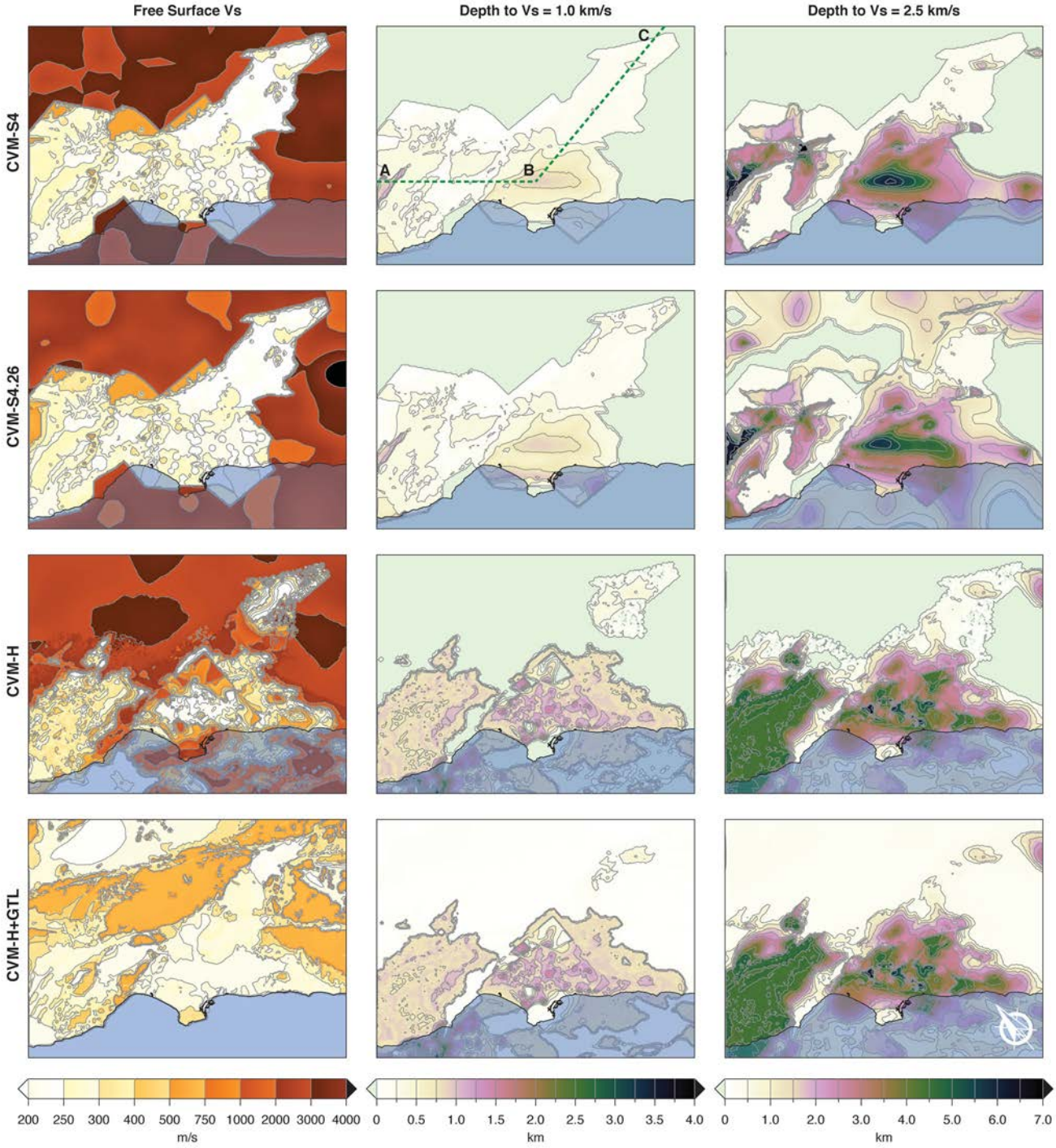

Figure 2. Comparison between the southern California community velocity models considered. Left: free-surface shear wave velocity, V_S. Centre: depth to the isosurface at which $V_S = 1.0$ km s^{-1}. Right: depth to the isosurface at which $V_S = 2.5$ km s^{-1}. The segments $\overline{\text{AB}}$ and $\overline{\text{BC}}$ in the top-centre frame are used as reference for Fig. 3.

this model was derived from an F3DT that had $V_{S_{\min}} = 1000$ m s^{-1}. Meaning that the changes near the surface are the result of the interpolation done when merging the inversion perturbations with the reference model, and not a direct consequence of the F3DT. On the other hand, we do not expect to see any difference between CVM-H and CVM-H+GTL beyond $z = 350$ m, as this is the depth at which the interpolation done when activating the GTL model matches the values of the original CVM-H model. Perhaps the larger contrasts between CVM-S4 and CVM-S4.26 are observed off shore and north of the San Andreas fault, in the Mojave desert (see Fig. 1 for reference). Other changes are observable in the Santa Clara river valley and Ventura basin, and in the vicinity of the San Bernardino basin, especially to the East (top-right corner of the simulation domain). In all these cases, CVM-S4.26 exhibits deeper structures than CVM-S4.

The strongest contrast between the CVM-H and CVM-H+GTL models with respect to CVM-S4 and CVM-S4.26 is in the San Fernando, Santa Clara river and Simi valleys and the Ventura basin.

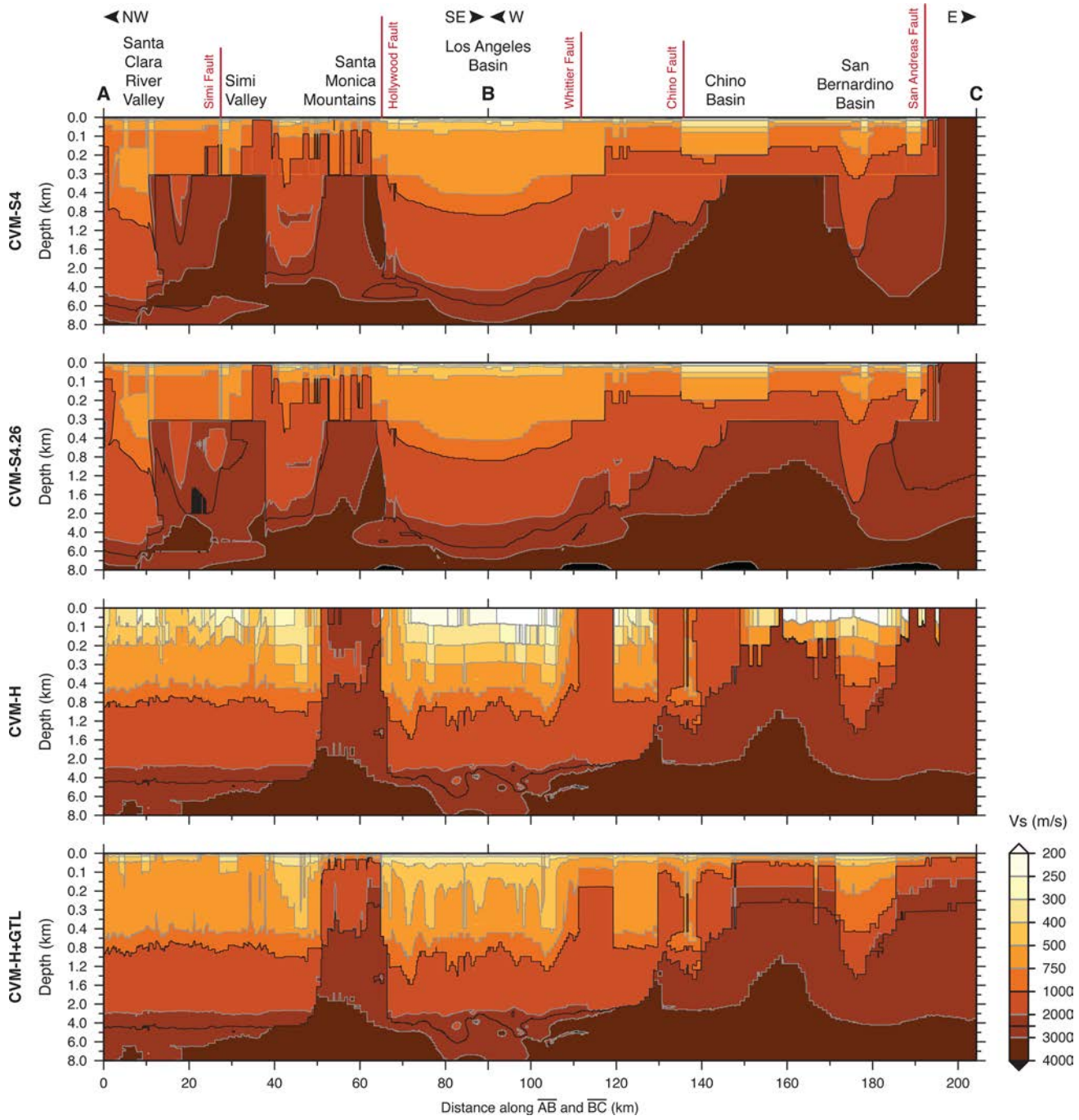

Figure 3. Vertical profiles of shear wave velocity (V_S) along the segments $\overline{AB}$ and $\overline{BC}$ shown in Figs 1 and 2 for each of the models considered. Markers and labels at the top indicate crossings of the profiles through significant geological structures and seismic faults. The vertical scales of the profiles are unevenly exaggerated in three segments from 0 to 0.4 km, 0.4 to 2 km and 2 to 8 km to highlight the differences between the models, especially near the surface.

Here, the structure in CVM-H is deeper and wider than in CVM-S4 and CVM-S4.26. Between CVM-H and CVM-H+GTL, the most relevant changes are along the edges of the basins and, in particular, in the shape and depth of the San Bernardino basin. These differences between the models are seen more clearly in Fig. 3, where we compare two contiguous vertical profiles (V_S) along the segments $\overline{AB}$ and $\overline{BC}$ shown in Figs 1 and 2. Note that most of the changes in CVM-S4.26 with respect to CVM-S4 are in the deeper structures, and to the northwest of the of segment $\overline{AB}$ and to the east of segment $\overline{BC}$. This figure also highlights the effects of the GTL model in CVM-H+GTL with respect to CVM-H, with the most relevant changes in the western section of the San Bernardino basin, the softer profiles introduced between the Chino basin and the Whittier fault, as well as the beneath the Santa Monica mountains and the reduction of the thickness of the deposits.

3 EVENTS AND SOURCE PROPERTIES

We consider 30 earthquakes in the region within the simulation domain. The selected events, scattered throughout, occurred between 1998 and 2014, and had magnitudes between 3.6 and 5.4,

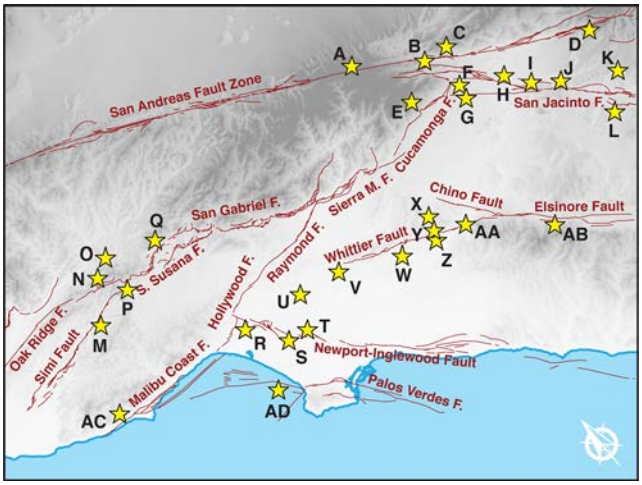

Figure 4. Location of the epicentres within the simulation domain for the 30 events considered here, along with some of the major quaternary faults in the region. The events are labeled with a sequential letter code. Details about the magnitude, date and focal mechanism are provided in Table 1.

and hypocentre depths that vary between 3.6 and 21.1 km. The largest events considered are the 2008 M_w 5.4 Chino Hills earthquake, followed by the 2014 M_w 5.1 La Habra earthquake, which are the strongest earthquakes registered in the region since the 1994 M_w 6.7 Northridge earthquake. Fig. 4 shows the location of the epicentres, which we labeled with a sequential letter code from A to Z and AA to AD. Also shown in the figure are the main seismic faults in the region. Table 1 provides detailed information about each of the selected events, including the event's ID (SCEDC 2013), magnitude, hypocentre coordinates and depth, focal mechanism (strike, dip and rake angles) and date and time (in coordinated universal time, or UTC). The coordinates, depth and focal mechanisms were provided by Lee (private communication, 2013), who inverted for the moment tensor using the approach developed by Lee *et al.* (2011). All the 30 events considered here were used in the F3DT inversions done by Lee *et al.* (2014b) towards the development of CVM-S4.26, and a subset of them (17, as indicated in Table 1) were also used by Tape *et al.* (2009) in the adjoint tomography study that led to updates in CVM-H.

As seen in Fig. 4, 12 of the events have epicentres near the northeastern portion of the simulation domain, to the east of the Los Angeles basin. Four of these events are located in the vicinity of the San Andreas fault zone (A–D), although not necessarily on the fault itself; one is on the Cucamonga fault (E) and seven in the vicinity of the San Jacinto fault zone (F–L). The following five events considered (M–Q) are located within the Simi, Santa Susana and San Gabriel faults, to the northwest of San Fernando valley. Next, beneath the Los Angeles basin, we consider one event at the junction between the Hollywood fault and the northern section of the Newport–Inglewood faults (R) and two more on the Newport–Inglewood fault (S, T); with the addition of one event right beneath the Los Angeles downtown area (U). To the east and southeast, six

Table 1. Selected events and description of source location, magnitude, focal mechanism, date and time and the number of stations used for validation. All the events listed here were used in the F3DT inversions done by Lee *et al.* (2014b) towards the development of CVM-S4.26. Those events identified with a star (*) were also used by Tape *et al.* (2009) in the adjoint tomography study that led to updates in CVM-H. The event ID numbers in this table correspond to those set by SCEDC (2013).

