

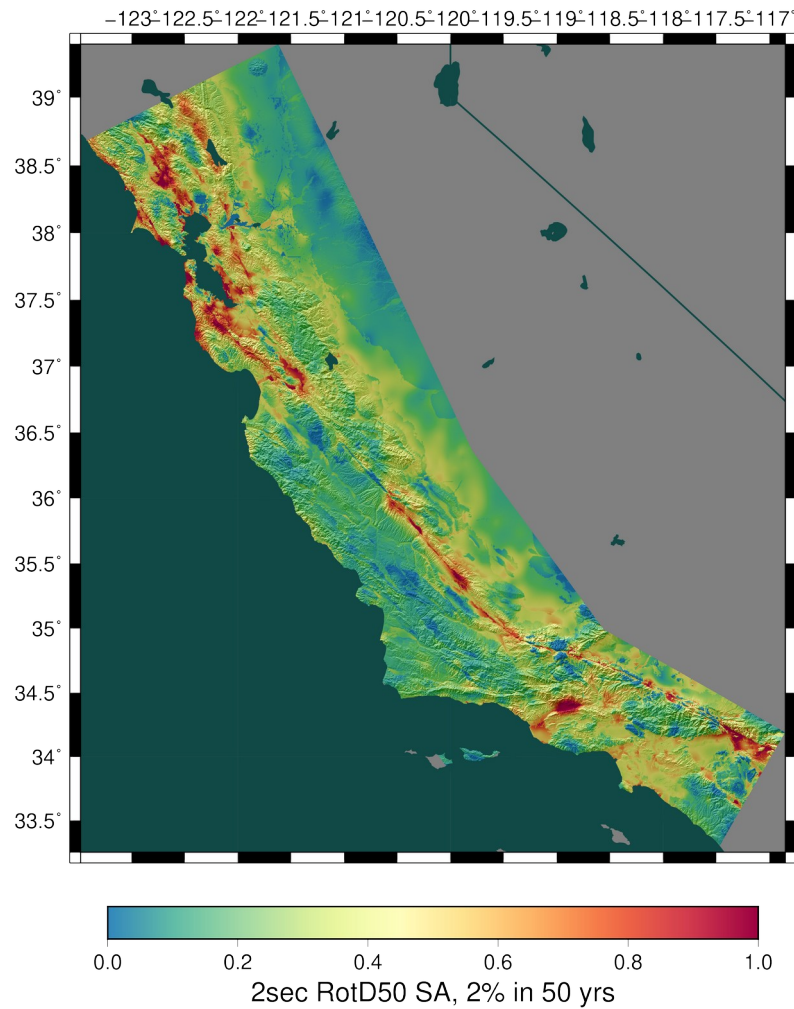
The CyberShake Platform

Scott Callaghan (SCEC)

October 1, 2026

2026 DEFORM workshop

scottcal@usc.edu

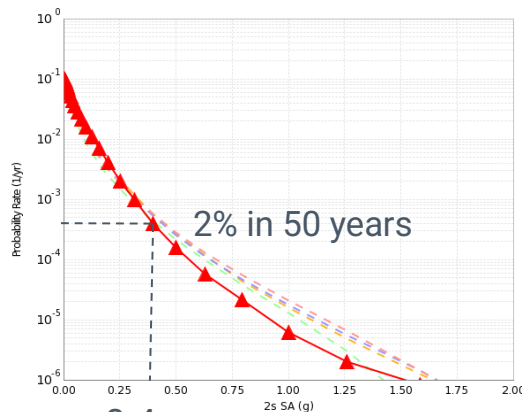


-
- PSHA Overview
 - CyberShake platform
 - Key elements
 - Computational requirements
 - Hazard data products
 - Results in California

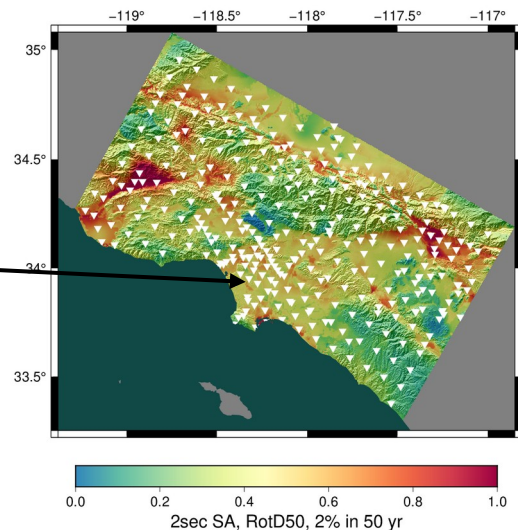


Probabilistic Seismic Hazard Analysis

- What ground motions can I expect at a location over the next N years?
 - Building codes
 - Insurance rates
 - Disaster planning
- The PSHA process:
 1. Obtain a list of all earthquakes of interest (ERF)
 2. Determine the shaking each earthquake causes
 3. Combine shaking with probabilities to produce hazard products
- Step 2 often performed using empirical GMMs
- CyberShake uses physics-based simulation

