2025 SCEC CVM Workshop

Multi-Scale Seismic Velocity Models for the San Andreas Fault System in the Western US

Date: April 4, 2025
Conveners: Patricia Persaud (U Arizona), Kim Olsen (SDSU), Artie Rodgers (LLNL), Philip Maechling (USC), and Yehuda Ben-Zion (USC)

Website: www.scec.org/events/2025-SCEC-CVM-Workshop

Abstract

SCEC hosted a workshop to accelerate the development and improvement of multi-scale seismic velocity models that are essential for accurate ground motion simulations and seismic hazard assessments in California and beyond. While significant progress has been made, challenges remain in integrating new data, improving resolution, merging models, and assessing uncertainties. This workshop focused on planning a coordinated research approach to develop multi-scale P- and S-wave velocity models for the San Andreas fault system in the Western US, targeting seismic wavefield simulations up to 5-10 Hz, a frequency range critical for engineering applications. Participants discussed tomography workflows, strategies for merging high-resolution local models and regional models, integrating diverse geophysical data, and developing robust uncertainty quantification methods. The workshop also addressed the development of essential IT tools for model management and access, ensuring practical implementation of workshop recommendations. Discussions at the workshop will guide the development of the next generation of SCEC CVMs, which will advance our understanding of crustal structures and seismic hazards.

Workshop Objectives

Community Velocity Models (CVMs) provide foundational information for many basic and applied topics including the determination of earthquake locations and other source properties, imaging of the subsurface, and simulations of ground motions for use in seismic hazard models. However, current large-scale velocity models within the state of California lack the deterministic information needed for ground motion simulations at the high frequencies required in engineering applications. To address this limitation, the next generation of CVMs should resolve fine-scale seismic structure, in particular in basins, the shallow crust, and around fault zones to allow simulations of ground motion in areas of high population density and critical infrastructure to frequencies of interest to engineers. The next generation of CVMs should be multi-scale while also covering the full spatial domain of the San Andreas fault system (SAFS) in the Western US. This workshop focused on advancing the development of such multi-scale velocity models that would be needed to uniformly develop a statewide seismic hazard model and for other applications. The workshop was divided into three main parts. The first session aimed at summarizing the state-of-the-art in current velocity (V_P and V_S) models available for the SAFS in the Western US with a focus on the different methodologies, resolutions, datasets, and spatial domains of these models. This session also covered the topic of critical data gaps and data needs such as underexplored geographic domains, full seismic characterization of crustal properties (seismic anisotropy and seismic attenuation), and model uncertainties. The second session discussed current approaches for combining velocity models to produce multi-scale models and the methods used to evaluate these models with an emphasis on improving future ground motion predictions across a broad range of frequencies. The third session introduced current IT tools available for interacting with, maintaining, querying, and visualizing models. Workshop discussions focused on future directions and the short- and near-term goals for critical data gathering, development of new methodologies, and resource needs.

The workshop topics and speakers are listed in the agenda below.

Agenda

Presentation materials may be viewed by clicking the links below. PLEASE NOTE: Files are the author's property. They may contain unpublished or preliminary information and should only be used for reviewing the talk. Only the presentations for which SCEC has received permission to post publicly are included below. All times are Pacific Daylight Time (or UTC-7).

08:30 - 08:40	Welcome and Overview of Workshop Objectives	Yehuda Ben-Zion

Session 1: Current Seismic Velocity Models for the SAFS in the Western US
Moderators Kim Olsen and Patricia Persaud

08:40 - 09:00	Use cases for velocity models & required components of statewide velocity model	Artie Rodgers
09:00 - 09:15	Overview from SCEC March '24 workshop report	Brad Aagaard
	Resolution of Current Velocity Models	
09:15 - 09:30	Regional tomographic models	Cliff Thurber
09:30 - 09:45	Geology-based models	Oliver Boyd
09:45 - 10:00	Basin-scale models	John Shaw
10:00 - 10:15	Models of the near surface material	Domniki Asimaki
10:15 - 10:30	Discussion: needs and opportunities	All
10:30 - 10:50	Break	
	Observational Data Gaps	
10:50 - 11:05	Geographic areas with critical data needs	Albert Kottke
11:05 - 11:20	Seismic attenuation	Chiara Nardoni
11:20 - 12:00	Discussion: needs and opportunities	All
12:00 - 13:00	Break	

Session 2: Methods to Combine and Evaluate Velocity Models
Moderators Kim Olsen and Cliff Thurber

