Geophys. J. Int. (0000) **000**, 1–7

A Comparative Study between Different Versions of Southern California Seismic Velocity Models: CVMS-4 and CVM-SI.26, through Simulation and Validation of Multiple Events

Ricardo Taborda,[1,2⋆] Shima Azizzadeh-Roodpish,[1,2]
and Naeem Khoshnevis,[2]

[1] *Department of Civil Engineering, University of Memphis, Memphis, TN* 38152*, USA*
[2] *Center for Earthquake Research and Information, University of Memphis, Memphis, TN* 38152*, USA*

To be submitted by June, 2016

1 INTRODUCTION

Developing community velocity models (CVMs) that can be broadly applied and continuously updated by a community of users has became of interest in simulation-based analysis. CVM-S and CVM-H models, released by Southern California Earthquake Center (SCEC), are good examples of community models, evolved over time. CVM-S, also known as CVM-S4, was originally developed by **?**) and later updated by **?**) and **?**). Recently, a new version of CVM-S, called CVM-S4.26 was built, based on the original model and the results of a sequence of 3D full-waveform tomographic inversions done by **?**) and **?**). CVM-S is known as an acceptable representations of the crustal structure in southern California at low frequencies ($f \leq 0.2$ Hz). According to **?**) CVM-S is getting better results than the alternative model CVM-H for the region understudy, even upto 1 Hz, which is in good agreement with previous works done in the area for particular cases (e.g. **?**) and **?**)). There are currently three alternative CVM-SI.26 models (2.2.1, 2.2.2 & 2.2.3) available which vary depending on how the perturbations were applied to the original model. There have been some researches with the aim of evaluation of different velocity models for southern California region in lower frequencies or for limited number of events. Such differences, however, have never been studied extensively for multiple events and/or at frequencies beyond the upper limits set by the underlying inversions used to construct the models.

Here, we design a systematic procedure, through quantitative comparisons among synthetics results of four versions of CVM-S and recorded data of thirty moderate-magnitude events, to evaluate the overall improvement of accuracy in predicting ground motion within simulation domain. The simulations are performed at 1 Hz frequency and minimum shear wave velocity of 200 m/s using Hercules, a finite element application for solving forward wave propagation problems due to kinematic faulting (**??**). Comparisions are ranked quantitatively by means of a goodness-of-fit (GOF) criteria at 0.5 Hz, obtained from a modified version of the criteria introduced by **?**). In the light of the comparision of the regional distribution of the GOF results for all events and all models, We conclude that CVM-SI26.223 consistently yields better results which confirms the improvements in this latest versions.

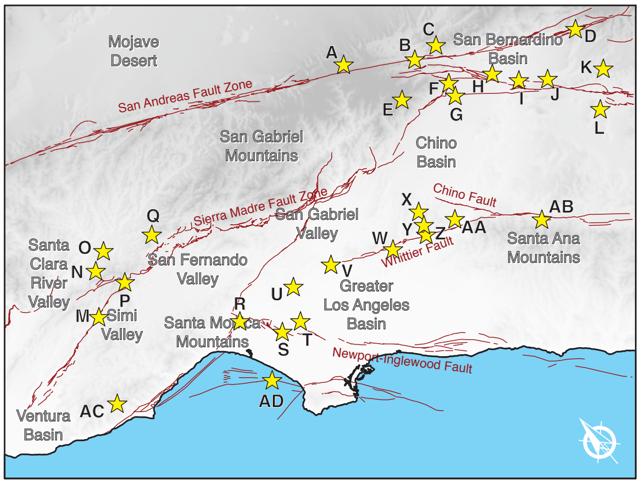

Figure 1. Major geologic structures in the region of interest including basins, valleys and mountains, along with the main quaternary faults.

2 STUDY DOMAIN AND EARTHQUAKE DATA

Thirty moderate-magnitude earthquakes ($3.5 > M_w > 5.5$), within period of 1998 to 2014, are considered for this study. The simulation domain, with a volume projection of 180 km × 135 km × 62 km, covers the entire Los Angeles metropolitan area and most of the significant geologic structures around. Fig. 1 shows the area of interest in which we labled the events with sequential letter code from A to Z and AA to AD and Table 1 provides their detailed information.The unprocessed recorded data from Southern California Earthquake Data Center (SCEDC) and Center for Engineering Strong Motion Data (CESMD) are downloaded for each earthquake to take advantage of the numerous time series available from different data centers. We obtain records for more than 800 stations overall, which provide us with enough available stations, even after removing the common stations or unusable ones, to make the results of evaluation acceptable.