Code	Earthquake name	Event ID	M_w	Coordinates (lon., lat.)	Depth (km)	Strike/Dip/Rake	Date (yyyy/mm/dd)	UTC time (hh:mm:ss)	Number of stations
A*	Wrightwood	9064568	4.40	−117.6480, 34.3740	8.99	285/57/86	1998/08/20	23:49:58.198	17
B*	NW of Devore	10972299	3.79	−117.4642, 34.2655	10.91	98/58/68	2001/07/19	20:42:36.470	52
C	NNE of Devore	14494128	3.72	−117.3838, 34.2587	7.18	344/69/-33	2009/08/01	12:55:55.317	77
D*	Yucaipa	14155260	4.88	−117.0113, 34.0580	11.61	75/59/55	2005/06/16	20:53:26.225	172
E	N of Rancho Cucamonga	10216101	3.60	−117.5762, 34.2058	4.92	54/69/16	2006/11/04	19:43:44.376	55
F*	2002 Fontana	13692644	3.74	−117.4288, 34.1613	6.54	233/72/-28	2002/07/25	00:43:14.872	55
G*	2005 Fontana	14116972	4.42	−117.4387, 34.1250	4.15	222/88/-25	2005/01/06	14:35:27.593	83
H*	San Bernardino	10370141	4.45	−117.3042, 34.1073	14.22	87/70/28	2009/01/09	03:49:46.051	159
I*	N of Loma Linda	9140050	4.37	−117.2525, 34.0500	15.36	270/90/-6	2000/02/21	13:49:43.017	38
J	Redlands	10541957	4.10	−117.1797, 34.0045	8.53	33/46/-68	2010/02/13	21:39:06.349	97
K	2010 Beaumont	10530013	4.28	−117.0232, 33.9322	13.93	234/89/9	2010/01/16	12:03:25.345	76
L*	2006 Beaumont	14239184	3.90	−117.1122, 33.8560	11.53	45/31/-25	2006/07/10	02:54:43.809	66
M*	Simi Valley	14000376	3.59	−118.7530, 34.2722	13.81	234/62/60	2003/10/29	23:44:48.206	54
N*	WSW of Valencia	9753489	3.90	−118.6678, 34.3705	14.21	83/62/57	2002/01/29	06:00:39.140	52
O*	N of Pico Canyon	9096972	3.98	−118.6090, 34.3980	11.53	287/55/54	1999/07/22	09:57:23.502	26
P	Chatsworth	14312160	4.66	−118.6195, 34.2995	7.58	82/27/51	2007/08/09	07:58:48.888	109
Q	Newhall	15237281	3.86	−118.4580, 34.3508	3.59	236/58/33	2012/10/28	15:24:23.172	120
R*	Beverly Hills	9703873	4.24	−118.3885, 34.0590	7.90	262/81/4	2001/09/09	23:59:17.695	130
S	Inglewood Area	10410337	4.70	−118.3357, 33.9377	13.86	243/60/25	2009/05/18	03:39:36.126	213
T*	NW of Compton	9716853	3.98	−118.2702, 33.9290	21.13	116/68/71	2001/10/28	16:27:45.388	55
U*	Downtown Los Angeles	9093975	3.77	−118.2180, 34.0100	9.53	125/49/79	1999/06/29	12:55:00.371	25
V	Whittier Narrows	14601172	4.44	−118.0817, 33.9923	18.85	282/36/73	2010/03/16	11:04:00.026	180
W	La Habra	15481673	5.10	−117.9300, 33.9220	5.00	239/70/38	2014/03/29	04:09:42.970	311
X*	Chino Hills	14383980	5.39	−117.7613, 33.9530	14.70	47/51/32	2008/07/29	18:42:15.960	335
Y*	2002 Yorba Linda	9818433	4.75	−117.7758, 33.9173	12.92	34/84/-10	2002/09/03	07:08:51.675	67
Z	2009 Yorba Linda	10399889	3.98	−117.7892, 33.8940	4.23	208/65/26	2009/04/24	03:27:49.840	91
AA*	ESE of Yorba Linda	9644101	3.64	−117.6882, 33.8777	3.59	56/65/37	2001/04/13	11:50:11.916	53
AB	Lake Elsinore	10275733	4.73	−117.4770, 33.7322	12.60	65/59/58	2007/09/02	17:29:14.827	116
AC	Westlake Village	10403777	4.42	−118.8825, 34.0667	14.17	254/73/30	2009/05/02	01:11:13.084	94
AD	Hermosa Beach	14738436	3.69	−118.4578, 33.8572	11.23	57/41/54	2010/06/07	23:59:27.165	93

events have epicentres located along the Whittier and Chino faults (V–AA) and one on the Elsinore fault (AB). Last, we included two earthquakes in the northwestern sections of the Malibu Coast fault and the Palos Verdes fault (AC and AD, respectively).

We modeled each event considering a point source analogy with rupture parameters scaled according to the magnitude of each earthquake. First, we computed the rupture dimensions using the general expressions proposed by Wells & Coppersmith (1994) for all types of faults to obtain the subsurface rupture length (l_s) and downdip rupture width (w_d). Then, following the analogy of a circular fault area, and assuming that in the vicinity of a point source the material is homogeneous, we estimated the rise time to be

$$t_r = \frac{16}{7}\frac{l_s f^{0.5}}{V_S \pi^{1.5}}, \tag{1}$$

where f is a shape factor equal to the ratio w_d/l_s, such that $A = f l_s^2$ (e.g. Stein & Wysession 2003; Shearer 2009).

Having estimated the rise time, we used the formulation proposed by Tinti *et al.* (2005) to express the source slip-rate function, $\dot{s}(t)$, as

$$\dot{s}(t) = D \int_{-\infty}^{\infty} W(t-\tau) Y(\tau) d\tau. \tag{2}$$

Here, $W(t)$ and $Y(t)$ are a triangular (smoothing) function and the Yoffe (1951) function, respectively. D is the final (maximum) slip, which we obtained based on the seismic moment of each event, $M_o = \mu DA$. The values of the shear modulus, μ, were extracted from the velocity models ($\mu = \rho V_S^2$) at the corresponding hypocentral locations. In turn, the rise time associated with the Yoffe function, τ_r, was estimated based on the fact that from eq. (2) we know that the total rise time of the slip-rate function is

$$t_r = \tau_r + 2\tau_s \tag{3}$$

where τ_s is the duration of the triangular function and t_r corresponds to the value we obtain from eq. (1). For this, in all cases, we assumed a value of $\tau_s = 0.1 t_r$, which is consistent with some of the tests done by Tinti *et al.* (2005). The resulting source time functions are shown in Fig. 5.

While we recognize that the procedure just described carries in it significant simplifications—for instance, the database used by Wells & Coppersmith (1994) was limited to earthquakes with a lower bound of M 4.7—we found the slip-rate functions and rise times we obtained to be acceptable. For a handful of events, we tested other alternatives and observed the results did not changed significantly, or were better using the chosen procedure. This is due in part to the fact that we are only considering $f_{\max} = 1$ Hz. As it can be seen in Fig. 5, with the exception of the Chino Hills and La Habra earthquakes, the frequency content of the slip-rate functions is nearly flat below 1 Hz, thus any change to the rise time or the shape of the slip functions could only have a marginal effect on the results for earthquakes of the magnitudes considered here.

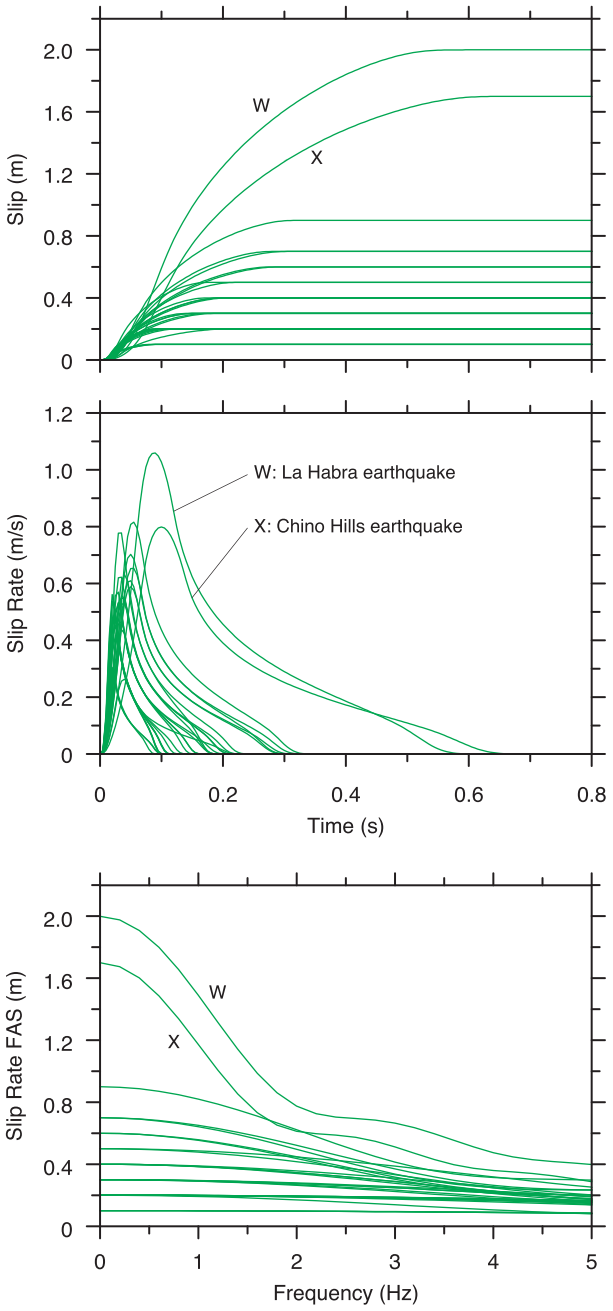

Figure 5. Source time functions (top: slip; middle: slip-rate) and slip-rate Fourier amplitude spectra (bottom) of all the events. Slip-rate functions were computed based on the total rise time estimated from eqs (1) and (2). The source functions of events W and X, corresponding to the 2014 La Habra and 2008 Chino Hills earthquakes, are singled out in the figure because for reference. These two events are the largest of all earthquakes considered.

4 SIMULATION METHOD AND PARAMETERS

We simulate the ground motion for each earthquake and velocity model combination using a 3-D finite-element approach to solve the anelastic wave equation. Similar numerical methods including finite differences, finite elements and spectral elements have been successfully used in equivalent problems in the past (e.g. Graves 1996; Bao *et al.* 1998; Komatitsch & Vilotte 1998). In particular, we use a parallel code called Hercules, which implements an octree datastructure for representing unstructured hexahedral meshes in memory (Tu *et al.* 2006). The solution approach used by Hercules relies on a standard Galerkin method for discretizing the equations of elastodynamics in space, and advances explicitly in time to obtain the next-step state of nodal displacements. The time integration scheme uses first-order backward and second-order central differences to approximate the velocity and acceleration, respectively (Taborda *et al.* 2010). Attenuation is introduced by means of a viscoelastic model composed of a set of parallel Maxwell elements

and a Voigt element (Bielak *et al.* 2011), and the absorbing boundaries are modeled using a plane-wave approximation (Lysmer & Kuhlemeyer 1969). Hercules has been used in multiple verification and validation exercises before, and has shown to be a reliable tool for 3-D ground motion simulation (e.g. Bielak *et al.* 2010; Taborda & Bielak 2013).

It is important to note that none of the velocity models provides information about the quality factors Q_P and Q_S, which are necessary for modeling the effects of intrinsic attenuation. In this regard, for the case of Q_S, we adopt the empirical rule

$$Q_S = 10.5 - 16V_S + 153V_S^2 - 103V_S^3 + 34.7V_S^4 - 5.29V_S^5 + 0.31V_S^6 \quad (4)$$

used in Taborda & Bielak (2013, 2014). This rule is an extension of that proposed by Brocher (2005, 2008), which was in turn inspired by other similar but simpler approximations used in the past (e.g. Olsen *et al.* 2003; Graves 2008). On the other hand, for the case of Q_P, we use

$$Q_P = \frac{3}{4}\left(V_P/V_S\right)^2 Q_S, \quad (5)$$

which is derived from the special case in which one considers no attenuation due to dilatational deformation, that is, $Q_\kappa \to \infty$ (e.g. Stein & Wysession 2003; Shearer 2009). In both cases, we consider the values of Q to remain constant within the range of frequencies considered.

In Hercules, the finite-element meshes are built at run-time, and the variable size of the elements is set such that it satisfies the rule

$$e \leq \frac{V_S}{f_{\max} p}, \quad (6)$$

where p is the number of points per wavelength. We set $p = 8$ as a minimum requirement, but due to the octree structure of the mesh, for most elements with properties transitioning from one element size to the next, the effective number of points per wavelength varies between 8 and 15.

Table 2 describes the main simulation parameters. All simulations were done on Blue Waters at the National Center for Supercomputing Applications. Table 2 also includes information about the average performance of the code for every simulation model. Note that the influence of the models CVM-H and CVM-H+GTL having larger areas/volumes of softer deposits is reflected by the number of elements, and consequently, by the time necessary to run each simulation. All simulations considered, we used over 0.27 million of CPU hours.

Table 2. Simulation parameters and numerical model details.

Domain size	
Length (km)	180
Width (km)	135
Depth (km)	61.875
Domain corners[a]	
Southwest	−119.2888°, 34.120549°
Northwest	−118.3540°, 35.061096°
Northeast	−116.8460°, 34.025873°
Southeast	−117.7809°, 33.096503°
Numerical parameters	
$f_{\max}$ (Hz)	1.0
$V_{S_{\min}}$ (m s^{-1})	200
Points/wavelength	8–15
Simulation Δt (s)	0.005
Simulation time (s)	100
Number of steps	20 000
Mesh details	
Minimum element size (m)	21.97 m
Maximum element size (m)	351.56 m
Number elements (millions)	
CVM-S4	106.4
CVM-S4.26	110.3
CVM-H	249.7
CVM-H+GTL	282.6
Average simulation wall-clock time (hr:mm:ss)[b]	
CVM-S4	1:03:46
CVM-S4.26	1:03:58
CVM-H	2:17:24
CVM-H+GTL	2:40:51

[a]The corners of the domains are given in longitude and latitude.
[b]Corresponding to 1280 cores on NCSA's Blue Waters.