13:00 - 13:15	Fusion methods - embedding and smoothing	Alan Juarez-Zuniga
13:15 - 13:30	Fusion methods - machine learning (PDF)	Te-Yang Yeh
13:30 - 13:50	Estimating epistemic uncertainty in ground motions from 3-D simulations	Norm Abrahamson

13:50 - 14:30	Discussion: needs and opportunities	All
14:30 - 14:50	Break	

Session 3: Tools for Using Models
Moderators Patricia Persaud and Artie Rodgers

14:50 - 15:10	Practical aspects of seismic velocity model tools	Phil Maechling
15:10 - 16:00	Discussion: needs and opportunities	All
16:00 - 16:30	Wrap-up	
	Workshop Adjourns	

Outcomes & Recommendations

General Summary of Outcomes

The workshop saw a wide range of user needs and interests from the attendees including discussions on improving the spatial resolution of the CVMs; accurately representing the offshore regions, sedimentary basins and near-surface seismic velocities; implementing realistic seismic attenuation in ground motion simulations; using non-seismic datasets in model development and validation; and developing multiple velocity models that capture the epistemic uncertainty in the 3D seismic structure. In contrast, the pre-workshop survey showed, fewer registrants had previously developed models, datasets, or merging algorithms to contribute to enhancing ground motion simulations and seismic hazard assessments in the Western US. This suggests an emerging interest in multi-scale CVMs, of which only a few examples currently exist. Many of the discussions pointed to a long-term research effort noting that recent improvements in the spatial density of datasets and novel techniques including machine learning based methods now facilitate underexplored approaches for building CVMs as well as new applications for the CVMs. Such applications ranged from near-surface to deep-crustal and lithospheric-scale studies. Based on the broad range of applications for multi-scale CVMs highlighted at the workshop including the seismic hazard applications, there is a need for cross-disciplinary discussions and collaborations in the early-stages of the development of these models that should also take into consideration how they are validated and maintained.

Discussions at the workshop emphasized four essential areas listed below where a coordinated research effort in developing multi-scale P- and S-wave velocity models for the SAFS in the Western US should focus on making progress.

Combining Models

There are two approaches for merging velocity models. (1) The traditional approach is to embed small-scale velocity models within regional models and smooth the boundaries (e.g., Fichtner et al., 2018; Ajala & Persaud, 2021; Yeh and Olsen, 2023). While this approach is straightforward and has the benefit of being integrated in the SCEC Unified Community Velocity Model (UCVM) software, the seismic velocities near the boundaries between models may produce artifacts despite the smoothing. Furthermore, embedding procedures only update the low-resolution model in the area covered by the high-resolution results, without enhancing the outer region, and they disregard useful information from the low-resolution results in the area covered by the high-resolution model. (2) Machine learning techniques can develop transformations between seismic velocity models of different resolution and spatial extent. A specific example is the sparse dictionary learning approach that enhances a regional lower-resolution model to match the characteristics and resolution of local higher-resolution results while preserving its regional coverage (e.g., Zhang and Ben-Zion, 2024). The developed sparse dictionaries may also provide an efficient way of representing the velocities of a multi-scale statewide velocity model. Another promising machine learning approach for merging velocity models is the Probabilistic Graphical Model (e.g., Zhou et al., 2024a,b), which enhances details in the low-resolution model regions by solving a maximum likelihood problem with prior knowledge from high-resolution models.

A technique in development referred to as a bias-informed refinement approach (Yeh and Olsen, 2025) starts with conventional embedding and smoothing and continues with iterative validation that strategically preserves the outperforming model regions, informed by errors (bias) between simulated and observed ground motions. The refinement procedure in this approach can be applied to assemble outperforming regions in merged models developed by machine learning and other techniques and can be utilized to produce a combined model over a large region of interest (e.g., California statewide model).

Additional techniques may be developed and applied to address specific needs such as gap areas or combining models that were developed with and without topography. Ultimately, the relative performance of all models and their uncertainties should be established quantitatively with rigorous validation studies using simulations of earthquakes and/or signals between virtual sources associated with stations that were not employed in the model development .