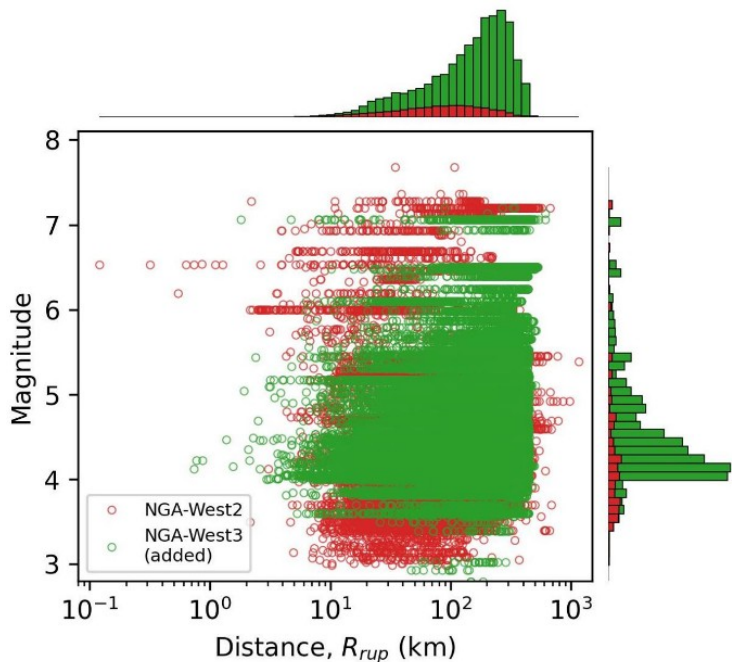


0.4 g



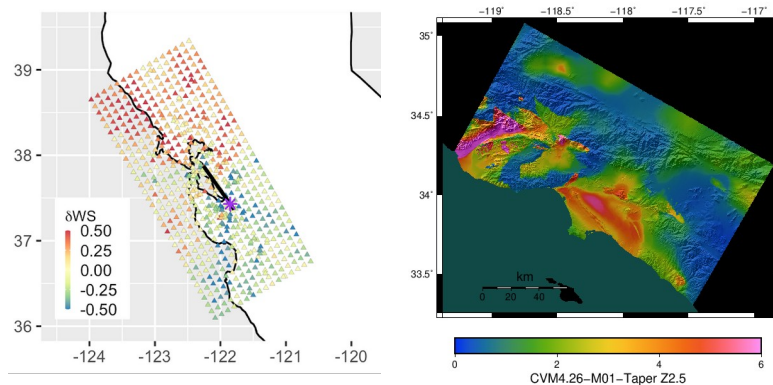
What does physics-based simulation offer?

Fills data gaps

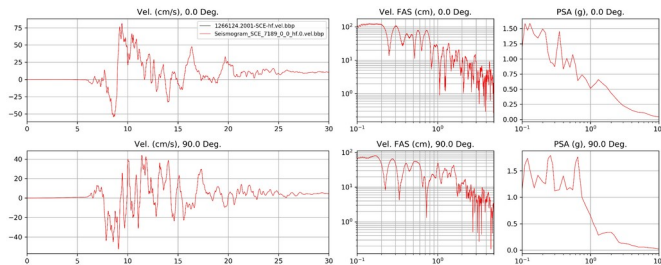


Stewart et al., 2024

Captures complex effects



Produces synthetic seismograms

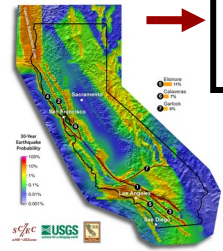




CyberShake Workflow Steps

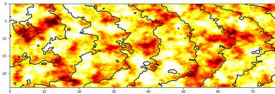
Broadband CyberShake workflow

Low-frequency CyberShake workflow



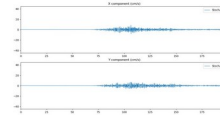
Uniform California
Earthquake Rupture
Forecast

Graves-Pitarka
kinematic rupture
generator



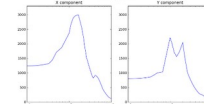
600,000+ events

High-frequency
seismogram
synthesis (BBP)

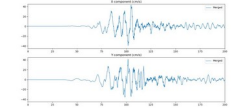


1-50 Hz
seismograms

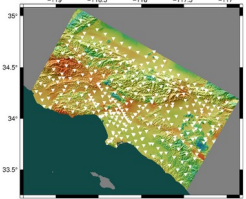
Merge and
combine



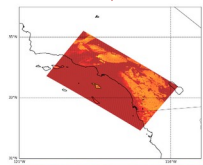
RotD50, PGA,
PGV, durations



0-50 Hz broadband
data products

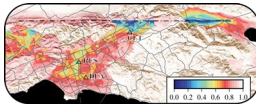


UCVM



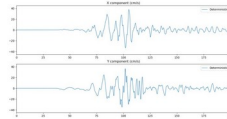
Velocity Mesh

AWP-ODC-SGT
wave propagation



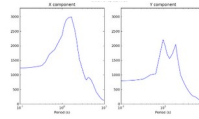
Strain Green
Tensors

Low-frequency
seismogram
synthesis



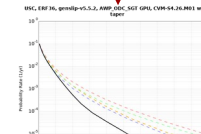
0-1 Hz low-frequency
seismograms

Intensity
measures

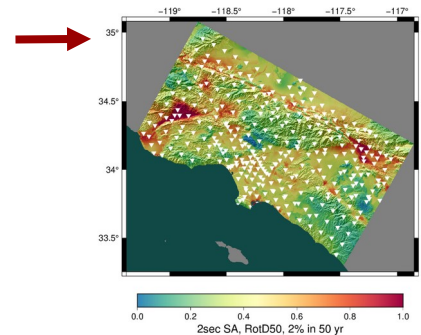


RotD50, PGA,
PGV, durations

Aggregate
data products



Hazard Curve



Hazard Map

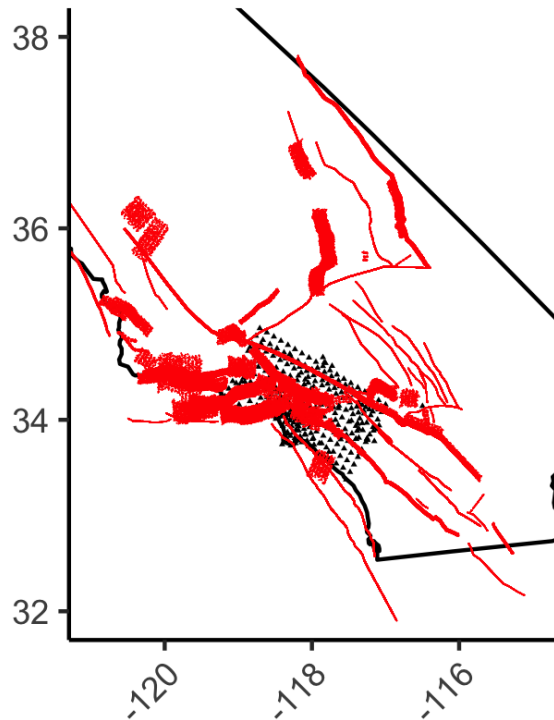


Earthquake Rupture Forecast (ERF)