5 GROUND MOTION SIMULATION RESULTS

Before we address the evaluation of the models, we present results from the simulations and offer a general perspective on the ground motion characteristics obtained for the events considered. Fig. 6 shows the peak horizontal magnitude of velocity on the free surface for all events, and for the particular case of simulations done using the model CVM-S4.26. This figure illustrates how the basins in the region respond to earthquakes originating at various locations within the simulation domain. Although interpretations in this regard are biased by the choice of the colour scale, it is fair to say that earthquakes of magnitude less than 3.9 remain local, showing only a marginal ground response in areas outside their immediate epicentral surrounding (e.g. events L, M, U). Events of magnitude greater than 4.3, on the other hand, show stronger response all throughout the domain, and exhibit more clearly the effects of the basins (e.g. events D, P, Y). Events with magnitudes between 3.9 and 4.3 are in a transition zone. In such cases, the shallower events register stronger ground motions (e.g. events Q and Z).

All events considered, the largest ground motions are obtained for the 2014 M_w 5.10 La Habra and 2008 M_w 5.39 Chino Hills earthquakes (events W and X, respectively), and the areas with most significant shaking are the greater Los Angeles basin, the San Bernardino basin and the region between Simi valley and the Ventura basin. While Fig. 6 only includes results obtained using CVM-S4.26, these observations are consistent across velocity models. We, however, now focus our attention to the discrepancies observed when using different velocity models.

Fig. 7 shows the peak horizontal magnitude of velocity on the free surface for all velocity models, for the particular cases of the 2005 M_w 4.42 Fontana, 2007 M_w 4.66 Chatsworth and 2014 M_w 5.10 La Habra earthquakes (events G, P and W, respectively). These three events were selected because their locations sample different paths into the various basins and they are also large enough to generate significant motions throughout most of the simulation domain. The Fontana (G) earthquake epicentre was located in the northern section of the San Jacinto fault zone, not far from the junction with the San Andreas fault zone and the Cucamonga fault (see Fig. 4). This earthquake shows clear influence in the area of the San Bernardino basin for all four models, although with some differences. In the case of CVM-S4, for instance, a significant level of the response is channeled into the Chino and the greater Los Angeles basins, as well as far into the Simi and the Santa Clara

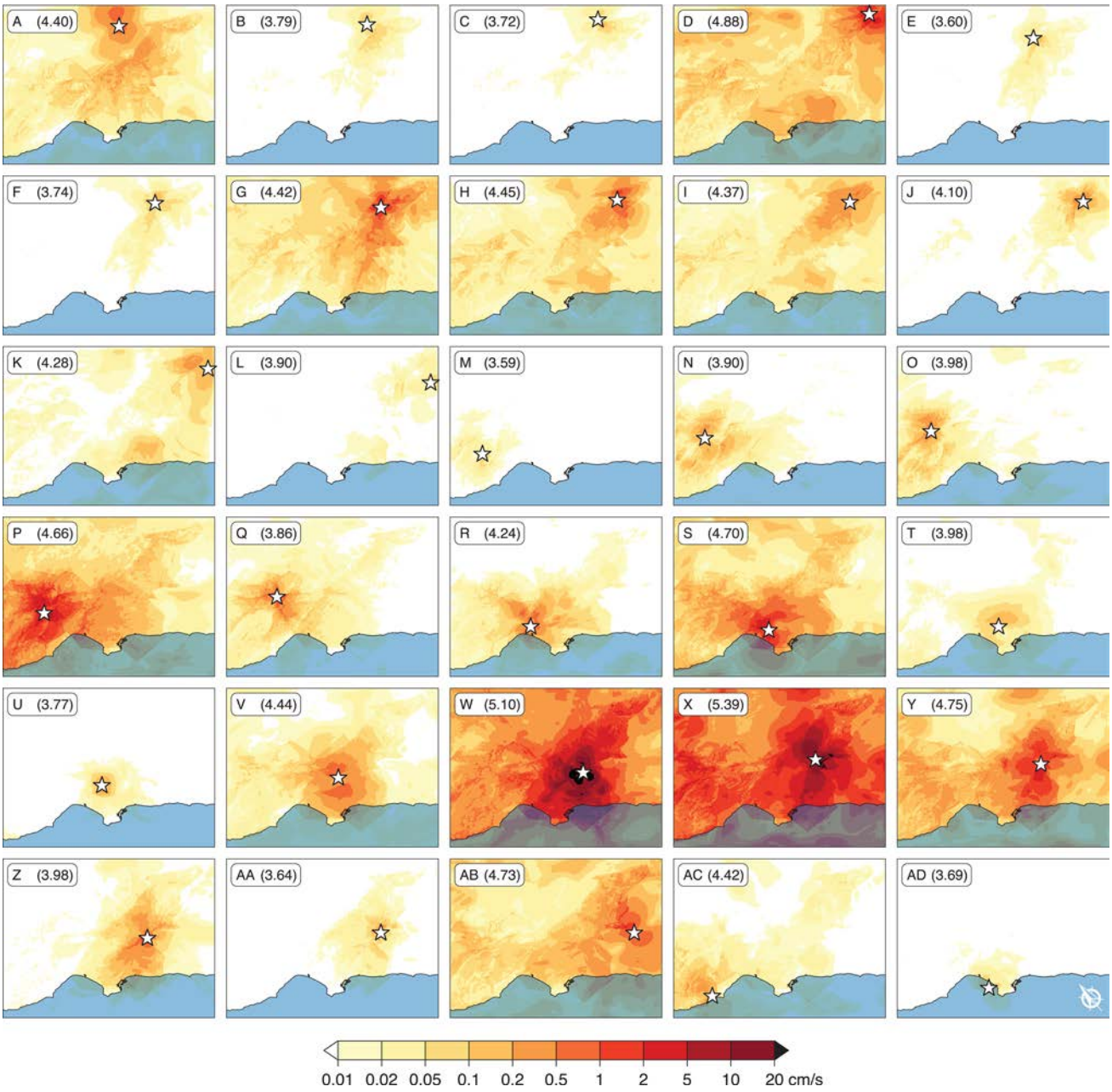

Figure 6. Free-surface peak horizontal magnitude velocity from simulations for all the events using the velocity model CVM-S4.26. The letter code used to identify each earthquake is shown at the top-left corner along with the event's magnitude, M_w (see Table 1). In each case, the star indicates the epicentre of the event (see also Fig. 4). Although smaller and larger values than those shown in the colour scale were obtained from the simulations, these were truncated for visual convenience.

river valleys. A similar response is seen in the case of CVM-S4.26, but to a lesser extent, with only marginal response in the Santa Monica area. CVM-S4.26 also exhibits lower values southwest and northwest from the epicentre than CVM-S4. In the cases of CVM-H and CVM-H+GTL, the ground motion is more localized around the epicentral area. Both models, however, seem to channel more energy along the north flank of the San Andreas fault and to the southeast and southwest, in particular along the west edge of the Santa Ana mountains, where both CVM-H and CVM-H+GTL have deeper structures than CVM-S4 or CVM-S4.26 (see Fig. 2 for reference).

In the case of the Chatsworth earthquake, the strongest response concentrates in the Simi and San Fernando valleys, along the Santa Clara river valley and into the Ventura basin. For the simulations done with CVM-S4 and CVM-S4.26, however, a greater amount of energy is channeled west into the Ventura basin and southeast towards the greater Los Angeles area. Both CVM-S4 and CVM-H show some level of basin effects in San Bernardino, an area where CVM-H+GTL shows the least level of amplification of all the models, due to the changes introduced by the GTL model as highlighted in Fig. 3. Both CVM-H and CVM-H+GTL show a stronger contrast between the Simi and San Fernando valleys and the Los Angeles basin, marked by the influence of the Santa Monica mountains, which seem to be more sharply defined in these two models than in CVM-S4 and CVM-S4.26 (see Fig. 2).

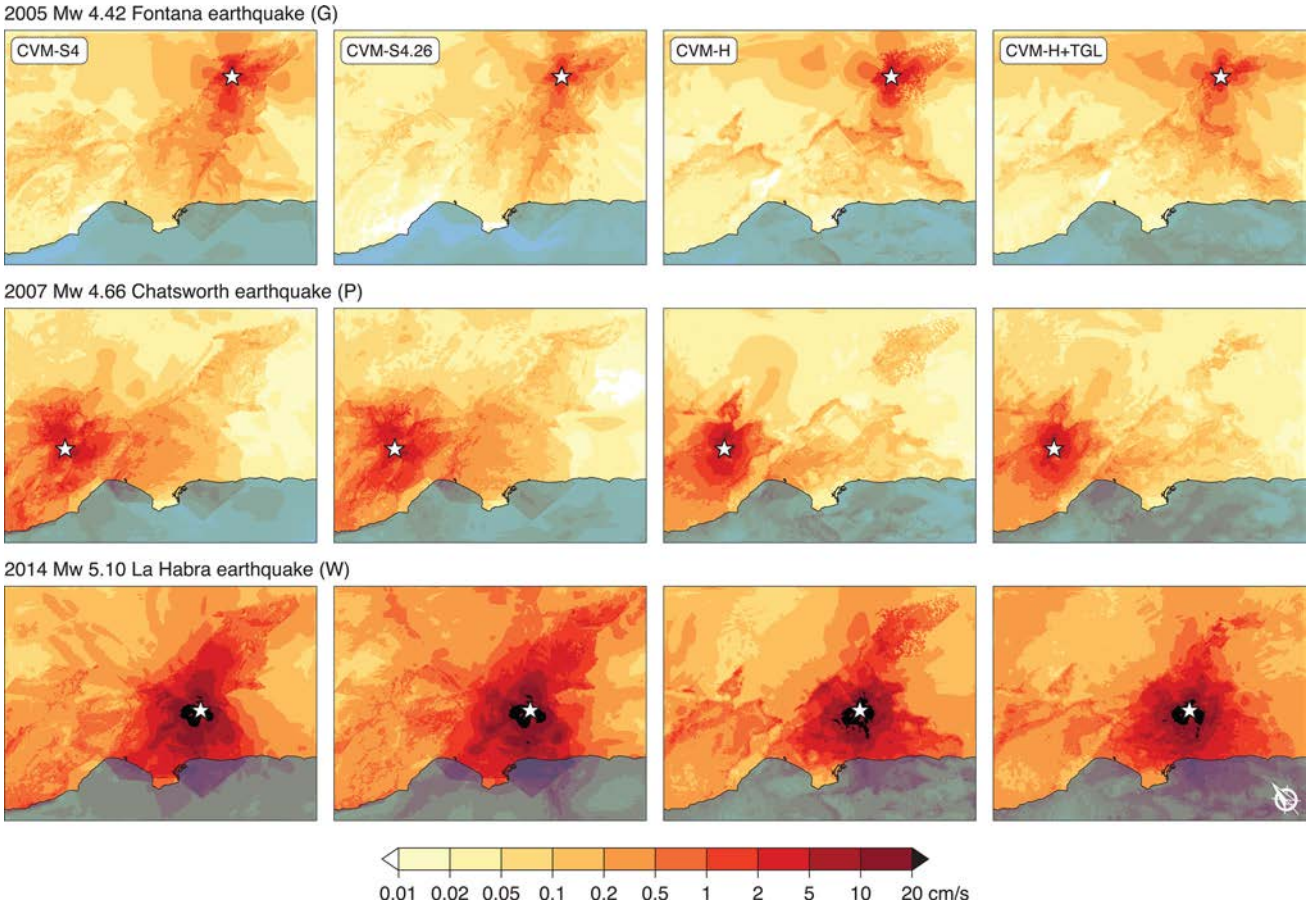

Figure 7. Free-surface peak horizontal magnitude velocity for three representative events (from top to bottom: G, P and W) using all four velocity models (from left to right: CVM-S4, CVM-S4.26, CVM-H and CVM-H+GTL). The stars indicate the epicentre locations (see Table 1 and Fig. 4 for reference). In each case, smaller and larger values than those shown in the contour maps were obtained, but truncated for visual convenience.

Last, in the case of the La Habra earthquake, the ground motions are mostly concentrated in the greater Los Angeles basin, though with some significant differences among the models. We first note again the fact that the model CVM-H+GTL introduces strong changes in the response of the San Bernardino basin with respect to CVM-H. As in other cases, CVM-H and CVM-H+GTL yield larger ground motions in the area of Irving (see Fig. 1) southeast from the epicentre. CVM-S4 and CVM-S4.26, on the other hand, exhibit larger ground motions within the Los Angeles basin itself and in the Chino basin. In turn, CVM-S4.26 yields larger shaking near the San Gabriel valley and mountains, and beyond in the Mojave desert. CVM-S4.26 also exhibits stronger response in the area of the Santa Ana mountains, as a result of the contrast in the model in this area with respect to CVM-S4. Of all four models, CVM-S4 has stronger response in the Ventura basin and CVM-H along the Santa Clara river valley—which likely reflects a better coupling between the crustal structure and fault representation in the model, considering the weak zone along the Santa Clara river due to the presence of the Oak Ridge fault beneath it.

Although not in equal measure for all events, the differences just highlighted were common among the models in the simulation results of other earthquakes. We further analyse the implications of these model discrepancies and their consequence in simulation results through the comparison of synthetics against data in the following sections.

6 DATA PROCESSING AND VALIDATION METHOD

We evaluate the accuracy of the simulations, and thus that of the velocity models, based on a quantitative validation of the simulated ground motions. The validation process consists of comparisons between synthetics and data, at locations where records were available for the simulated events. For this, we compiled a large collection of broad-band and strong-motion records from two data centres, the Southern California Earthquake Data Center (SCEDC) and the Center for Engineering Strong Motion Data (CESMD). SCEDC and CESMD archive records from various seismic networks in the southern California region. In total, we obtained records from more than 800 stations spread throughout the simulation domain (see top map in Fig. 8). Unfortunately, not all the stations recorded all the events, thus the number of available data points for comparisons varies between events. In addition, some stations were discarded for reasons explained below. The total number of stations with records used for validation, for each event, is shown in Table 1.