Model Validation

We propose a general workflow/system for validating 3D velocity models using comparisons of simulations and recordings for small magnitude earthquakes (~M4, to avoid uncertainties related to finite-fault effects) in different frequency bands. The validation consists of a combination of waveform evaluation using cross correlation and additional metrics to test the accuracy of various aspects of candidate velocity models, including body and surface wave travel times and waveforms at select bandwidths, duration, and peak amplitudes in both time and frequency domains. In the case of uneven spatial density and distribution of recordings, one could use virtual sources (e.g., Lu and Ben-Zion, 2022). Model validation should allow for the inclusion of constraints from ground truth via measured Vs30 and borehole data, if available. The comparison of simulated and recorded waveforms and intensity measures should be rated using a goodness-of-fit (GOF) scale for target-specific selection of intensity measures (IMs).

The simulations used for validation should support anelastic attenuation (frequency-dependent Q(f)) via select IMs such as duration, peak amplitudes, Fourier Amplitude Spectra (FAS), and topography, e.g., via curvilinear grids. Such a workflow offers the flexibility needed for a comprehensive validation of the efficacy of candidate multi-scale velocity models to reproduce key features used for different purposes, such as source estimation and ground motion prediction at different frequency bands. Validation procedures should build on existing methodologies (e.g., Lu and Ben-Zion, 2022; Olsen and Mayhew, 2010; Anderson, 2004). For validation of velocity models that include anisotropy (e.g., azimuthally dependent S-wave velocities), the validation procedure should include simulation methods that can model shear-wave splitting; the validation may be obtained through cross correlation of azimuthally dependent S-wave arrivals. The workflow should include methodology to estimate the uncertainty of the P- and S-wave velocities as well as Q(f) in the tested models (e.g., using Bayesian, Metropolis-Hasting algorithms, Ely et al., 2018; Barbosa et al., 2020).

2-year recommendations

Build the validation system and conduct a small-scale proof-of-concept for validating multi-scale velocity models under development in southern California. The validation workflow should be integrated with methodology to merge velocity models of different scales, retaining only the model parts that pass the validation criteria. Document the extent to which a merged shallow taper (constant or spatially variable depth, anchored by Vs30 values) improves the GOF values. If existing velocity models in specific areas underperform, initiate new data collection and the development of alternate models in those areas.

5-year recommendations

Apply the validation workflow as detailed in the 2-year plan to state-wide multi-scale velocity models. Develop new or alternate V_P, V_S, Q_P, Q_S, models in key areas that underperform in the 2-year plan.

Tools for Using Models

Some of the existing archives of California seismic velocity models provide software tools for accessing model properties. These software tools typically provide an interface that inputs a georeferenced point in the Earth, and outputs the material properties (typically V_P, V_S, and density) for that point. Such tools include the SCEC UCVM software (Small et al., 2017) and GeoModelGrids (Aagaard, 2020; https://geomodelgrids.readthedocs.io) that provide access to collections of seismic velocity models. SCEC's UCVM has registered more than 20 existing 3D California seismic velocity models, making these models available for visualization through the SCEC CVM Explorer, and making them accessible for use in ground motion simulation codes including SW4 and AWP-ODC.

Workshop participants discussed both model development and use. We acknowledged that several existing velocity models (e.g. USGS SFBR v21.1, CVM-H v15.1) were developed using commercial velocity model development tools (e.g. EarthVision, GoCAD). Use of commercial tools by CVM developers, rather than the prevailing "open source" software preferences of the community, indicates these commercial tools provide essential capabilities that may be required in the future. We also discussed standardization of model storage format. Existing models are delivered in multiple storage formats. GeoModelGrids defines a model storage format that may provide a flexible format supporting multi-resolution models, topography, and metadata capture. However, few existing models are currently in GeoModelGrids format, and further evaluation is needed to determine if it will support all existing and future models. Model query tools, particularly tools for populating ground motion simulation meshes with material properties, need to be high-speed to support building large simulation meshes. UCVM provides MPI-based parallel extraction methods that provide the required access speed, which greatly improves the code's scalability, but makes the UCVM software more difficult to maintain and develop.

Future CVM software tools will need to work with velocity models at multiple scales with irregular spatial domains, while constructing simulation meshes. SCEC's UCVM provides tools for combining, also called tiling, multiple velocity models, a feature that is likely to be required to combine models with different, or overlapping, coverage areas. In addition, UCVM software tools are used to modify models such as modifying near-surface properties, and smoothing interfaces at model boundaries, while building meshes based on multiple models.