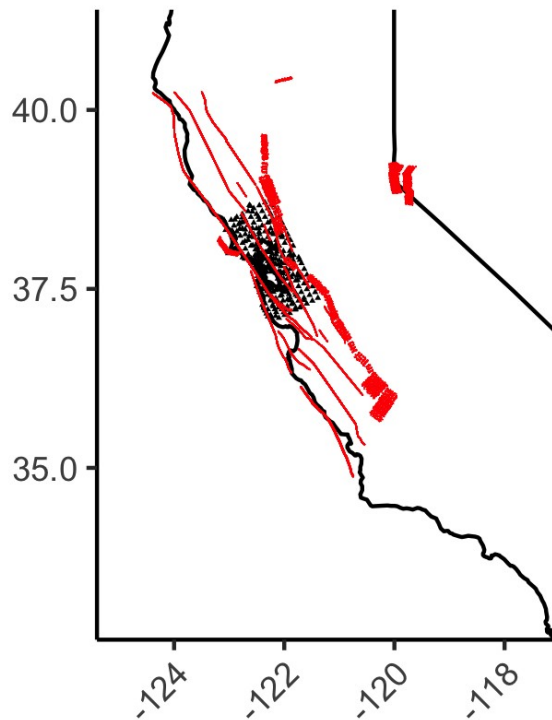
- Derived from the UCERF2 model for California
- Prescribes fault surfaces with associated probabilities and magnitudes to be consistent with data
- Magnitudes permitted to vary +/- 2 sigma to capture aleatory variability in potential magnitude
 - Median magnitudes range from 6.55 to 8.15
 - Full range is 5.95 to 8.45
- Events with median magnitude > 6.5 and with some part of the fault is within 200 km of the site are included in the simulation

CyberShake ERF

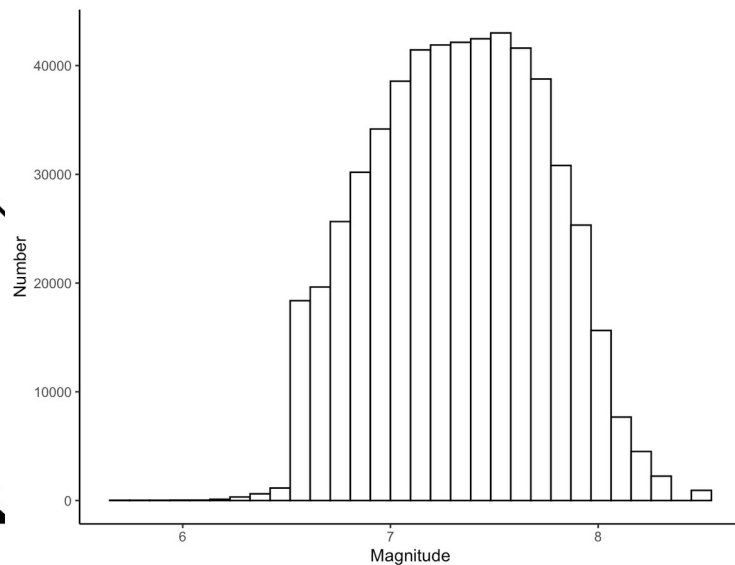
CS22.12



CS24.8

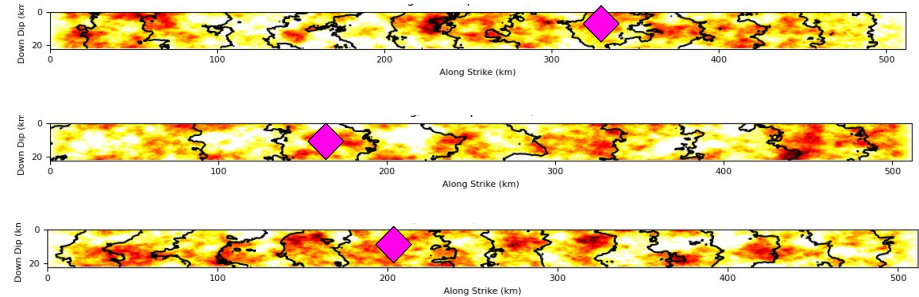
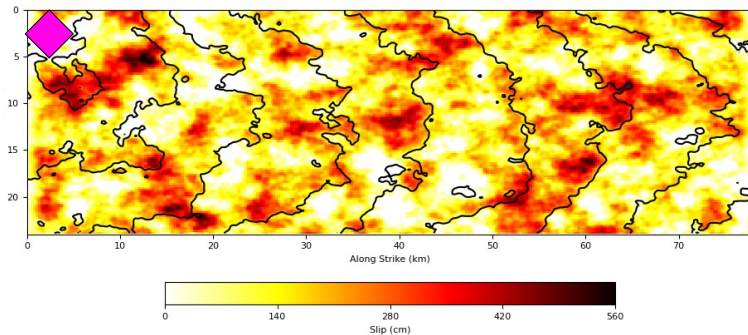


Magnitude distribution
from UCERF2



Rupture Generator

- CyberShake uses the Graves-Pitarka kinematic rupture generator
- Takes in a fault geometry, hypocenter location, rupture velocity (67.5%-87.5% of shear wave velocity), random seed
- Hypocenters are placed every 4 km along the fault surface
- Each realization or “rupture variation” has a unique rupture velocity, hypocenter, and slip distribution