Records from SDEDC and CESMD were processed and selected as follows. For each event, we first downloaded all the stations that recorded the earthquake and identified those that fell within the simulation domain boundaries. From this initial set, we kept only free-surface stations with records in three orthogonal components, two horizontal and one vertical. Although the majority of stations have instruments oriented as positive in the North (NS), East (EW)

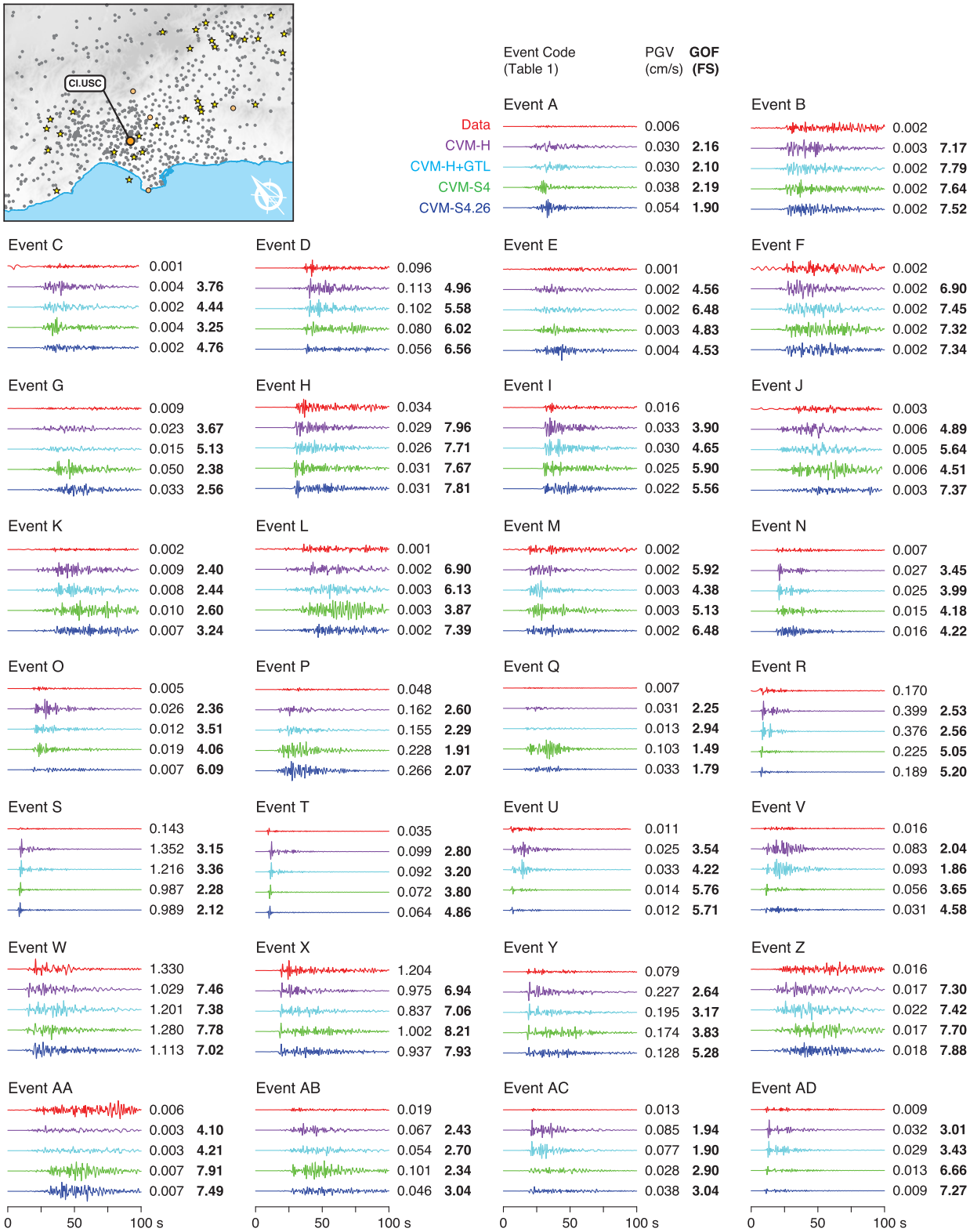

Figure 8. Comparison of ground motion data and synthetics for all the events at the CI.USC, in the NS component of motion. For reference, the event codes are shown at the top of each set of signals (see Table 1). The signal at the top corresponds to the data, followed by synthetics generated using the CVM-H, CVM-H+GTL, CVM-S4 and CVM-S4.26 velocity models, respectively. The numbers to the right of each signal are the peak velocity (first column) and for the synthetics, the average FS from eq. (8) (second column, in bold). In the map, grey dots indicate the locations of the complete collection of stations (842 total) used for validation, with those in common among all events highlighted in larger coloured circles. Stars indicate the location of the epicentres (see also Fig. 4).

and up directions (UD), we rotated or sign-flipped those signals that had different orientations in order to bring them all to a common standard. We also rotated the synthetics to be oriented in the NS, EW and UD directions. The selection of only free-surface stations means that we discarded the channels at depth in geotechnical arrays, as well as the channels that are part of structural monitoring arrays. In the case of strong ground motion records from CESMD, we gave priority to records available in their raw V1 format. For these records, whose original data correspond to accelerations, we performed gain and baseline corrections, and applied a high-pass filter at 0.05 Hz before integrating to obtain velocities and displacements. In the rare occasions, when V1 records were not available but processed V2 formats were, we used the latter. In the case of records downloaded from SCEDC, we used both strong-motion acceleration (HN) channels and broad-band velocity (BH) channels. In general, we gave priority to HN channels, which we processed similarly to the V1 records from CESMD, checking the base line, then high-pass filtering, and finally integrating to obtain velocity and displacement records. We used BH channels only when HN were not available and the BH channels were far enough from the epicentres to avoid saturation. In such cases, we differentiate to obtain acceleration, and high-pass filter and integrate the records to obtain displacements.

Once we processed all data and synthetics, we proceeded with the validation of the simulations for each event, and each velocity model. The validation process was carried out using the GOF method proposed by Anderson (2004), with minor modifications introduced by Taborda & Bielak (2013). The method compares synthetics against data using 11 individual parameters, namely: Arias intensity integral (C1), energy integral (C2), Arias intensity value (C3), total energy (C4), peak acceleration (C5), peak velocity (C6), peak displacement (C7), response spectrum (C8), Fourier amplitude spectrum (C9), cross-correlation (C10) and strong-phase duration (C11). Each parameter is mapped onto a numerical scale ranging from 0 to 10, where a score of 10 corresponds to a perfect match between two signals. The values obtained for a given pair of signals for each of the 11 scores are combined using the expression:

$$\mathrm{S} = \frac{1}{9}\left(\frac{\mathrm{C1}+\mathrm{C2}}{2} + \frac{\mathrm{C3}+\mathrm{C4}}{2} + \sum_{i=5}^{11} \mathrm{C}i\right) . \tag{7}$$

Following the guidelines suggested by Anderson (2004), this scoring procedure is applied to each pair of data and synthetics using compatible 'broad-band' sets, and a series of bandpass filtered versions of the signals or subbands, SB_1, SB_2 and SB_3. The results of S from eq. (7) for the broad band (BB) and each of the subbands (SB_i) are then combined to obtain a final score (FS), define as:

$$\mathrm{FS} = \frac{1}{4}\left(\mathrm{BB} + \sum_{i=1}^{3} \mathrm{SB}_i\right) . \tag{8}$$

We used signals bandpass filtered between 0.1 and 1.0 Hz for BB, and between 0.1 and 0.25 Hz, 0.25 and 0.5 Hz and 0.5 and 1 Hz for SB_1, SB_2 and SB_3, respectively. The upper frequency limit of 1 Hz was based on the maximum frequency of the simulations. The lower limit of 0.1 Hz, on the other hand, was determined by the need for minimizing instrumental and numerical processing issues at the low frequencies (0–0.1 Hz), which can be introduced when processing strong-motion and BB records to obtain velocities and displacements. This also helps eliminate permanent displacements in the synthetics extracted at locations near the epicentre, especially for shallower earthquakes.

Before running the comparison process, signals are also (sub)sampled to have the same time-step size, Δt. We chose a common Δt equal to 0.1 s, and used a decimation corner (low-pass) frequency of 2 Hz for the records. This corresponds to a Nyquist frequency of 5 Hz, or five times the simulation maximum frequency. In addition, both data and synthetics were synchronized using the earthquake times (shown in Table 1) as reference. Synthetics were synchronized assuming that the earthquake time corresponds to the instant at which the source slip-time function is at half the total rise time. Data were synchronized using the time stamp on the record. We cut or zero-padded the records at the beginning depending on whether the station channel started recording before or after the earthquake time. When zero-padding, a tapper filter in time was applied at the beginning of the record. Once the signals have been synchronized in terms of start time, they are also matched to have the same length, using the shortest of the two to determine the length used for comparisons. Applying these filters equally to data and synthetics, using a common Δt, and synchronizing the start and end time of each pair of signals provides a maximum level of consistency in both the frequency and time domains, which minimizes the numerical discrepancies that could arise from the comparisons performed using the different metrics C1 through C11.

Additional details about the original parameters proposed by Anderson (2004) and the modifications introduced to the scoring criterion are given in Taborda & Bielak (2013). Taborda & Bielak (2013, 2014) also include brief discussions regarding the choice of the method proposed by Anderson (2004), in light of other available procedures such as those introduced by Kristeková *et al.* (2006, 2009) or Olsen & Mayhew (2010). In summary, previous experience in verification and validation efforts tells us that the method of Anderson (2004) works better for validation, as opposed to verification, mainly because it is not restricted to matching waveforms, but is rather oriented at conveying information of physical meaning to both seismologists and engineers. We also prefer Anderson (2004) because it provides additional consistency with respect to previous work done by Taborda & Bielak (2013, 2014).

7 VALIDATION RESULTS

The most basic form of validation is to perform a visual comparison of synthetics against recorded seismograms. Fig. 8, for instance, presents a graphical validation of synthetics against data (NS component) for all the events at station CI.USC. This station is located near the downtown area of Los Angeles, and is one of five stations for which there are records available for all the earthquakes considered in this study. While these types of comparisons are useful to corroborate that the general parameters of the simulation are correct (i.e. time synchronization of synthetics with data and amplitude order), they are of limited use when it comes to quantifying the overall validity of the results. This is better accomplished through quantitative comparisons as those given by the evaluation method described in the previous section. As an example, in Fig. 8 we include the GOF scores (FS values) for each of the synthetics obtained at the CI.USC station. This parallel between qualitative and quantitative comparisons shows the correlation that exists between what can be visually considered a good match and the higher FS values (e.g. events B, H, W, Z), or vice versa (e.g. events A, P, Q). This has been illustrated before in other studies (e.g. Taborda & Bielak 2013), which gives us confidence in the use of the chosen evaluation approach.

A comparison like the one shown in Fig. 8 for every set of signals, components and events at each station would be, however,

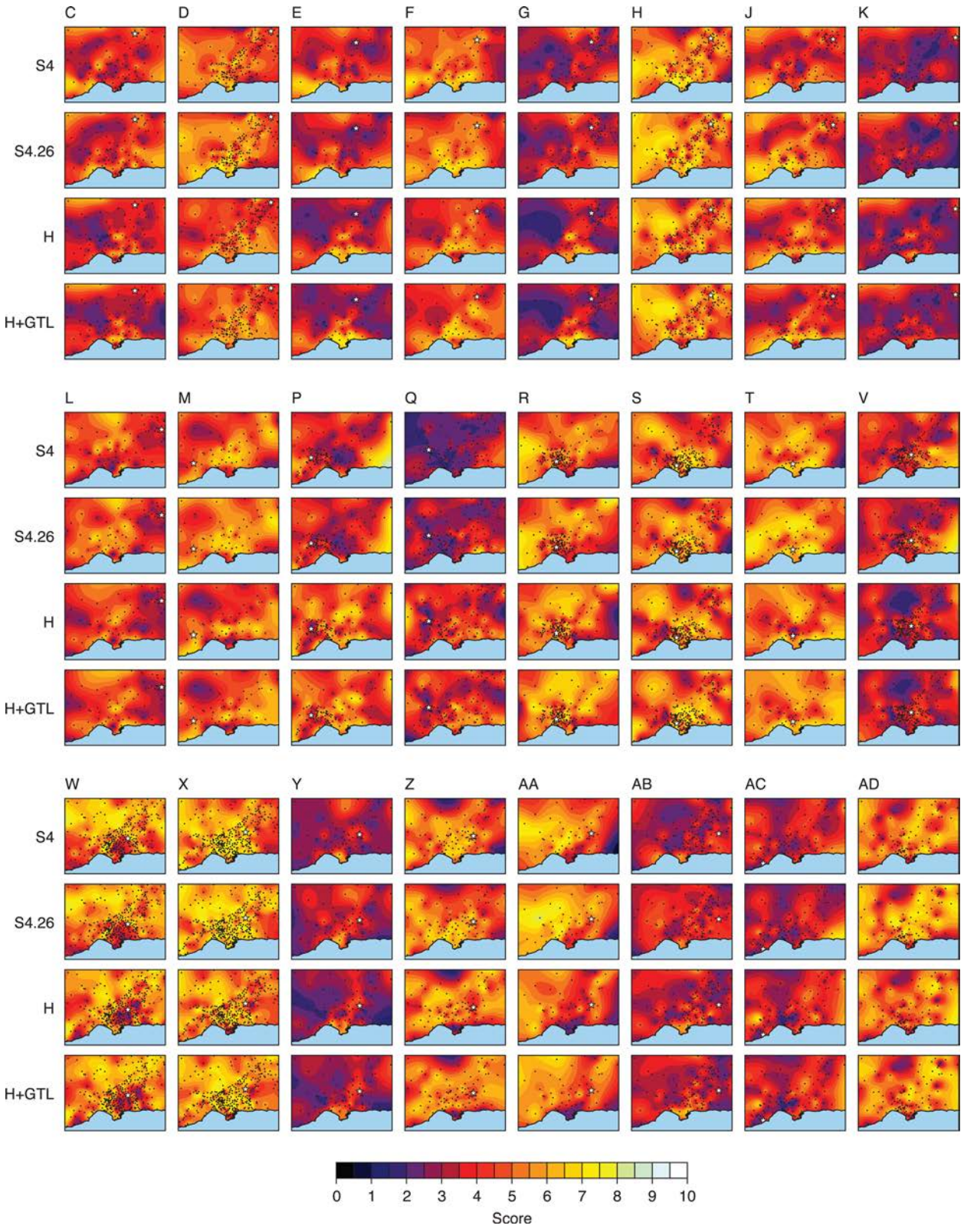

Figure 9. GOF maps for all events with 53 or more stations used for validation. Contours indicate the score obtained by averaging the FS values for all three components of motion (EW, NS and UD). Dots correspond to the location of stations and stars indicate the epicentres for each event. Event labels are placed at the top of each set of four maps corresponding to the results obtained using the CVM-S4, CVM-S4.26, CVM-H and CVM-H+GTL, as indicated with labels on the left margin. The selection of events (which corresponds to 24 out of 30 and leaves out events A, B, I, N, O, U) was done to facilitate the arrangement of the figure.

unpractical. Therefore, we condense the results for all stations in GOF maps for all events and velocity models. Fig. 9 shows the results for 24 of the 30 events considered, for the four velocity models. The selection of a subset was made based on the arrangement of the figure and the events left out were those with the least number of station records available for validation (A, B, I, N, O, U). Those included in the figure have 53 or more stations (see Table 1). The contour surfaces in this figure are built based on the FS values computed at each station (see eqs 7 and 8) after being averaged for the three components of motion. While we recognize that the spatial interpolation of the coloured surface in this and other similar subsequent figures is artificial (in the sense that the stations are not evenly spaced throughout the simulation domain), we have opted for this format because it facilitates the interpretation of results and offers consistency with previous studies (e.g. Taborda & Bielak 2013, 2014). The colour scale here is such that darker patterns indicate a poorer match (FS $<$ 4) while lighter colours are indicative of better fits (FS $>$ 6). According to Anderson (2004), a score above 8 can be considered an excellent fit.