3 VELOCITY MODELS

A comparative study is performed on CVM-S4 and three alternative CVM-SI.26 models named as CVM-SI.26.221, CVM-SI.26.222

Table 1. Information for 30 events including their ID ?), magnitude and final number of stations.

Code	Event ID	M_w	Num.	Code	Event ID	M_w	Num.
A	9064568	4.40	17	P	14312160	4.66	109
B	10972299	3.79	52	Q	15237281	3.86	120
C	14494128	3.72	77	R	9703873	4.24	130
D	14155260	4.88	172	S	10410337	4.70	213
E	10216101	3.60	55	T	9716853	3.98	55
F	13692644	3.74	55	U	9093975	3.77	25
G	14116972	4.42	83	V	14601172	4.44	180
H	10370141	4.45	159	W	15481673	5.10	311
I	9140050	4.37	38	X	14383980	5.39	335
J	10541957	4.10	97	Y	9818433	4.75	67
K	10530013	4.28	76	Z	10399889	3.98	91
L	14239184	3.90	66	AA	9644101	3.64	53
M	14000376	3.59	54	AB	10275733	4.73	116
N	9753489	3.90	52	AC	10403777	4.42	94
O	9096972	3.98	26	AD	14738436	3.69	93

and CVM-SI.26.223. These models are developed using gathered information from oil-well samples, gravity observations, boreholes, and seismic refraction surveys in the major southern California basins and include elastic properties and material density of particles within the domain which are required for simulation process. CVM-S model is basically based on 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 3D seismic tomography model proposed by ?), and an upper-mantle model based on teleseismic inversions introduced into the model by ?).

Using a full 3D tomographic (F3DT) inversion process involved a sequence of 26 iterations over a reference model extracted from CVM-S4, SCEC developed its recent model CVM-S4.26 (?). This is similar to the procedure applied previously on the Los Angeles region by ?). In ?), 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, $V_S = 1000$ m/s and ρ = 2000 km/m^3. The truncation was smoothed until the values extracted from the model reached 3000 m/s, 2000 m/s and 2300 km/m^3, respectively. To compute the perturbations to the initial model, ?) 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; ?) and the scattering-integral method (SI-F3DT; ?). Each iteration in the procedure involved the computation of a forward simulation done using a staggered-grid finite-differences approach (?). The forward simulations were done for a maximum frequency, $f_{\max} = 0.2$ Hz; and the misfits were computed using seismograms band-pass filtered at 0.02–0.2 Hz. After the 26th iteration perturbations were obtained, the results were merged with the original CVM-S4 model using an interpolation scheme to recover the truncated values. This is the CVM-S4.26 model.

Three alternative CVM-SI.26 models are rooted from the original one depending on the way that perturbations were applied. CVM-SI.26.221 applies an integration scheme in which negative perturbations are only used outside the basins, and positive perturbations are used everywhere. CVM-SI.26.222, on the other hand, applies negative perturbations only if outside the basins, and disregards positive perturbations when inside the basins. Finally, the latest version, CVM-SI.26.223, applies both negative and positive perturbations everywhere, but sets the original value inside the basins as a floor limit. Here, basin describes any structure with $V_s < 1000m/s$ using the cutoff value from ?) to build the starting model. In all cases, the schemes reverse the process used to build the starting model first, in order to recapture the original structures in CVM-S with values of $V_s < 1000m/s$, and apply the perturbations later (as described). These schemes have been implemented in the SCEC Unified Community Velocity Model (UCVM) software framework (?), which has already been used to produce unstructured (etree) meshes ?),?);?); ?) and ?) at resolutions finer than that used in the inversion. Fig. 2 compares the four models in 3 parts:free-surface V_S and the depths to the isosurfaces at which V_S values reach 1.0 and 2.5 km/s. The isosurfaces plots are mostly similar except at off shore in which CVM-S4.26 shows some deeper target depth. Differences through models from original version to the recent one are better visible in free surface shear velocity plots.