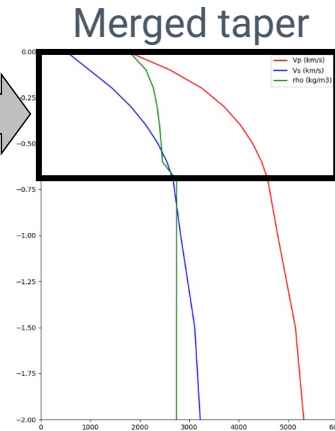
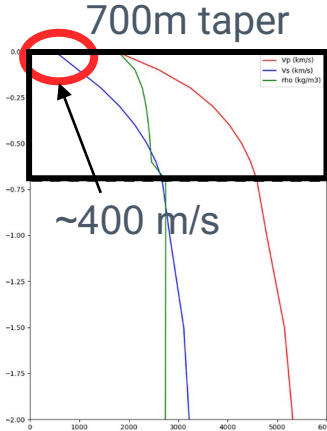
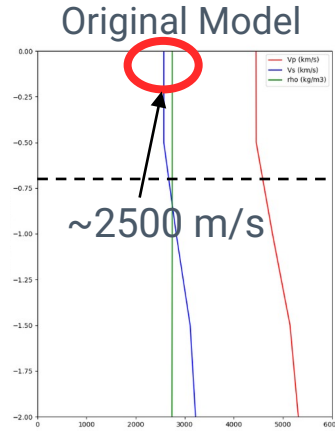
Velocity Model Development

- Begin with “best available” 3D velocity model for a region
 - May depend on the key metric
 - May require tiling together multiple models
- Consider near-surface for modification if V_s is too high
 - In California, frequently apply a V_{s30} -based taper
- Apply minimum V_s floor
- Check V_p/V_s ratio
- Perform verification simulations to check for unusual features



Merged Taper Near-Surface Modification

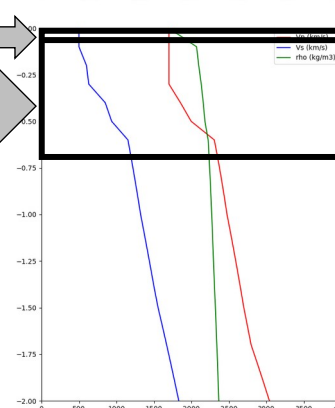
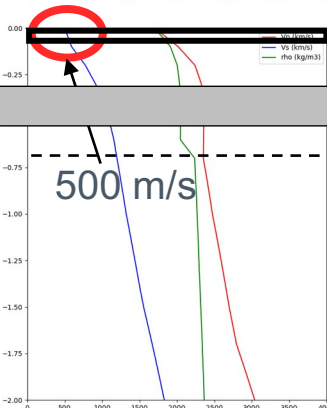
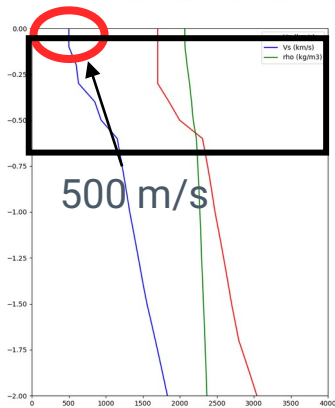
LAPD
(outside basin)



Taper is selected at all depths to 700m

Vs30: 2573 m/s -> 400 m/s
Z1.0: 0m -> 110m

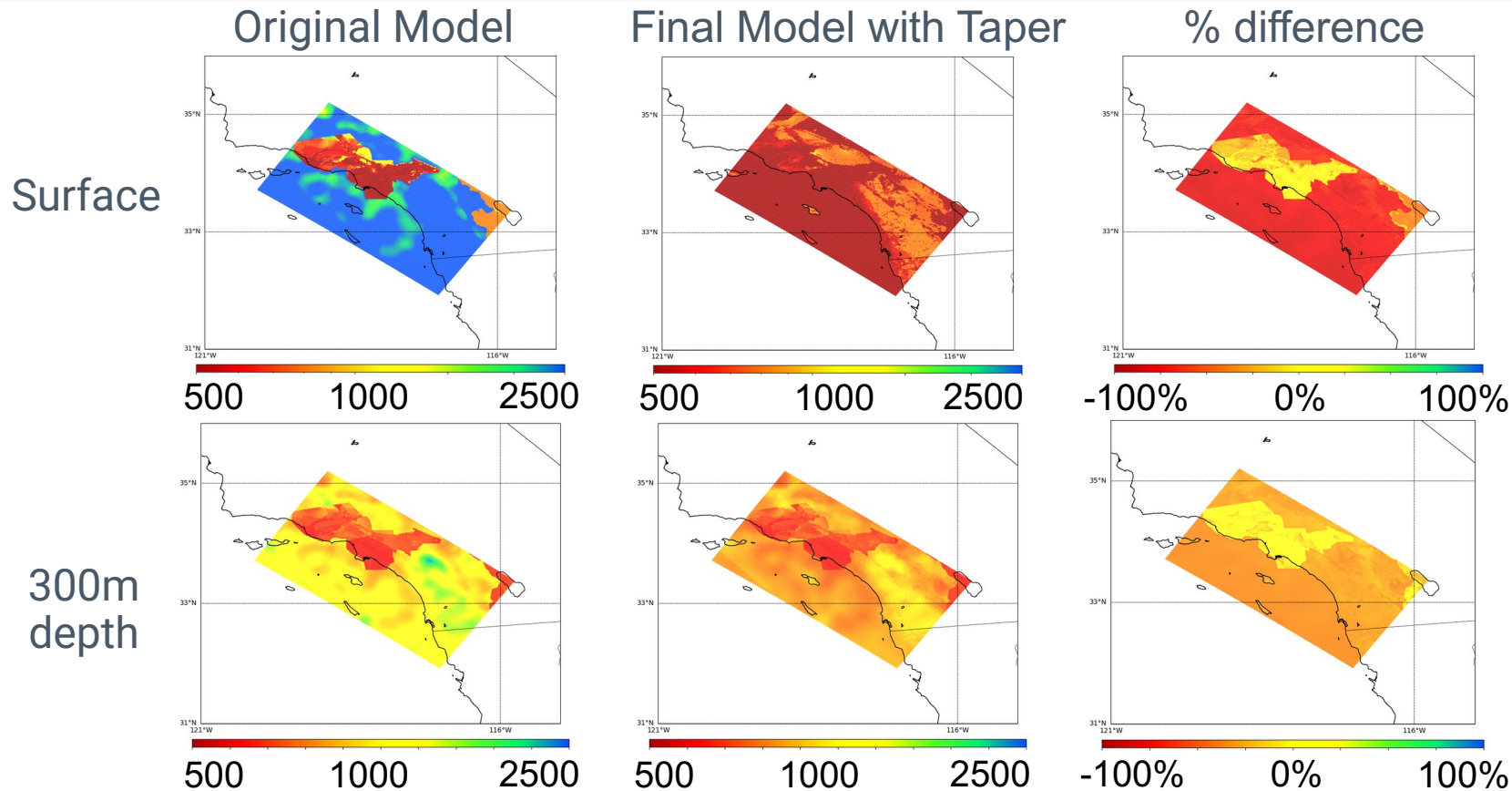
USC
(in basin)