This figure illustrates the wide range of results. There are events with GOF maps strongly dominated by acceptable fits (FS $>$ 5) and scattered very good fits (FS $>$ 7), while others are strongly dominated by poor fits (FS $<$ 4) and scattered bad fits (FS $<$ 3). This can be appreciated more clearly by examining the distribution of FS values within the collection of stations associated with each event. Fig. 10 shows this for a selection of five events: X, W, S, V and D. These are the events with the largest number of stations available for validation. This figure shows the same GOF maps included in Fig. 9 (for the events selected), but adds at the bottom of each map a histogram of the scores in intervals of 0.5 points on the 0–10 scoring scale. Ideally, one would like the histograms to have a non-Gaussian negatively skewed distribution with a median above 8. We are therefore interested in the models and simulations that show a tendency toward the right-hand side of the scale, which correspond to light-coloured GOF maps (i.e. good fits). Although the validation results shown here are far from this ideal, some interesting trends can still be observed.

Note, for instance, that the simulation histograms of events X and S with CVM-S4.26, and event W with CVM-H+GTL show negatively skewed distributions with median values near 7. The histograms obtained for the simulations done with other models for these same events, however, tend to show lower median values and normal-shape distributions. In general, most histograms—those shown in Fig. 10 as well as those of all other events and models omitted here for brevity—tend to have a normal distribution with median values between 3.5 and 6. This is discussed in greater detail in the next section.

In addition to the overall results, we are also interested in the validation of the simulations on different frequency bands. In two previous studies on simulations of the 2008 M_w 5.4 Chino Hills earthquake, Taborda & Bielak (2013, 2014) showed that simulations produced results closer to observations in the lower frequency bands, and that GOF scores tend to decline for the higher frequency bands. Taborda & Bielak (2013), however, only considered the model CVM-S4, whereas Taborda & Bielak (2014) considered the models CVM-S4, CVM-H and CVM-H+GTL but did not included CVM-S4.26 because this model was released after that study was concluded. To illustrate the analysis by frequency bands, Fig. 11 shows the GOF maps corresponding to the Chino Hills earthquake (event X). These results are equivalent to those presented by Taborda & Bielak (2014) but here, for uniformity in the comparisons across all events, we use a point source model—as opposed to the extended source model used by Taborda & Bielak (2014).

The histograms in Fig. 11 show that the number of stations with low GOF scores tend to increase as one moves from the lowest frequency band (SB_1) to the higher ones (SB_2 and SB_3). Subsequently, while the histograms in SB_1 tend to be narrower and have medians in the upper half of the validation scale, the histograms in SB_2 and SB_3 are more spread over the scoring scale and have lower medians, especially in SB_3. While the results shown in this figure correspond only to event X, they are fairly representative of most other events. The following section explores these and other statistical characteristics of the validation results for all events and models.

8 VELOCITY MODELS EVALUATION RESULTS

The main purpose of the validation process conducted on the simulations is the final evaluation of the velocity models available for southern California. To that end, we further condense the validation results to facilitate their analysis. Fig. 12 shows the average GOF scores for all the events, for each velocity model. Here, the average scores are obtained as the simple arithmetic mean of the GOF scores (FS values) of all the stations for each event. In the figure, we highlight the average scores corresponding to the models CVM-S4.26 and CVM-H+GTL. As it can be seen in the figure, out of the four models, CVM-S4.26 yields the highest scores in most of the events. In all, in 20 of the 30 events considered here, the top GOF average was obtained with this model, followed by CVM-H+GTL, which yielded the top values in 7 of the remaining 10 events, and CVM-S4 dominating the last 3. Also in this figure, we include a table on the margin with single-value scores for the velocity models. These are obtained by averaging the mean scores of all events for each model. Although once averaged the differences between the values become marginal, they are consistent with the individual values. That is, the top score (4.87) corresponds to CVM-S4.26, followed by CVM-H+GTL (4.82), CVM-S4 (4.75) and CVM-H (4.60).

These observations remain consistent if the equivalent analysis is done on the GOF scores obtained for the BB and the different frequency bands analyses. This is shown in Fig. 13. Here, the results also indicate that the model CVM-S4.26 leads to the top scores in all frequency bands for a majority of the events (i.e. in at least 17 out of the 30). We also highlight the fact that the results in the BB analysis are fairly representative of the overall results shown in Fig. 12, and that the relative average values are fairly consistent among the different bands. That is, the shape of the plots is about the same and the most significant change is simply the fact that the lower frequency band (SB_1) has higher scores than those of the upper bands. Similarly, in terms of the single-value score for each model, the top values correspond to CVM-S4.26, with the highest mean score of 5.11 obtained for the SB_1 band (0.1–0.25 Hz).

Mean values, however, are not always representative of the distribution of scores across stations for each event, and they do not allow to have a sense of the shape of the associated histograms derived from the GOF maps. Since showing the histograms for all cases would not be practical, we look directly at the mode, median and standard deviation in a condensed form. Fig. 14 shows these statistical properties for the validations done for all events and all models. Here, the coloured circles are centred at the value of the mode (measured as the middle point in the bin with the top count

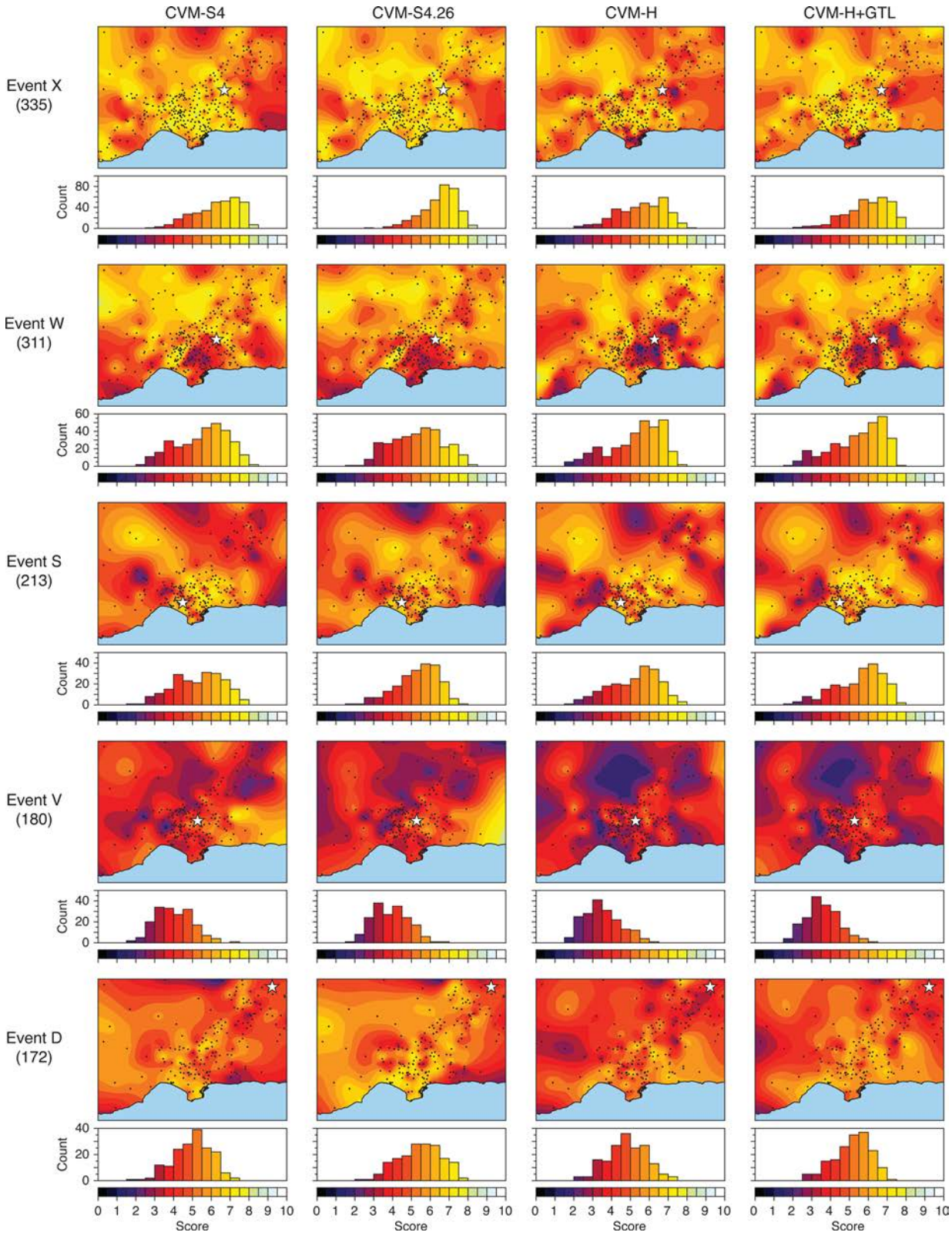

Figure 10. GOF maps and histograms for the top five events with the largest number of stations used for validation, arranged top to bottom. GOF maps are the same as shown in Fig. 9, but displayed horizontally for CVM-S4, CVM-S4.26, CVM-H and CVM-H+GTL from left to right, respectively. The histograms beneath each map show the distribution of the FS values averaged for all three components of motion in intervals widths of 0.5 points. The number of stations is shown on the left, in parenthesis, below each event label.

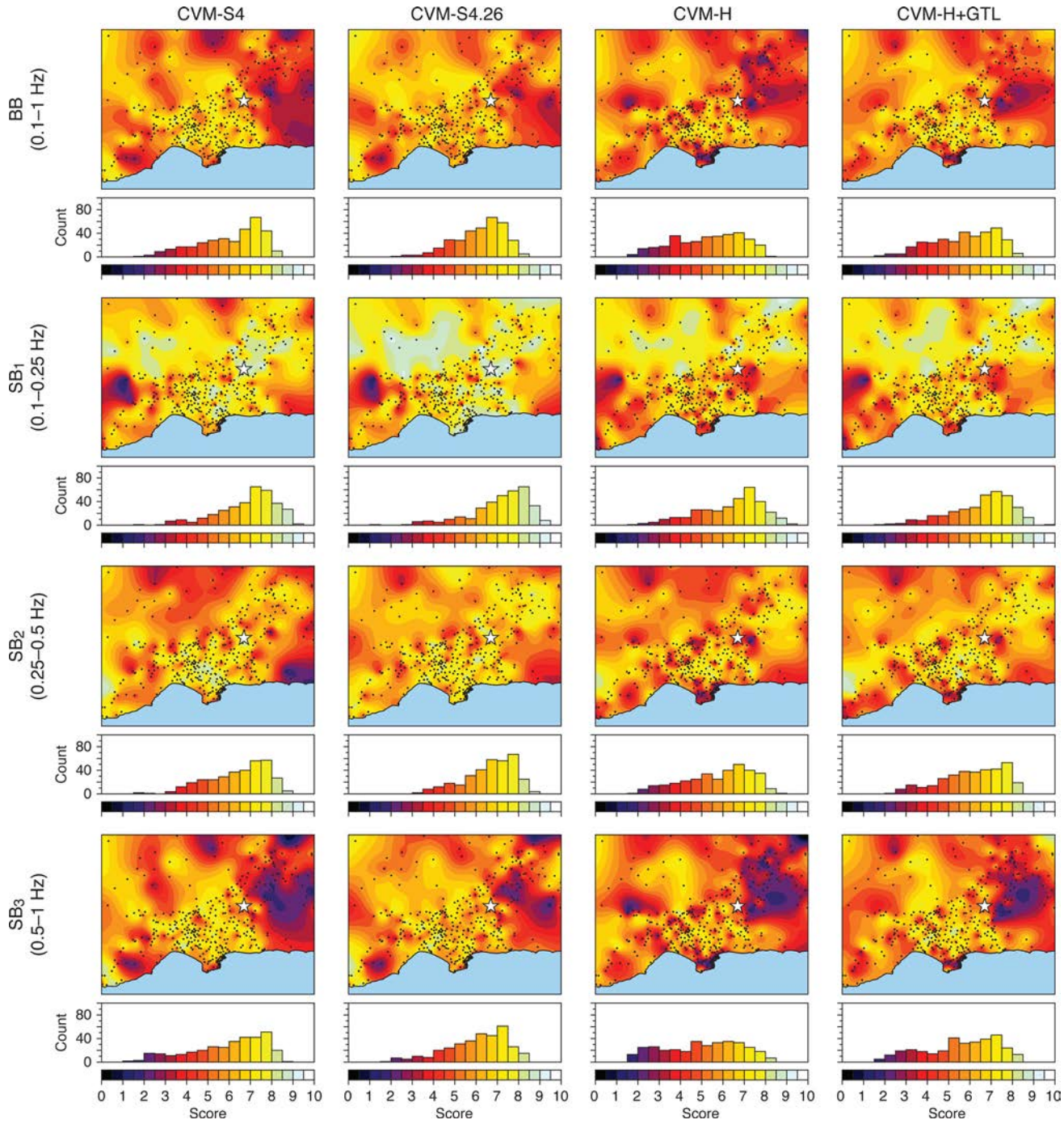

Figure 11. GOF maps and histograms for Event X (2008 M_w 5.4 Chino Hills earthquake) classified by frequency bands, BB, SB_1, SB_2 and SB_3 (top to bottom, respectively), and all velocity models (left to right). Here, GOF values correspond to scores obtained with eq. (7), averaged for all three components of motion. Histograms show the distribution of values in intervals of 0.5 points.

of stations). The size of the circles corresponds to the percentage of stations associated with the mode value (i.e. the number of stations in the tallest column of the histogram normalized by the total number of stations). In turn, the colour of the circles indicates the value of standard deviation of the distribution, and the black small dots linked by a dashed line correspond to the median of the histograms for each event and model. The plots in this figure also include a reference line at a GOF value of 5.