4 SIMULATION APPROACH

The simulations are performed by Hercules on Blue Waters system (at the National Center for Supercomputing Applicationsat) at maximum frequency of 1 Hz and minumum shear wave velocity of 200 m/s for 30 earthquakes (as point sources) and 4 velocity models discussed before. Hercules is a parallel 3D finite element computer application for solving forward anelastic wave propagation problems which implements an octree datastructure for representing unstructured hexahedral meshes in memory (?) with a solution approach using 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. To approximate the velocity and acceleration, the time integration scheme uses first-order backward and second-order central differences respectively (?).

Since the models do not provide information for quality factors, we consider it constant within the 1Hz frequency range. Q_S is calculate by the empirical rule

$$Q_S = 10.5 - 16V_S + 153V_S^2 - 103V_S^3 + 34.7V_S^4 - 5.29V_S^5 + 0.31V_S^6 \quad (1)$$

used in ??) and for Q_P we use equation

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

derived from the situation with no attenuation due to dilatational deformation ((?) and (?)).

The finite element meshes of Hercules are built at run-time, and the variable size of the elements satisfies the rule

$$e \leq \frac{V_S}{f_{\max} p}\ , \quad (3)$$

where p is the number of points per wavelength. We consider minimum requirement for p equal to 8. However, 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 due to the octree structure of the mesh.

5 EVALUATION METHOD

Using the goodness-of-fit (GOF) method proposed by ?) and modified by ?), GOF scores results for all velocity models and earthquakes combinations are prepared. The method is basically based on comparions of synthetics records with data, at locations where records are available. Parameters involved in calculating GOF