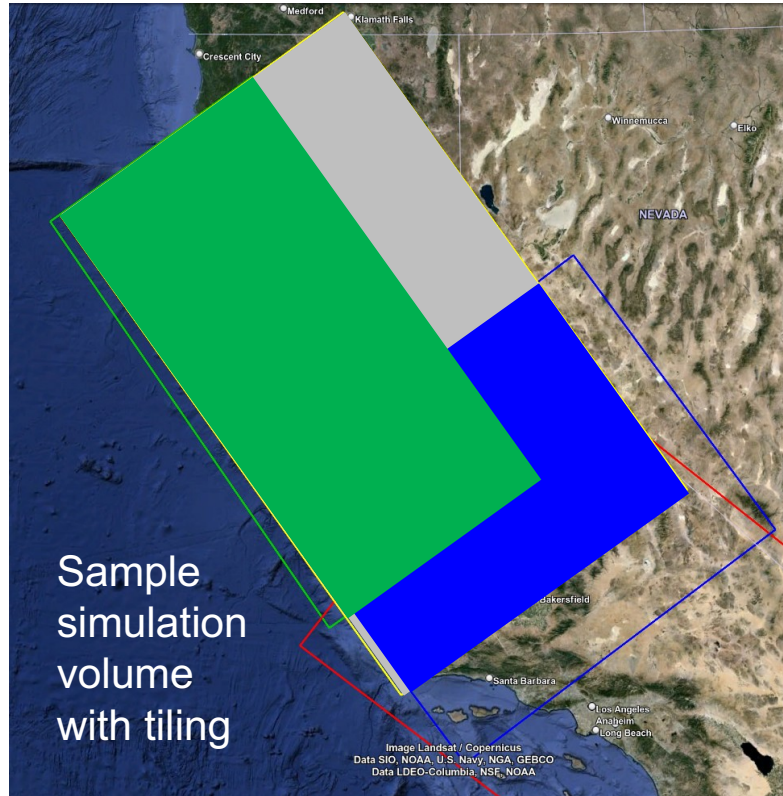
Taper is only selected at surface point

Vs30: 500 m/s -> 500 m/s
Z1.0: 580m -> 580m

Southern California Velocity Model

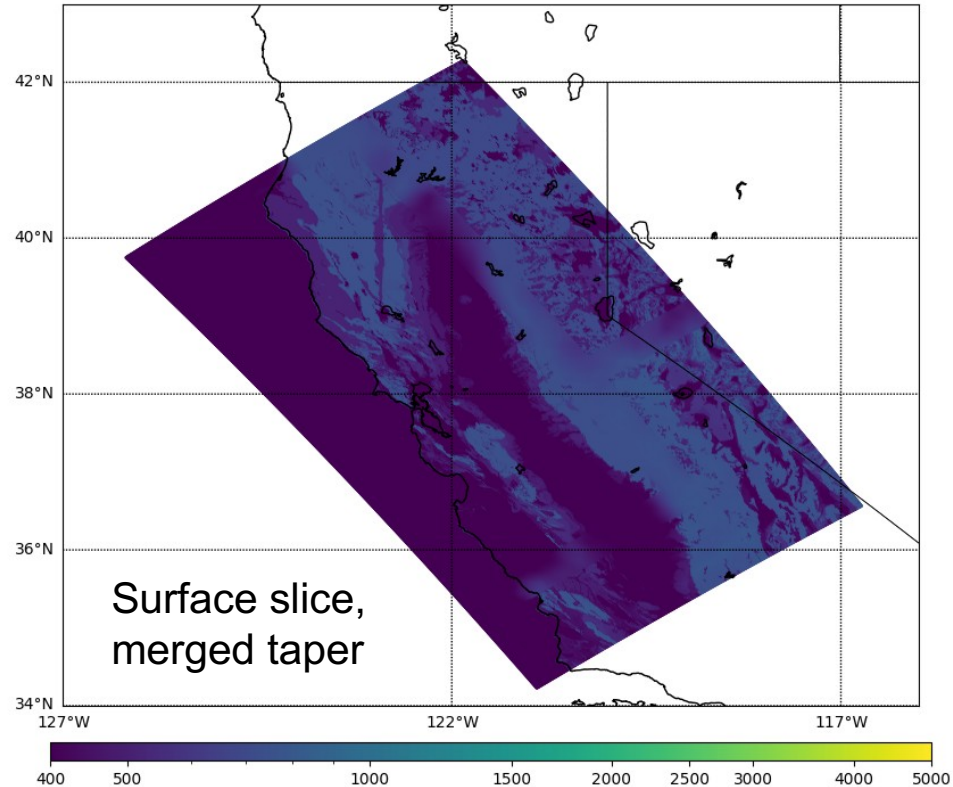
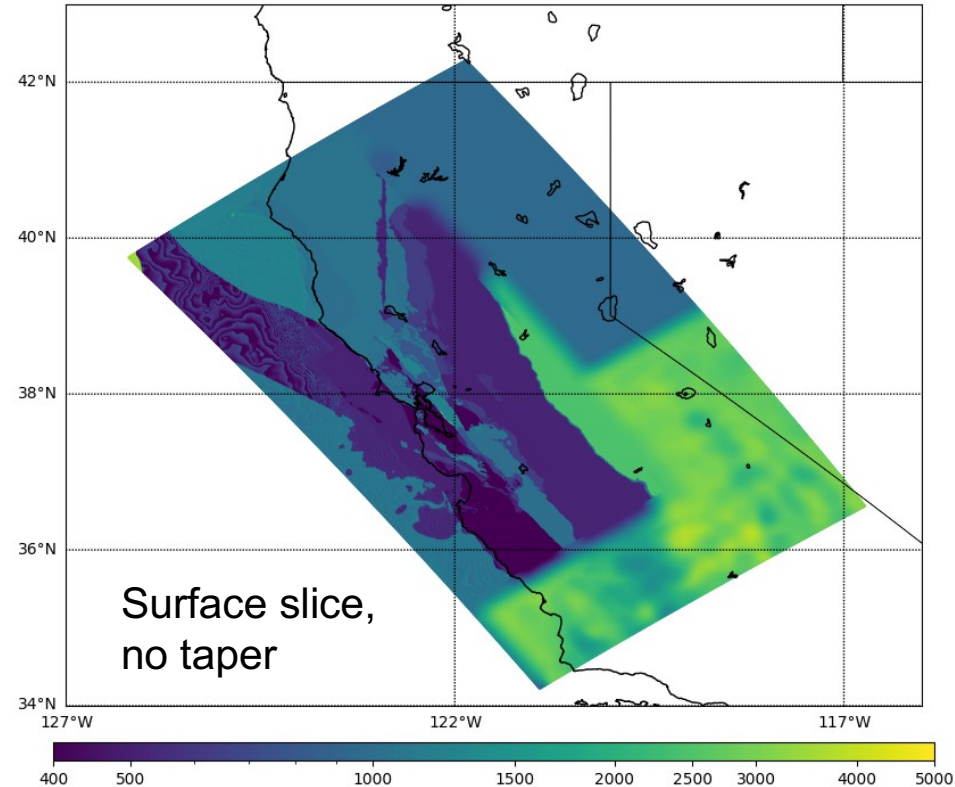


Northern California Velocity Model



- Model consists of 3 tiled models
 - USGS SFCVM, v21.1
 - CCA-06 (tomographic model)
 - 1D background model, based on Sierra geologic region in SFCVM