In simple terms, we would like to obtain large circles, filled with dark colour, centred above the black dots and in the upper half of the scoring scale (above the reference line). This would be indicative of negatively skewed histograms with dominant good fits between synthetics and data, or in other words, narrow histograms tilted to the right (or upper half) of the scoring scale. Conversely, circles that are centred below the black dots would indicate that the histograms are positively skewed, which is undesirable because such distributions are the result of bad fits. In turn, circles that coincide with the black dots are representative of normal-shape distributions. These are neither desirable nor undesirable, but are definitely preferable if centred at values above 5.

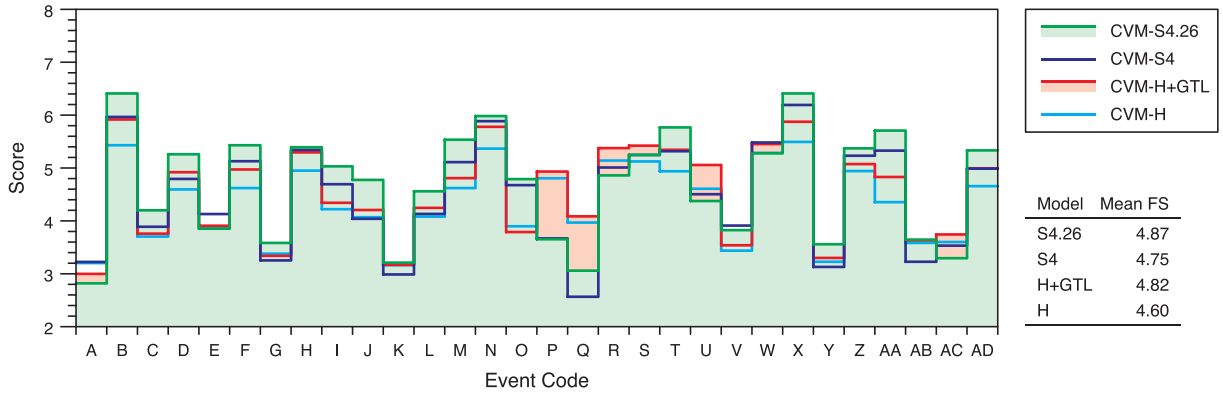

Figure 12. Summary of GOF results for all events. Values correspond to the average FS (eq. 8) of all stations, in all components and frequency bands, discretized by velocity models. The values in the table to the right correspond to the mean value of the score for each velocity model considering all events. The areas under the lines for the values obtained with simulations using the velocity models CVM-S4.26 and CVM-H+GTL are filled to highlight that CVM-S4.26 yields the largest number of events with the top value (20 out of 30 events), followed by CVM-H+GTL with seven events.

Fig. 14 shows these results for both the final combined scores (eq. 8) and the scores corresponding to the lowest frequency band (SB_1). In both cases, the model CVM-S4.26 shows, again, the best overall performance with about 15 events with circles (mode values) and dots (median values) above 5, for both the combined and the low-frequency analyses. The relative positions of the mode and median values indicate that histograms tend to be of normal shape or negatively skewed. We note, however, that a good number of events (about 10) tend to fall on the lower half of the scoring scale and have a positively skewed histogram (i.e. circle mode values below the black dot median values). See also that in the low-frequency band analysis, the results indicate that the models CVM-S4 and CVM-S4.26 tend to yield histograms with narrower distributions, as indicated by the darker colours of the circles (or lower values of standard deviation). This is desirable because it is indicative of uniformity in the validation of an event with these modes. This also means that, if anything in the simulation could be improved (e.g. source model and attenuation parameters) then one should expect all GOF values to improve.

We also look at how the collective results of the validations work together for each model. Fig. 15 shows GOF maps similar to those presented in the previous section, but in this case we average the results of the top 20 events with the largest number of available records (validation stations) for each station. Of course, not every station recorded the top 20 events, so the average value obtained at each location combines only the scores from the simulations within the subset that recorded any given event. This is done for each velocity model. The figure includes these maps and their corresponding histograms for the cases of the final score combining all frequency bands and components (the set at the top) and the results for the lowest frequency band SB_1 (the set at the bottom). As it can be seen, once these scores are averaged, the histograms tend to have a normal distribution. Mode values are always centred at about 4.75 and 5.25, and mean and median values vary between 4.83 and 5.34 and 4.86 and 5.23, respectively. The proximity between these values confirms their normal distribution. The GOF maps in this figure are also contrasted with the regional basin structures as represented in each model. The white contour lines on the maps represent the edges of the basins or large deposit areas as extracted from Fig. 2. Although it is difficult to draw definite conclusions from these figures, it does appear that better scores are obtained inside the basin/deposit areas than outside, with the exception of some of the stations in the Mojave desert.

To provide additional context to these results, Fig. 16 shows the standard deviation values corresponding to the averages shown in Fig. 15 for each station. As done before, the results in this figure include the analysis for each velocity model and for the scores combining all frequency bands and all components (top set), and those of the lowest frequency band SB_1 (bottom set). Values of standard deviation oscillate, for the most part, between (score) values of 0.5–2.5 points, with normal distributions centred around 1.5 and 2. Large values of standard deviation are only present in isolated locations for some but not all the models. We note here that the model CVM-S4.26 in the FS values shows the sharpest concentration about 1.25–1.5 points, which we interpret as a positive sign of the consistency in the results obtained with this model.

We further searched for correlations between the GOF scores and the models' properties or events' characteristics. In terms of the events, in general, the validation scores do not correlate with parameters such as hypocentral depth, epicentral location, or earthquake magnitude. We did find, however, that the events that were better recorded within the region of interest (i.e. with larger numbers of stations used for validation comparisons) yielded somewhat better average scores. This is shown in Fig. 17. In particular, Fig. 17(a) shows the correlation between the average combined final score and the number of stations for each event and velocity model. In turn, Fig. 17(b) shows the same correlations, but this time for the average score obtained at the lowest frequency band analysis. In each of these plots, we include linear regressions that show there are similar trends independently of the models. In addition, we looked at the material properties in the models themselves. We explored correlations between near-surface shear wave velocity parameters (e.g. V_{S30}) but found no clear relationship, with the arguable exception of basin depth. Fig. 17(a) shows the relationship between all the stations average GOF scores for all events and the depth to $V_S = 2.5$ km s^{-1}. Here, the data seem to suggest a tendency to higher values with increasing depth, but the data are spread over a wide range of score values. However, once the averages are taken only at those stations that registered five or more events, the relationship is somewhat clearer, as shown in Fig. 17(d).

A point of interest in the trend lines in Fig. 17(d) is the margin of separation between the regressions for the models

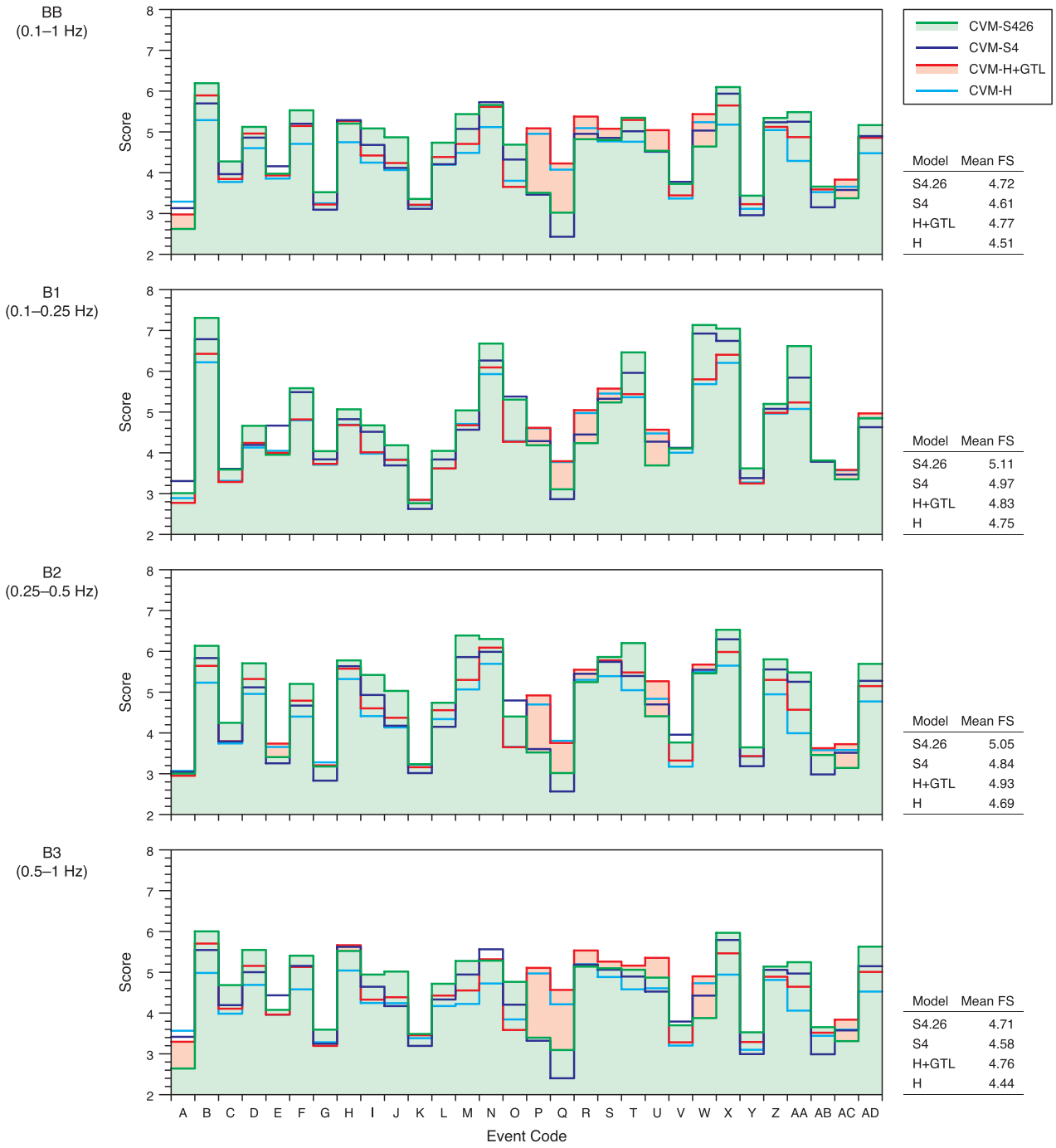

Figure 13. Summary of GOF results for all events, discretized by frequency bands and velocity models. Values correspond to the average FS scores of all stations, in all components, in the respective frequency band and velocity model. The areas under the lines for the values obtained with simulations using the velocity models CVM-S4.26 and CVM-H+GTL are filled to highlight that these models consistently yielded better scores, with CVM-S4.26 leading in the majority of events for all frequency bands.

CVM-H+GTL and CVM-H. This is indicative of the validation improvement yielded by the effect of the GTL model. While we noted that the GTL model may artificially change some of the structures as currently implemented in the CVM-H model, this reflects the positive effect that adding the GTL model has, regardless of these artefacts. This, in turn, indicates that a model such as CVM-S4.26 would benefit even more from incorporating a more comprehensive GTL model, similar to that used in CVM-H+GTL. This is an area of current work in progress.

9 CONCLUSIONS

This paper presents a comprehensive validation study of the simulation of 30 events recorded in recent years in the greater Los Angeles metropolitan area, using four different seismic velocity models currently available for southern California. The main objective of this effort was to evaluate the velocity models based on the GOF obtained between the synthetics and recorded data from strong-motion and seismic stations, using a standardized approach to model a large ensemble of earthquakes and validate the results.

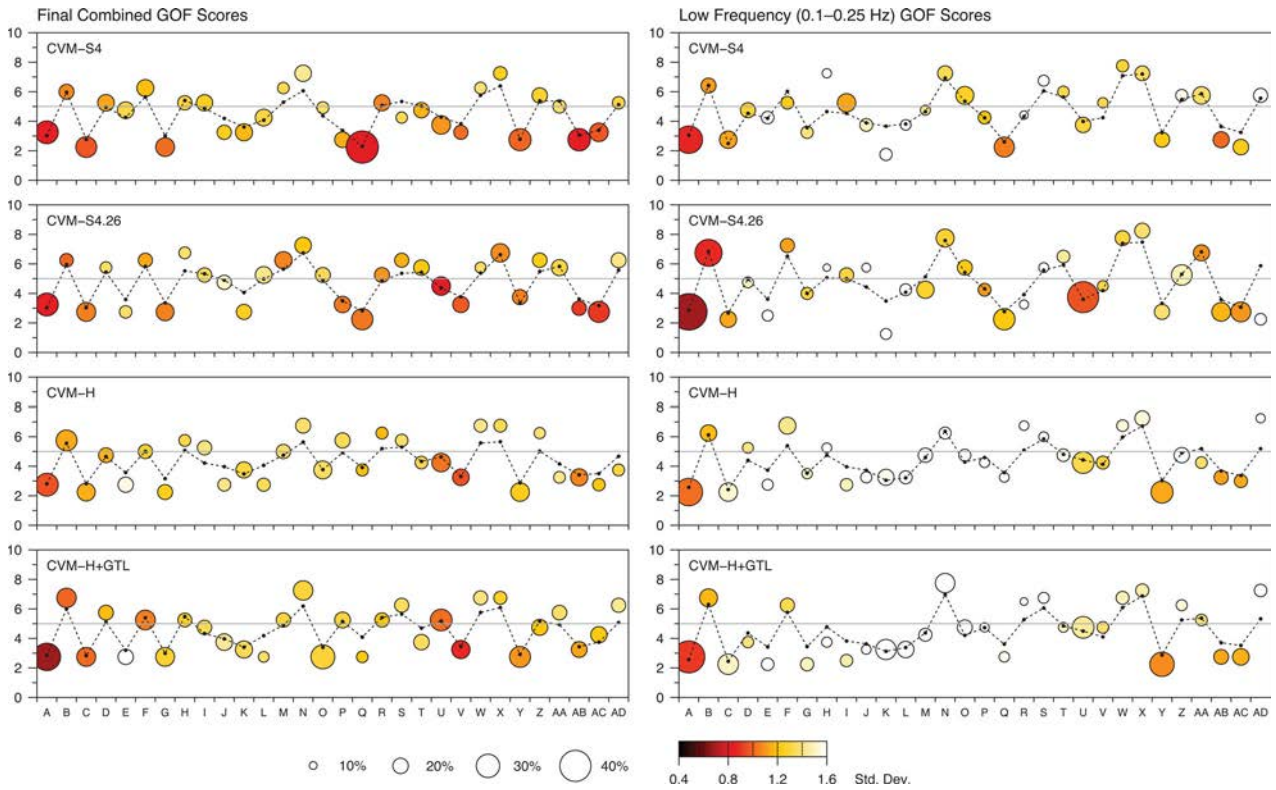

Figure 14. Statistical parameters of the validation results discretized by velocity model. Left: results from the combined final GOF score for all frequency bands and components. Right: results corresponding to the lowest frequency band used in the validation analysis (SB_1: 0.1–0.25 Hz). For each event in both plots, coloured circles are centred at the score value corresponding to the mode of the validation histogram. The size of these circles corresponds to the percentage of stations in the histogram bin corresponding to the mode. The colour of the circles is indicative of the standard deviation of the distribution and the black dots (artificially linked across events with a dashed line) correspond to the median of each event's validation histogram.