The workshop identified several essential features of future velocity model software tools including model storage, discretizing a volume, querying, merging, and modifying mesh properties after extraction. Based on the group discussion, it may make sense to divide the capabilities of UCVM into a more modular structure and develop specialized software tools for each required capability (storage, meshing, query, merging, modifying). This more modular software development approach should produce software that is more maintainable than UCVM. One potential path forward is to combine development efforts by several research community members who are working on mesh making, GeoModelGrids, UCVM, and CVM visualization tools. SCEC's CVM Explorer, and CRESCENT's CVM viewer, provide useful web-browser-based tools that can provide a visual overview and interactive investigation of existing models. Although the existing viewers serve different geographical areas, it may be possible to work together on these viewers in the future. As simulations codes require more complex meshes that include topography, multi-resolution regions, and irregular structures, existing mesh building codes such as Gmsh (https://gmsh.info/), Cubit (https://coreform.com/coreform-cubit/), and Los Alamos Grid Toolbox (LaGriT) (lanl.github.io/LaGriT/) may be needed to properly discretize complex simulation volumes. To support the needs of future statewide, multi-scale seismic velocity models and ground motion simulation codes, a future CVM toolkit might be developed which uses GeoModelGrids to provide a standard storage format and query interface and uses parts of UCVM to provide tiling, model merging, and model modifications after extraction.

Computational & Instrumental Resources

Fulfilling the objectives of this workshop will require computational resources and sensor data to build the constituent models and assess their validity. Data from open access sources will likely provide all the data needed to create and evaluate models. These will come from traditional permanent networks operating in California and Nevada and available through the Federation of Digital Seismic Networks or Southern California Earthquake Data Center web services (e.g. networks: CI, BK, NC, NP, CE, TA, II), as well as temporary networks. Unconventional data sets will contribute to model building and validation (e.g. borehole logs, ambient noise cross-correlations, fiber optic distributed acoustic sensing (DAS), and geodetic and potential field data) along with surface geology and geotechnical data sets.

Models will be built on different scales, covering different domain sizes and based on different data and analysis techniques. These models will be combined as described above. The multiplicity of model combinations immediately expands the number of cases to consider. This will require a robust computational framework to organize models, track metadata (e.g. parameters used to merge models, modify near-surface seismic velocities, and include topography and seismic attenuation), generate gridded models for use in waveform simulation or other codes, and track performance of models to various validation methods and data sets. This framework must be designed to be compliant with Findable, Accessible, Interoperable and Reusable/Reproducible (FAIR) principles.

Standardized data will need to be gathered and stored for model evaluations. Validation tools will operate to compute forward simulations of waveforms or other seismological observables and compare simulated with observed data.

Computing resources will be needed, especially for tasks involving waveform simulations, particularly full waveform inversion tomography and simulations for model validation. Computing demands increase as the fourth power of the maximum frequency and with the computational domain size, so high-performance computing will be a bedrock requirement of this effort. Linear seismic wave propagation favors starting with inexpensive simulation of low-frequencies and gradually increasing the maximum frequency similar to multi-scale waveform inversion methodologies. We can envision computing resource needs (cycles and storage) growing as this effort increases the maximum frequency content of waveform simulations and comparisons with observed data.

2-yr recommendations

Resources needed for several short-term efforts include:

Firstly, effort and computing resources for the proof-of-concept for validating multi-scale velocity models under development in southern California, and simultaneously build a waveform tomography model of the San Andreas Fault system or the entire state to serve as the background for higher resolution, smaller domains with near surface models. Waveform tomography offers an attractive methodology for the statewide, long-wavelength background model because by design it improves waveform simulations, which is a main objective of this effort.

Secondly, there must be an effort to create a cyberinfrastructure framework for statewide model development and evaluation. This effort must collect current models of the region and archive them within an emerging database of models with metadata for investigations described above and other applications described in the workshop such as geodynamic modeling. While the details of the framework and database of models remain to be determined, the basic outline can be defined presently based on the outcome of the workshop and surveys of current efforts. Resources for computing effort for processing and storage of models and validation data are needed in the near-term.

5-yr recommendations

For the longer term, a computational infrastructure is needed to systematically build and objectively evaluate the performance of models at multiple scales. The details remain to be determined, but they will emerge from the preliminary efforts led by SCEC staff and a focused working group with community input. Waveform tomography and other models will continue to be developed as new data sets are collected from new earthquakes and expanding sensor networks. The statewide model building effort must accommodate strategies to incorporate new data and measure improvement in model predictions as models evolve.