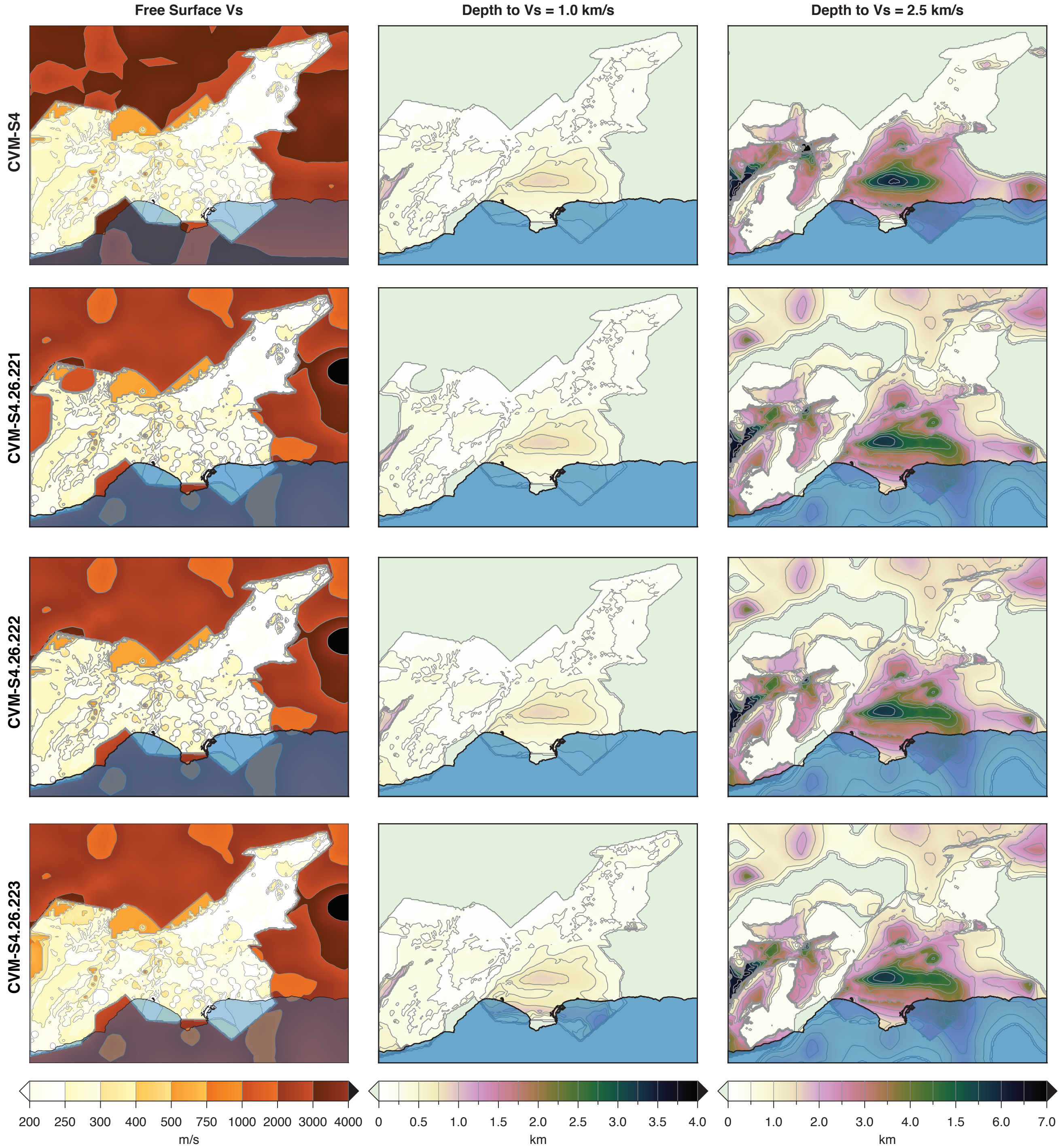

Figure 2. Comparison between the southern California community velocity models considered. Left: free-surface shear wave velocity, V_S. Center: depth to the isosurface at which $V_S = 1000$ m/s. Right: depth to the isosurface at which $V_S = 2500$ m/s.

scores (scaled from 0 to 10 showing improvement with increment) include 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). After calculating the eleven scores for each pair, the combined score can be obtained from:

$$\mathrm{S} = \frac{1}{9}\left(\frac{\mathrm{C1}+\mathrm{C2}}{2} + \frac{\mathrm{C3}+\mathrm{C4}}{2} + \sum_{i=5}^{11} \mathrm{C}i\right) . \tag{4}$$

The scoring procedure will be applied to different bands, following the guidelines suggested by **?**), using compatible “broadband” sets, and a series of band-pass filtered versions of the signals or sub-bands, SB_1, SB_2, and SB_3. The results of S from equation

(4) for the broadband (BB) and each of the sub-bands (SB_i) are then combined to obtain a final score (FS), define as:

$$\mathrm{FS} = \frac{1}{4}\left(\mathrm{BB} + \sum_{i=1}^{3} \mathrm{SB}_i\right) . \qquad (5)$$

We use signals band-pass filtered between 0.1 and 0.5 Hz for BB, and between 0.1 and 0.25 Hz, 0.25 and 0.35 Hz, and 0.35 and 0.5 Hz for SB_1, SB_2, and SB_3, respectively. Although we perform the simulations for 1 Hz frequency,the upper limit of 0.5 Hz is selected to better present the differences between various models. The lower limit is goverened by the need for minimizing instrumental and numerical processing issues at the low frequencies (0–0.1 Hz), which is applied duringcan processing on signals. More over, This is helpful to eliminate permanent displacements in the synthetics extracted at locations near the epicenter, especially for shallower earthquakes.

6 SIGNAL PROCESSING

We process the data and perform gain and baseline corrections, and apply a high-pass filter at 0.05 Hz before integrating to obtain velocities and displacements. In addition, signals are also (sub)-sampled to have the same time-step size, Δt. Plus, both data and synthetics were synchronized using the earthquake original times. Applying filters equally to data and synthetics, using a common Δt equal to 0.1 s and a decimation corner (low-pass) frequency of 2 Hz, and synchronizing the start- and end-time of each pair of signals provides a high 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 in this regard can be found in **?**)**?**);**?**).

7 RESULTS

Here, We have the result of quantitative measures to assess the level of agreement between the data and the simulation, using the GOF scores, which is discussed in details in evaluation method section. We begin by presenting the overall final result, and then break it down into some of its different components and combinations.