The four velocity models considered here (CVM-S4, CVM-S4.26, CVM-H and CVM-H+GTL) belong to the two distinct families of CVMs developed and supported by scientists affiliated with the SCEC and often simply referred as CVM-S and CVM-H.

Our analysis shows that among these models, the latest version of the CVM-S family, CVM-S4.26 yields the best results in a more consistent manner. This is true even at frequencies above the model's resolution (initially tailored to be valid only up to 0.2 Hz). This is important because it provides confidence in full tomography efforts, in the sense that even though inverse problems are more limited by the nature of the computational effort involved than forward simulations, they still can have a significant impact on the use of the resulting models in problems that extend their initial limitations. This, of course, is only true if the starting model constitutes a reasonably good model. In this regard, we can then say that our analysis also suggests that moving forward, CVM-S4.26 can be regarded as a reasonably good reference model.

We also found that the model CVM-H+GTL (i.e. the model CVM-H with the addition of a geotechnical model) leads to better results than the same model without the GTL addition. This suggests that the adoption of strategies to modify the properties of the geological structures in the uppermost layers is a positive addition to models in general. We, however, noted that a careful inspection of the model CVM-H+GTL also revealed that the GTL can introduce artificial changes in the geometry of basins and the velocity contrasts between the rock basement and the near-surface soft deposits. At this point, it is unclear if these changes have a strong influence on the results, which is an aspect that will require additional attention, but we note again the fact that in general, the consideration of a GTL model seems a positive addition in simulations.

Last, we recognize that while we find these conclusions to be sufficiently well supported by our analysis, the simulation approach in itself, as applied here, may still be significantly improved. The positive aspect of this is that any advance will undoubtedly lead to better GOF values, thus better, more accurate ground motion predictions. Potential improvements include: better source models, as those used in previous studies for the 2008 Chino Hills earthquake; better attenuation parameters, which is an area of active research; more precise numerical models (i.e. greater number of points per wavelength or higher order numerical methods); and, of course, better velocity models. A logical next step, for instance, will be the addition of a more detailed GTL model to CVM-S4.26, possibly including small-scale heterogeneities in the material, especially near the surface. Other efforts needed include the simulation of a selection of the events considered here at higher frequencies (above 1 Hz) possibly including frequency-dependent attenuation models.

Future challenges include the simulation and validation of stronger events (as sufficient data become available) and the consideration of more complex physics, such as off-fault and near-surface plasticity, rough faults and surface topography. All these are areas in which we are currently working ourselves and in collaborations with other research groups.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to Robert W. Graves and an anonymous reviewer for their detailed and constructive

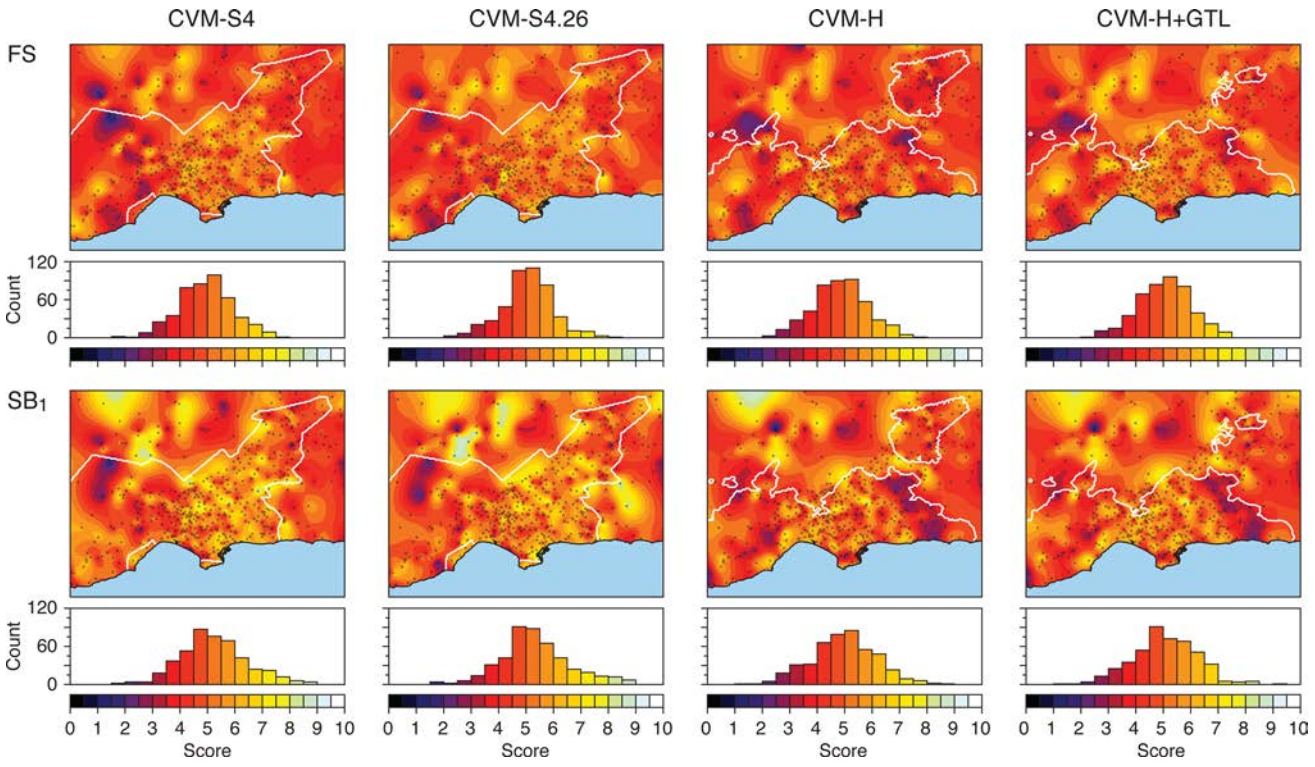

Figure 15. Average GOF maps for each velocity model obtained considering only the top 20 events with the largest number of available records (validation stations). The set at the top corresponds to GOF values obtained for the final score (FS), which combines all frequency bands and the broad-band analyses. The set at the bottom corresponds to GOF values obtained for the lowest frequency band (SB_1, 0.1–0.25 Hz). Each map is accompanied by the corresponding histogram showing the distribution of the scores in the scale from 0 to 10. On the maps, as reference, the edge of the basins based on a Z1.0 contour from Fig. 3 is shown for each model (white line).

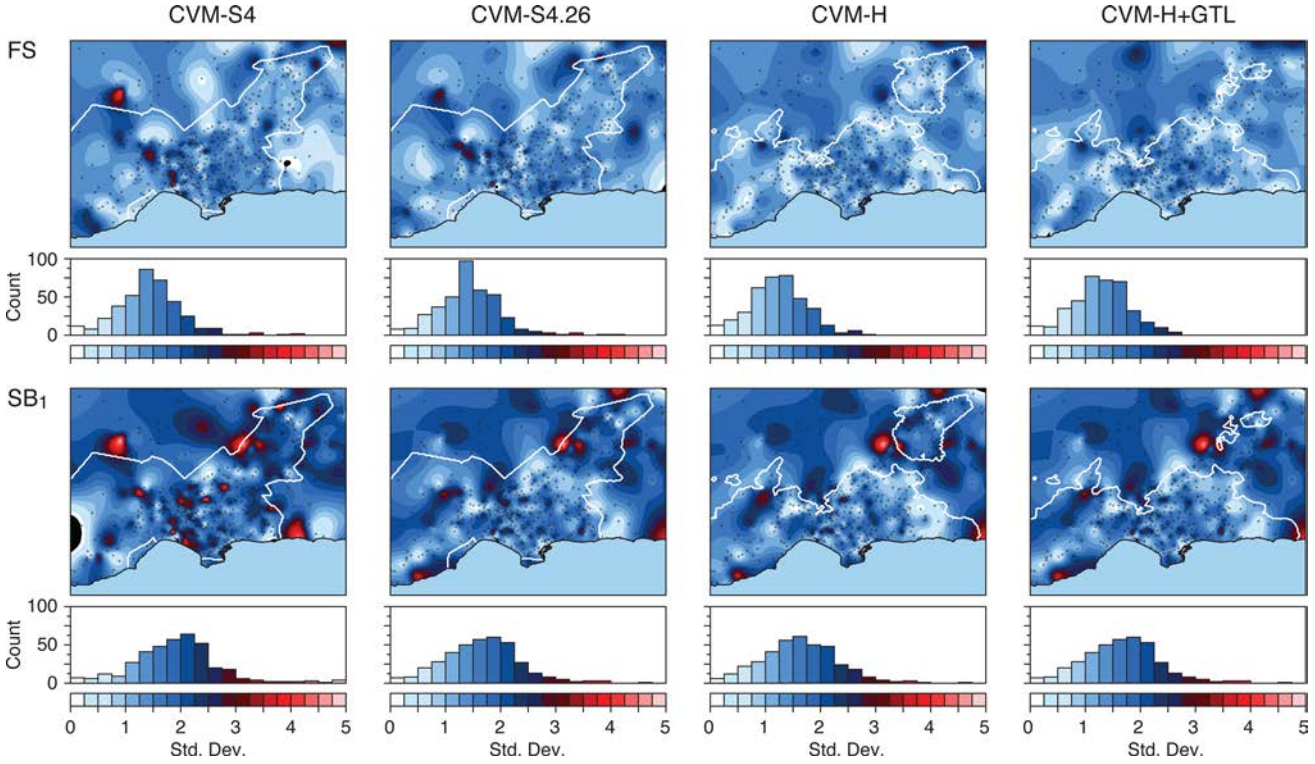

Figure 16. Standard deviation values at each station computed from the scores of the top 20 events with the largest number of available records, for each station. These standard deviation values correspond to the average scores shown in Fig. 15. The top set shows the results for the FS values and the bottom set for the SB_1 values.

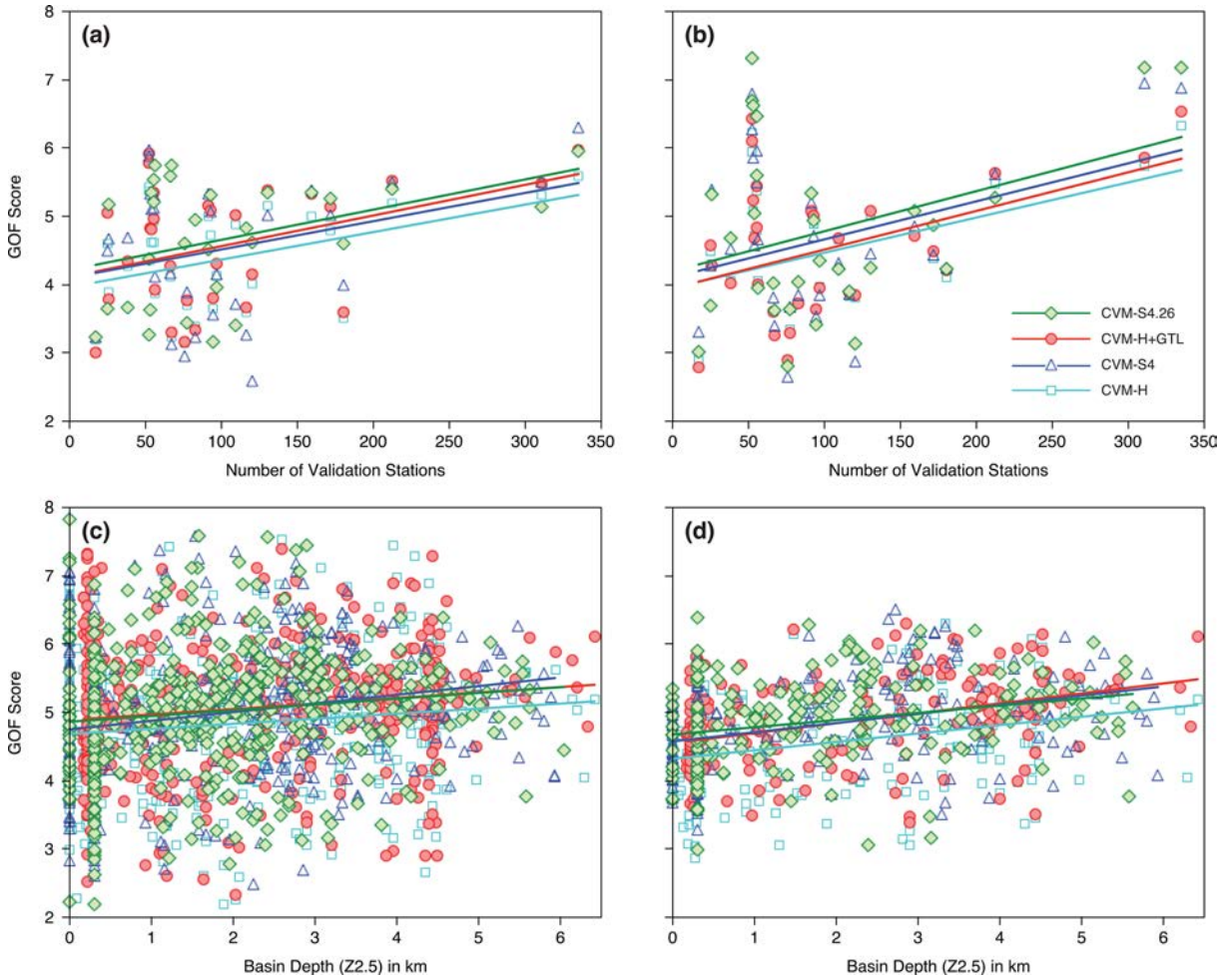

Figure 17. GOF scores correlations. (a) Average GOF scores computed as the combined value of all components and frequency bands per event and velocity model, versus the number of stations available for the validation analysis. (b) GOF scores obtained for the lowest frequency band (SB_1, 0.1–0.25 Hz) for each model and event, versus the number of stations available for the validation analysis. (c) Average of the GOF scores at each station for all the events that were recorded at a given station, discretized by velocity model, versus the basin depth at the location of the station (measured as the depth to $V_S = 2.5$ km s^{-1}). (d) Average of the GOF scores at each station for all the stations that recorded five or more events, discretized by velocity model, versus the basin depth at the location of the station (also measured as the depth to $V_S = 2.5$ km s^{-1}).