- Smoothing applied 20km from all interfaces
- Merged taper applied throughout model
- V_p/V_s ratio capped at 4
- Minimum V_s of 400 m/s

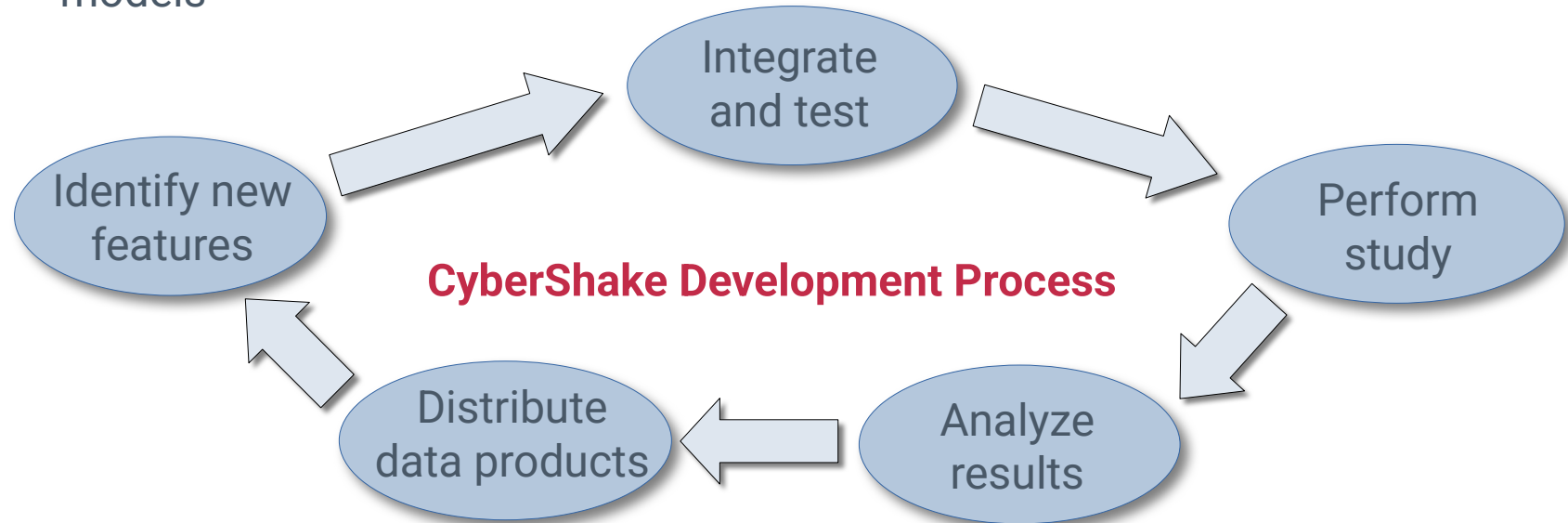
Merged Taper Impact





CyberShake Model Development

- Performs integrative science: advances in source descriptions, velocity models, and physics codes are integrated into the platform to improve hazard models