Participants

The workshop saw 79 attendees out of 151 registrants. Among the attendees, 62 were from the US, with over half hailing from California. The remaining 17 participants were from the international community, from Spain, New Zealand, Italy, Germany, Croatia, Mexico, Pakistan, Greece, Singapore, and China. Only 11% of attendees were graduate students or postdocs.

In terms of industry representation:

- 47 participants (59%) were from the education sector.
- 14 participants (18%) were from government organizations.
- 3 participants (4%) each from the energy, chemical, utilities sector; consulting; and tech/software industries.
- 9 participants (11%) were from various other sectors.

Full list of attendees is available at www.scec.org/events/2025-SCEC-CVM-Workshop.

References

Aagaard, B. T. (2020). Geomodelgrids: A storage scheme and API for multiscale Earth models. Poster Presentation at 2020 SCEC Annual Meeting.

Aagaard, B.T. and Hirakawa, E.T., (2021), San Francisco Bay region 3D seismic velocity model v21.1: U.S. Geological Survey data release, doi:10.5066/P9TRDCHE

Ajala, R. and Persaud, P. (2021). Effect of Merging Multiscale Models on Seismic Wavefield Predictions Near the Southern San Andreas Fault. *Journal of Geophysical Research: Solid Earth,* **126**(10), e2021JB021915. doi: 10.1029/2021JB021915

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.

Barbosa, C.H.S., L.N.O. Kunstmann, R.M. Silva, C.D.S. Alves, B.S. Silva, D.M.S. Filho, M. Mattoso, F.A. Rochinha, A. L.G.A. Coutinho (2020). A workflow for seismic imaging with quantified uncertainty, *Computers & Geosciences*, **145**, 104615, https://doi.org/10.1016/j.cageo.2020.104615.

Ely, G., A. Malcolm, and O. V. Poliannikov (2018). Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method. *Geophysics* **83** (2): R63–R75. doi: https://doi.org/10.1190/geo2017-0321.1

Fichtner, A., van Herwaarden, D.-P., Afanasiev, M., Simutė, S., Krischer, L., Çubuk-Sabuncu, Y., Taymaz, T., Colli, L., Saygin, E., Villaseñor, A., Trampert, J., Cupillard, P., Bunge, H.-P., & Igel, H. (2018). The Collaborative Seismic Earth Model: Generation 1. *Geophysical Research Letters*, ***45*(**9), 4007–4016. https://doi.org/10.1029/2018GL077338

Lu, Y. and Y. Ben-Zion (2022). Validation of seismic velocity models in southern California with full-waveform simulations, *Geophys. Journal Int.* **229**, 1232–1254, doi: 10.1093/gji/ggab534.

Olsen, K.B., and J.E. Mayhew (2010). Goodness-of-fit Criteria for Broadband Synthetic Seismograms, With Application to the 2008 Mw5.4 Chino Hills, CA, Earthquake, *Seism. Res. Lett.* **81** , 715-723.

Small, P., Gill, D., Maechling, P. J., Taborda, R., Callaghan, S., Jordan, T. H., Ely, G. P., Olsen, K. B., & Goulet, C. A. (2017). The SCEC Unified Community Velocity Model Software Framework. Seismological Research Letters, 88(5). doi:10.1785/0220170082

Zhang, H and Y. Ben-Zion (2024). Enhancing regional seismic velocity models with higher-resolution local results using sparse dictionary learning, *J. Geophys*. *Res.*, ***129***, e2023JB027016, doi: 10.1029/2023JB027016.

Yeh, T.-Y., and K.B. Olsen (2023). Fault damage zone effects on ground motions during the 2019 Mw 7.1 Ridgecrest, CA, earthquake, *Bull. Seis. Soc. Am.*, **113(4)**, 1724-1738.

Yeh, T.-Y. and K.B. Olsen (2025). Toward a Multi-scale Community Velocity Model for California, *Seism. Res. Lett.* **96**, 2B, 1307.

Zhou, Z., P. Gerstoft, and K.B. Olsen (2024a). Graph-learning Approach to Combine Multiresolution Seismic Velocity Models, *Geophys. Jour. Int.*, **238**, 1353 - 1365.

Zhou, Z., P. Gerstoft. and K.B. Olsen (2024b). 3D Multiresolution Velocity Model Fusion With Probability Graphical Models, *Bull. Seis. Soc. Am* **114** (3): 1279 - 1292. DOI: https://doi.org/10.1785/0120230271.