Fig. 3 presents GOF results of all four vesions of velocity models for some selected events with largest number of stations. The countour surface in these plots are built based on FS values computed at each section. Although the spacial interpolation for colored surface is artificial, it is facilitate understaing of results and it is consitant with previous similar works. Brighter colors corresponds to higher scores and consequently better matchs. Fig. 3 also includes histograms of the scores in interval of 0.5 points on the 0 to 10 scoring scale. Having histograms with a tendency toward right side of the scale indicates more stations with higher scores. We can see an overall improvement in this matter, while going from CVM-S4 to CVM-S4.26.223.

Since the models differ only in detailes, the scores corresponding to them would be close and the difference plots can better highlight the differences and would be more helpful to discuss the results. Fig. 4 shows differece between the distribution of total scores for all the validation stations in the region of interest for any of the events for 3 versions of CVM-S4.26 with respect to CVMS-4. This figure illustrates the wide range of results indicating events and areas in which any of the velocity models have better results.

In the next step, we investigate the results of improvent in goodness of fit scores of 3 different versions of CVM-S4.26 with respect to original version, CVM-S4, for the broad band and in different defined band widths. The results are presented for 2008 Chino Hills Earthquake. There were studies previously for this events showing that simulations produced results closer to observations in the lower frequency bands for velocity model CVM-S4 (**?**). Similar results were confirmed by **?**)for both models CVM-S4 and CVM-S4.26.223. Similar results can be noted for all different versions of CVM-S4.26. The overall improvement with respect to CVM-S4 model is more noticable at lower frequency band (0.1-0.25 Hz) than in higher frequncy bands.

8 DISCUSSION AND CONCLUSIONS

The main research goal of the proposed activities is to evaluate and compare the different available versions for CVM-S4 velocity models. In order to do that, series of simuations are done for the under study velocity models and 30 moderate magnitude events located in the simulation domain. The GOF scores are calculated for them and the validation results condensed here statistically to facilitate their analysis. Fig. 6 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) from all the stations for each event. CVM-S4.26.223 is the most consistent model to yield the highest scores, between all four models. 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 and overall mean values of each event. That is, the top score (4.87)corresponds to CVM-S4.26.223, followed by CVM-S4.26.222, CVM-S4.26.221, and CVM-S4.00. These observations are also consistent with the equivalent analysis when done based on the GOF scores obtained for broadband and the different frequency bands analysis. Our results for the goodness-of fit measure of the synthetics indicate that, CVMS4.26.223 shows an overall improvement for all the events.

ACKNOWLEDGMENTS

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 Southern California Earthquake Center (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) award “SI2-SSI: A Sustainable Community Software Framework for Petascale Earthquake Modeling” (ACI-1148493). 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 (PressOn-BlueWaters)” (ACI-0832698); and “Extending the spatiotemporal