comments. The authors also thank David Gill, at the Southern California Earthquake Center (SCEC) for his help with the generation of the etree versions of the velocity models using the SCEC UCVM platform; and En-Jui Lee at the University of Southern California (now at the National Cheng Kung University, Taiwan) for providing the events catalogue and focal mechanisms. This work was supported by the United States Geological Survey (USGS) through the award 'Evaluation of the southern California seismic velocity models through ground motion simulation and validation of past earthquakes' (G14AP00034); and by the SCEC through the award 'Evaluation of CVM-SI.26 perturbations integration involving undergraduate computer science and graduate earth sciences and engineering students' (14032). Additional support was provided through the U.S. National Science Foundation (NSF) awards 'SI2-SSI: A Sustainable Community Software Framework for Petascale Earthquake Modeling' (ACI-1148493) and 'SI2-SSI: Community Software for Extreme-Scale Computing in Earthquake System Science' (ACI-1450451). This research is also part of the Blue Waters sustained-petascale computing project, which is supported by NSF (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications (NCSA). Computational support was possible through PRAC allocations supported by NSF awards 'Petascale Research in Earthquake System Science on Blue Waters (PressOnBlueWaters)' (ACI-0832698); and 'Extending the spatiotemporal scales of physics-based seismic hazard analysis' (ACI-1440085). SCEC is funded by NSF Cooperative Agreement EAR-1033462 and USGS Cooperative Agreement G12AC20038. The SCEC contribution number for this paper is 6150.

REFERENCES

Aagaard, B.T. *et al.*, 2010. 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.

Anderson, J.G., 2004. Quantitative measure of the goodness-of-fit of synthetic seismograms, in *Proceeding 13th World Conference on Earthquake Engineering,* International Association for Earthquake Engineering, Vancouver, British Columbia, Canada, Paper 243, p. 14.

Bao, H., Bielak, J., Ghattas, O., Kallivokas, L.F., O'Hallaron, D.R., Shewchuk, J.R. & Xu, J., 1998. Large-scale simulation of elastic wave

propagation in heterogeneous media on parallel computers, *Comput. Meth. Appl. Mech. Eng.,* **152**(1–2), 85–102.

Bielak, J. *et al.*, 2010. The ShakeOut earthquake scenario: verification of three simulation sets, *Geophys. J. Int.,* **180**(1), 375–404.

Bielak, J., Karaoglu, H. & Taborda, R., 2011. Memory-efficient displacement-based internal friction for wave propagation simulation, *Geophysics,* **76**(6), T131–T145.

Brocher, T.M., 2005. Compressional and shear wave velocity versus depth in the San Francisco bay area, California: rules for USGS Bay Area Velocity Model 05.0.0. Tech. Rep. OFR-2005-1317, U.S. Geol. Surv.

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.

Brocher, T.M., Aagaard, B.T., Simpson, R.W. & Jachens, R.C., 2006. The USGS 3D seismic velocity model for northern California, *EOS, Trans. Am. geophys. Un.,* **87**(52), Fall Meet. Suppl., Abstr. S51B–1266.

Chaljub, E., Moczo, P., Tsuno, S., Bard, P.-Y., Kristek, J., Kaser, M., Stupazzini, M. & Kristekova, M., 2010. Quantitative comparison of four numerical predictions of 3D ground motion in the Grenoble Valley, France, *Bull. seism. Soc. Am.,* **100**(4), 1427–1455.

Chen, P., Zhao, L. & Jordan, T.H., 2007. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97**(4), 1094–1120.

Cui, Y. *et al.*, 2010. Scalable earthquake simulation on petascale supercomputers, in *SC'10: Proceeding of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis,* IEEE, pp. 1–20.

Ely, G.P., Jordan, T.H., Small, P. & Maechling, P.J., 2010. A Vs30-derived near-surface seismic velocity model, in *Abstract AGU Fall Meeting,* no. S51A-1907, San Francisco, California, December 13–17.

Fujiwara, H. *et al.*, 2009. A study on subsurface structure model for deep sedimentary layers of Japan for strong-motion evaluation, Tech. Rep. 337, National Research Institute for Earth Science and Disaster Prevention, Tsukuba, Japan (in Japanese).

Gill, D., Small, P., Taborda, R., Lee, E.-J., Olsen, K.B., Maechling, P.J. & Jordan, T.H., 2015. Standardized access to seismic velocity models using the unified community velocity model (UCVM) software, in *Abstract SSA Annual Meeting, Seismological Research Letters*, 86(2B), 679, Pasadena, California, 21–23 April.

Graves, R.W., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. seism. Soc. Am.,* **86**(4), 1091–1106.

Graves, R.W., 2008. The seismic response of the San Bernardino basin region during the 2001 Big Bear lake earthquake, *Bull. seism. Soc. Am.,* **98**(1), 241–252.

Graves, R. *et al.*, 2011. CyberShake: a physics-based seismic hazard model for Southern California, *Pure appl. Geophys.,* **168**(3–4), 367–381.

Hauksson, E., 2000. Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California, *J. geophys. Res.,* **105**(B6), 13 875–13 904.

Isbiliroglu, Y., Taborda, R. & Bielak, J., 2015. Coupled soil-structure interaction effects of building clusters during earthquakes, *Earthq. Spectra,* **31**(1), 463–500.

Kohler, M.D., Magistrale, H. & Clayton, R.W., 2003. Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3, *Bull. seism. Soc. Am.,* **93**(2), 757–774.

Koketsu, K., Miyake, H., Fujiwara, H. & Tashimoto, T., 2008. Progress towards a Japan integrated velocity structure model and long-period ground motion hazard map, in *Proceedings of 14th World Conference Earthquake Engineering,* no. S10-038, p. 7, October 12–17, Beijing, China.

Komatitsch, D. & Vilotte, J.-P., 1998. The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Kristeková, M., Kristek, J., Moczo, P. & Day, S.M., 2006. Misfit criteria for quantitative comparison of seismograms, *Bull. seism. Soc. Am.,* **96**(5), 1836–1850.

Kristeková, 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., 2013. Personal communication, University of Southern California, now at National Cheng Kung University, Taiwan (R.O.C.), October 17.

Lee, E.-J., Chen, P., Jordan, T.H. & Wang, L., 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., 2014a. Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes, *Seismol. Res. Lett.,* **85**(6), 1275–1284.

Lee, E.-J., Chen, P., Jordan, T.H., Maechling, P.J., Denolle, M. & Beroza, G.C., 2014b. Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods, *J. geophys. Res.,* **119**(8), 6421–6451.

Lysmer, J. & Kuhlemeyer, R.L., 1969. Finite dynamic model for infinite media, *J. Eng. Mech. Div., ASCE,* **95**(EM4), 859–877.

Magistrale, H., McLaughlin, K. & Day, S., 1996. A geology-based 3D velocity model of the Los Angeles basin sediments, *Bull. seism. Soc. Am.,* **86**(4), 1161–1166.

Magistrale, H., Day, S., Clayton, R.W. & Graves, R., 2000. The SCEC southern California reference three-dimensional seismic velocity model version 2, *Bull. seism. Soc. Am.,* **90**(6B), S65–S76.

Magistrale, H., Olsen, K.B. & Pechmann, J.C., 2006. Construction and verification of a Wasatch front community velocity model, Tech. Rep. 06HQGR0012, U.S. Geol. Surv.

Manakou, M.V., Raptakis, D.G., Chávez-García, F.J., Apostolidis, P.I. & Pitilakis, K.D., 2010. 3D soil structure of the Mygdonian basin for site response analysis, *Soil Dynam. Earthq. Eng.,* **30**(11), 1198–1211.

Maufroy, E. *et al.*, 2015. Earthquake ground motion in the Mygdonian basin, Greece: the E2VP verification and validation of 3D numerical simulation up to 4 Hz, *Bull. seism. Soc. Am.,* **105**(3), 1398–1418.

Olsen, K.B., 1994. Simulation of three-dimensional wave propagation in the Salt Lake basin, *PhD thesis,* University of Utah, Salt Lake City, Utah.

Olsen, K.B. & Mayhew, J.E., 2010. Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 M_w 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 basins, *Bull. seism. Soc. Am.,* **93**(2), 627–638.

Plesch, A., Suess, M.P., Munster, J., Shaw, J.H., Hauksson, E., Tanimoto, T. & Members of the USR, Working Group 2007. A new velocity model for southern California: CVM-H 5.0, in *Proceedings of SCEC Annual Meeting,* Palm Springs, CA, September 8–12.

Plesch, A., Shaw, J.H., Hauksson, E. & Tanimoto, T., 2008. SCEC community velocity model (CVM-H 5.5), in *Proceedings of SCEC Annual Meeting*, Palm Springs, CA, September 6–11.

Plesch, A., Tape, C., Shaw, J.H. & Members of the USR, Working Group 2009. CVM-H 6.0: inversion integration, the San Joaquin valley and other advances in the community velocity model, in *Proceedings of SCEC Annual Meeting,* Palm Springs, CA, September 13–16.

Plesch, A., Tape, C., Graves, R., Shaw, J., Small, P. & Ely, G., 2011. Updates for the CVM-H including new representations of the offshore Santa Maria and San Bernardino basin and a new Moho surface, in *Proceedings of SCEC Annual Meeting,* no. B-128, Palm Springs, CA, September 11–14.

Prindle, K. & Tanimoto, T., 2006. Teleseismic surface wave study for *S*-wave velocity structure under an array: southern California, *Geophys. J. Int.,* **166**(2), 601–621.

Ramírez-Guzmán, L., Boyd, O.S., Hartzell, S. & Williams, R.A., 2012. Seismic velocity model of the Central United States (Version 1): description and simulation of the 18 April 2008 Mt. Carmel, Illinois, earthquake, *Bull. seism. Soc. Am.,* **102**(6), 2622–2645.

SCEDC 2013. Southern California Earthquake Center, California Institute of Technology, Data set, doi:10.7909/c3wd3xh1.

Schlosser, S.W., Ryan, M.P., Taborda, R., López, J., O'Hallaron, D. & Bielak, J., 2008. Materialized community ground models for large-scale

earthquake simulation, in *SC'08: Proceedings of ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis,* p. 11, IEEE Computer Society.

Shaw, J.H. *et al.*, 2015. Unified structural representation of the southern California crust and upper mantle, *Earth planet. Sci. Lett.,* **415,** 1–15.

Shearer, P.M., 2009. *Introduction to Seismology,* Cambridge Univ. Press.

Small, P., Maechling, P., Jordan, T., Ely, G. & Taborda, R., 2011. SCEC UCVM—Unified California velocity model, in *Abstract AGU Fall Meeting,* no. S21B-2200, San Francisco, California, December 5–9.

Stein, S. & Wysession, M., 2003. *An Introduction to Seismology, Earthquakes, and Earth Structure,* Wiley-Blackwell.

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, doi:10.1029/2001JB001628.

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.

Taborda, R., López, J., O'Hallaron, D., Tu, T. & Bielak, J., 2007. A review of the current approach to CVM-Etrees, in *Proceedings of SCEC Annual Meeting,* Palm Springs, CA, September 8–12.

Taborda, R., López, J., Karaoglu, H., Urbanic, J. & Bielak, J., 2010. Speeding up finite element wave propagation for large-scale earthquake simulations, Tech. Rep. CMU-PDL-10-109, Carnegie Mellon University, Parallel Data Lab.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science,* **325**(5943), 988–992.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.,* **180**(1), 433–462.

Tinti, E., Fukuyama, E., Piatanesi, A. & Cocco, M., 2005. A kinematic source-time function compatible with earthquake dynamics, *Bull. seism. Soc. Am.,* **95**(4), 1211–1223.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

Tromp, J. *et al.*, 2010. Near real-time simulations of global cmt earthquakes, *Geophys. J. Int.,* **183**(1), 381–389.

Tu, T., López, J. & O'Hallaron, D., 2003. The etree library: a system for manipulating large octrees on disk, Tech. Rep. CMU-CS-03-174, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania.

Tu, T., Yu, H., Ramírez-Guzmán, L., Bielak, J., Ghattas, O., Ma, K.-L. & O'Hallaron, D.R., 2006. From mesh generation to scientific visualization: an end-to-end approach to parallel supercomputing, in *SC'06: Proceedings of ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis,* p. 15, IEEE Computer Society.

UCVM Developers 2013. Proposed final CVM-S4.26.GTL model, Southern California Earthquake Center. Available at: https://scec.usc.edu/scecpedia/CVM-S4.26.GTL, last accessed February 2016.

Wald, D.J. & Allen, T.I., 2007. Topographic slope as a proxy for seismic site conditions and amplification, *Bull. seism. Soc. Am.,* **97**(5), 1379–1395.

Wells, D.L. & Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. seism. Soc. Am.,* **84**(4), 974–1002.

Wills, C.J. & Clahan, K.B., 2006. Developing a map of geologically defined site-condition categories for California, *Bull. seism. Soc. Am.,* **96**(4A), 1483–1501.

Yoffe, E.H., 1951. LXXV. The moving Griffith crack, *Phil. Mag.,* **42**(330), 739–750.

Zhao, L., Jordan, T.H., Olsen, K.B. & Chen, P., 2006. Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Bull. seism. Soc. Am.,* **95**(6), 2066–2080.