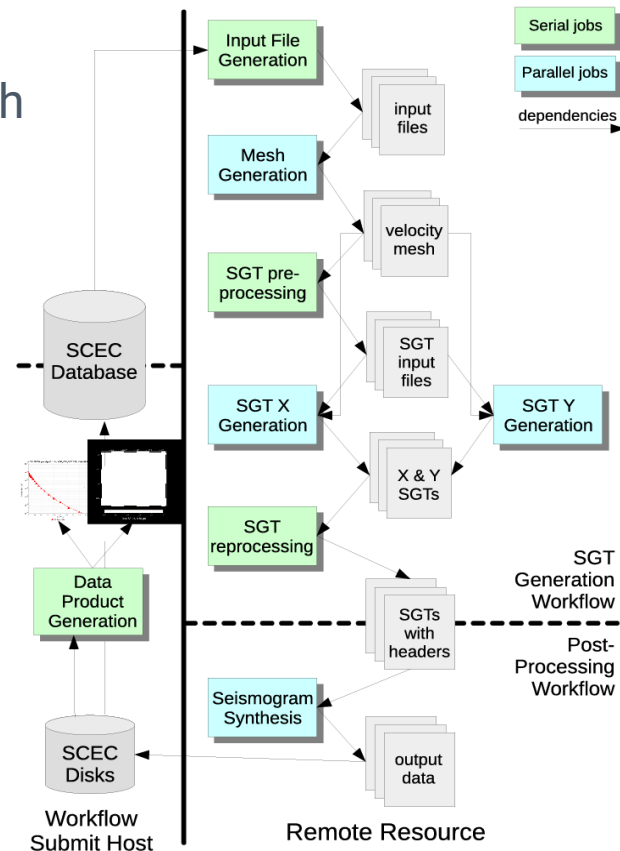
Computational Requirements

CyberShake Stage	Number of Tasks	Node-Hours	Output Data
Velocity mesh creation (parallel)	1	10 CPU	300 GB
AWP-ODC wave propagation (parallel)	3	200 GPU	1500 GB
Low-frequency seismogram synthesis (parallel)	1	230 CPU	38 GB
High-frequency seismogram synthesis (serial)	20,000	90 CPU	187 GB
Total, 1 site (including small jobs)	20,020	530	2025 GB
Total, full region	6.3 million	167,000	620 TB

- CyberShake runs require scientific workflows utilizing high performance computing resources

Scientific Workflow Stack

- Pegasus-WMS: describes workflow as directed graph of jobs with dependencies that make and use files
- HTCondor: manages queue of jobs and executes them on remote systems
- No changes to application codes
- Scientific workflows provide:
 - Automation
 - Data management
 - Job execution
 - Error recovery





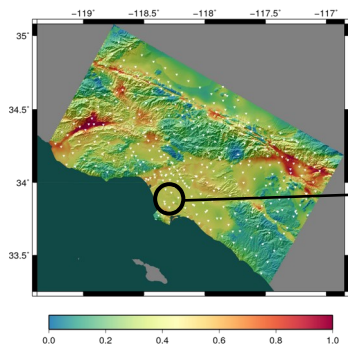
Study 24.8 (Northern California) Statistics

- September 24 – November 8 (45 days)
- Ran wave propagation calculations on OLCF *Frontier* and low-frequency synthesis and stochastic simulations on TACC *Frontera*
- Used about 180,000 node-hours, including up to 44% of *Frontier*
- Ran 27,800 jobs using Pegasus-WMS and HTCondor workflow tools
- Managed 1 PB of data
- Produced 36 TB / 9 million files of output data products
- Generated 126.8 million three-component seismograms and 34.3 billion IMs

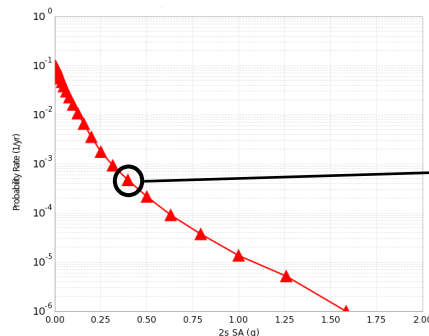


Layered Dataset of Hazard Products

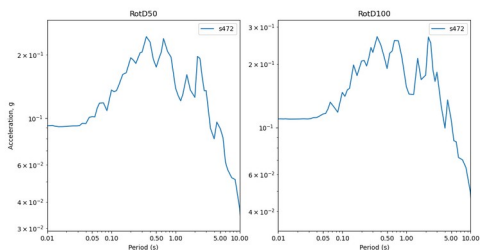
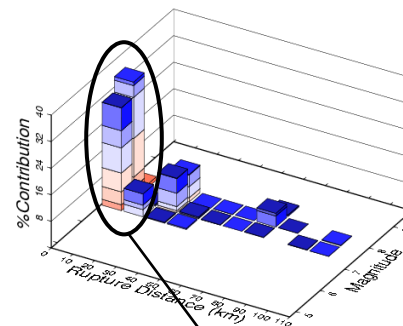
Hazard Maps
(multiple periods)



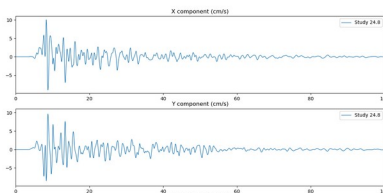
Hazard Curves
(100s of sites)



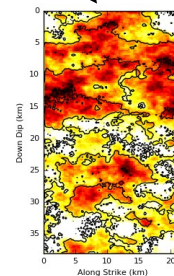
Disaggregation
(arbitrary level)



Intensity Measures
(100s per seismogram)



Seismograms
(~500,000 per site,
2- or 3-component)