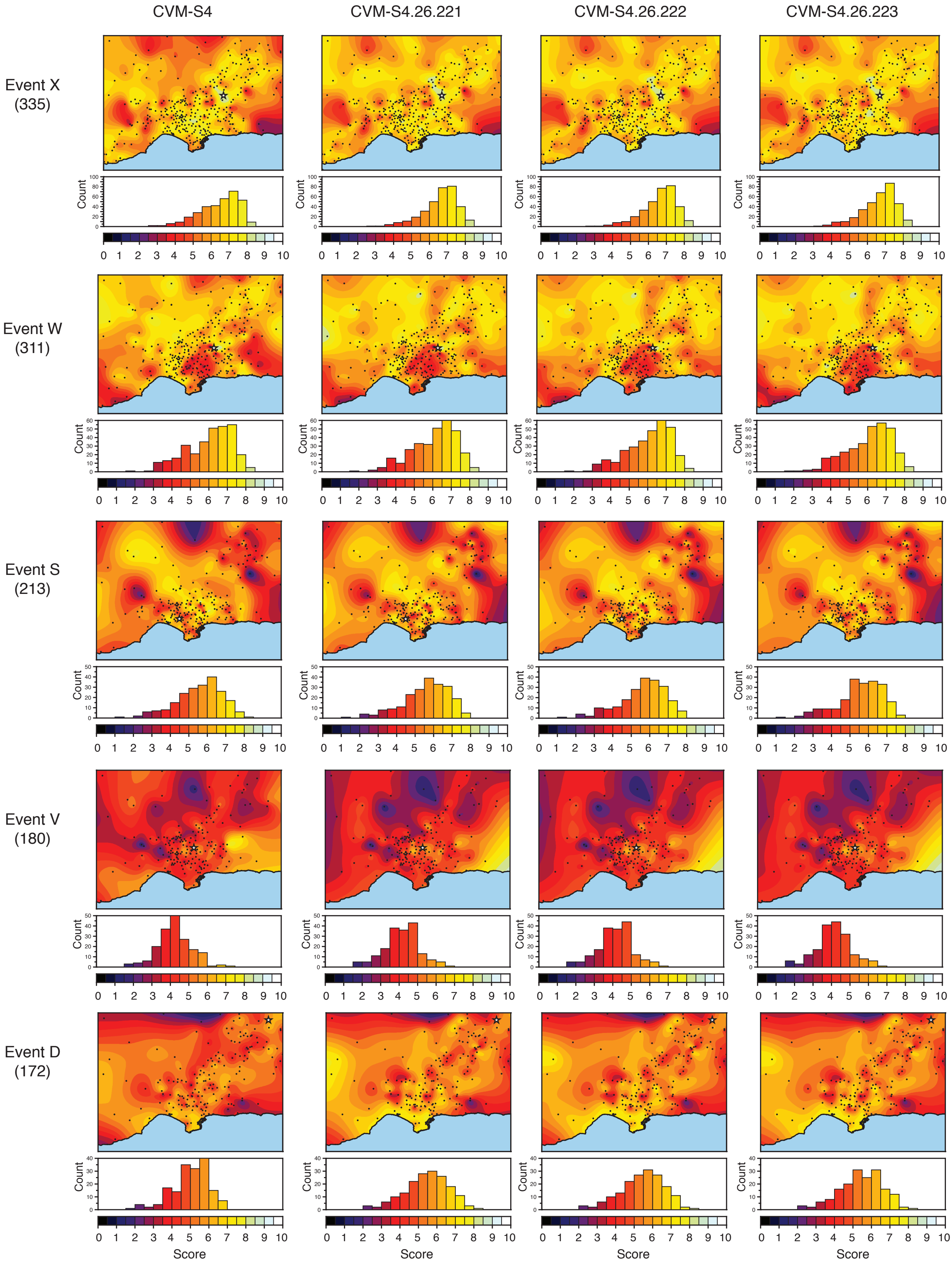

Figure 3. GOF map for selected events with higher number of stations. Countours indicate the total FS score coming from the average of results in 3 directions. Event codes and their number of stations are mentioned in the plots. . all the events comparing 3 versions of CVM-S4.26 with CVM-S4 velocity model. Stations are depicted by dots and epicenters are located by stars.

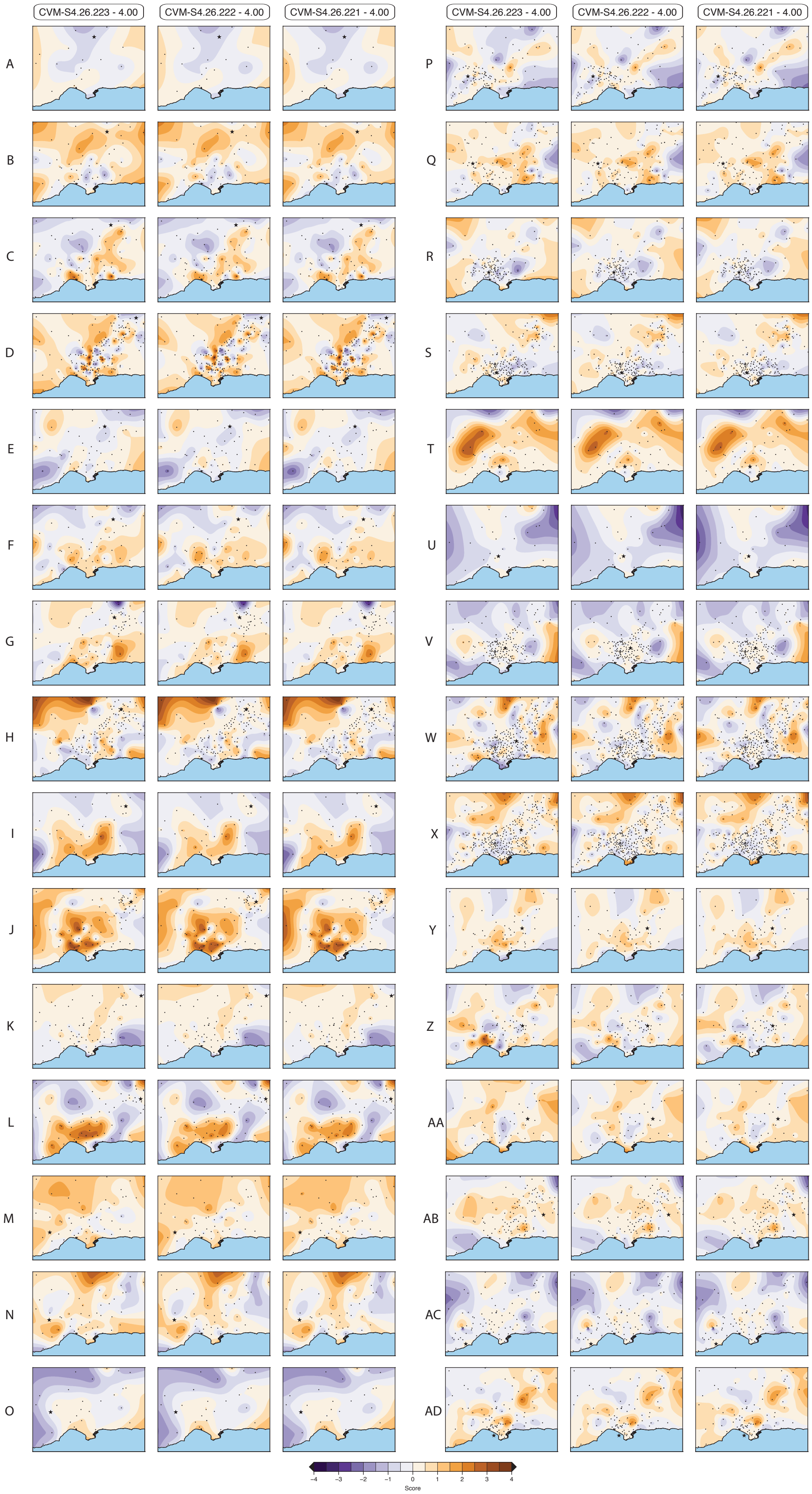

Figure 4. difference GOF map for all the events comparing 3 versions of CVM-S4.26 with CVM-S4 velocity model. Stations are depicted by dots and epicenters are located by stars.

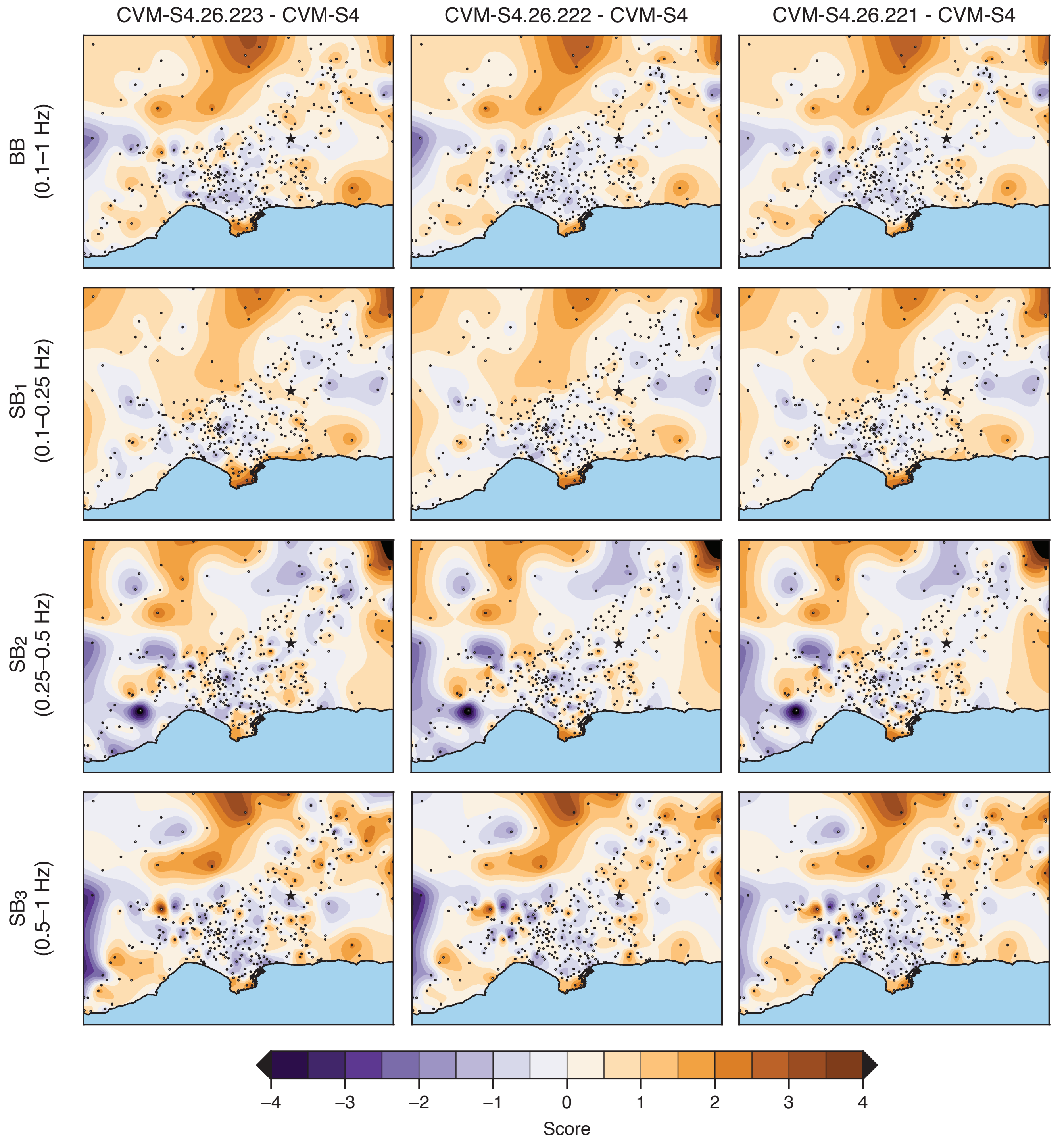

Figure 5. Improvement GOF map Chino Hills earthquake for broad band and for different bandwidths. Stations are depicted by dots and epicenters are located by stars.

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 PEDNING.

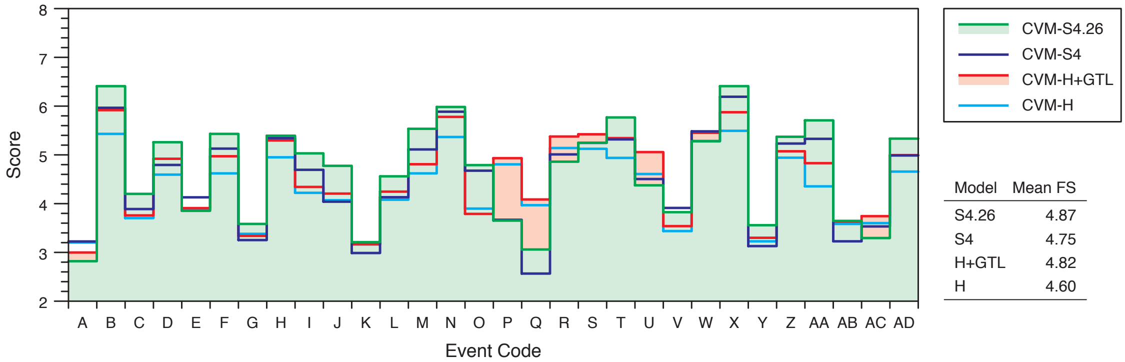

Figure 6. Summary of GOF results for all events. Values correspond to the average FS scores of all stations, in all components and frequency bands, discretized by velocity models. The values in the table to the right correspond to the mean value of the score for each velocity model considering all events. The areas under the lines for the values obtained with simulations using the velocity models CVM-S4.26.223, CVM-S4.26.222, CVM-S4.26.221 and CVM-4 are filled to highlight that CVM-S4.26.223 yields the largest number of events with the top value.