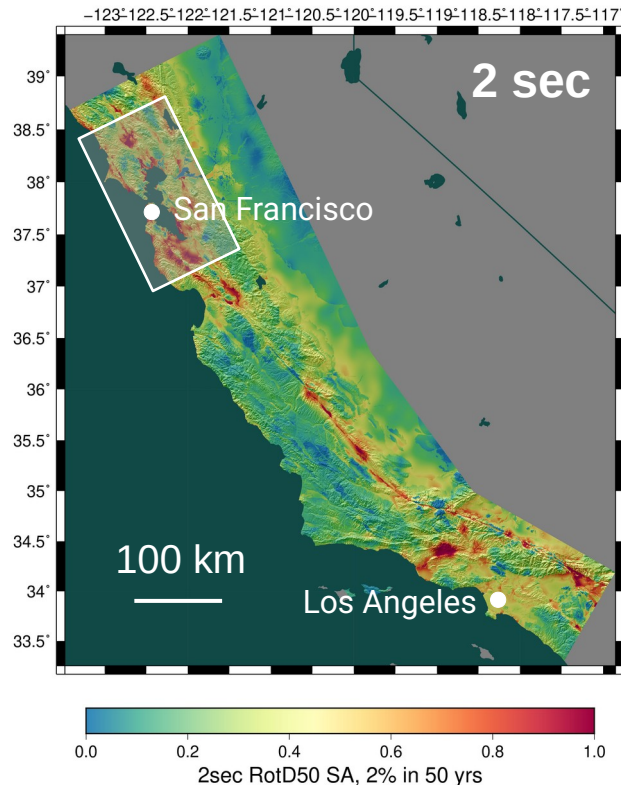
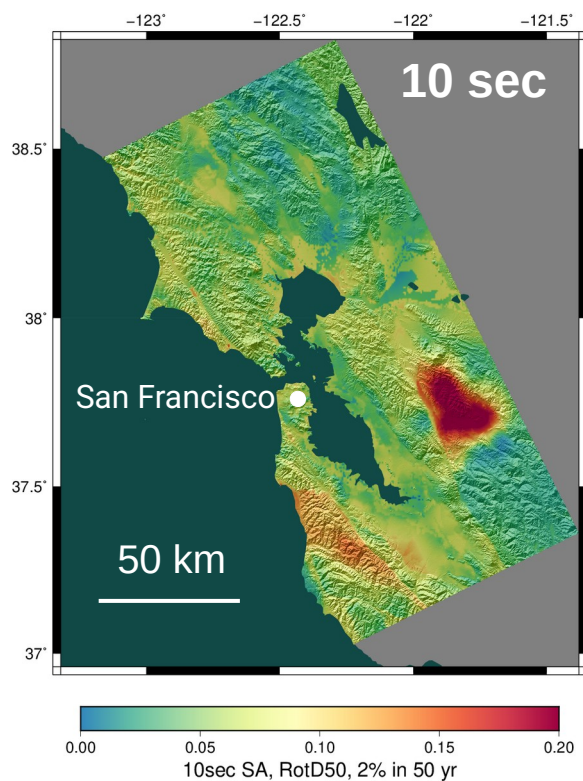
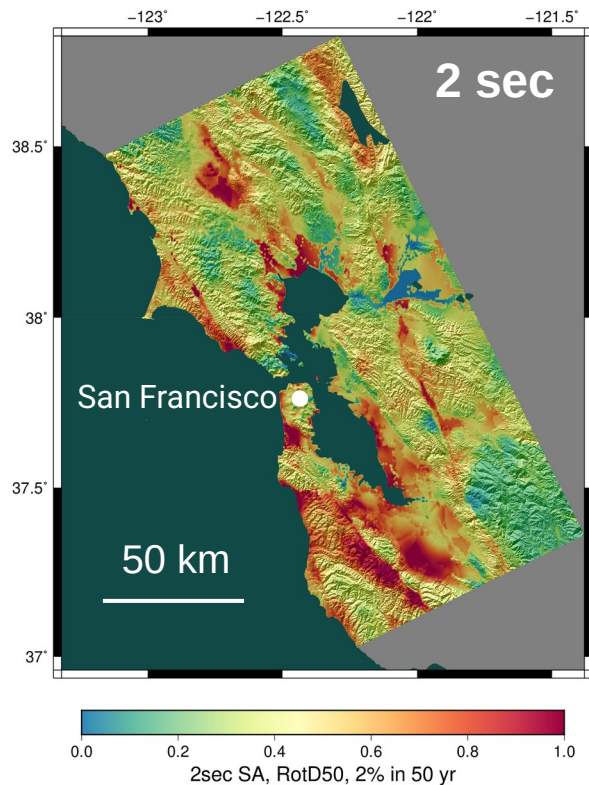
Earthquake Realizations
(~1M events per hazard model)



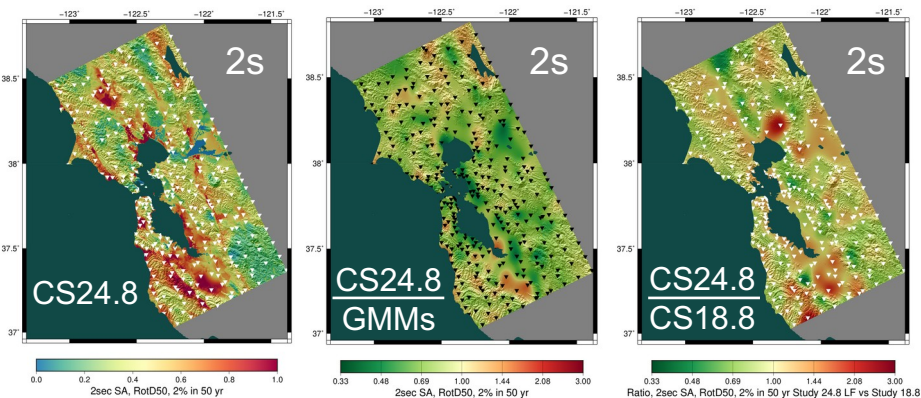
Data Products

- Seismograms for each event for each site
 - Base raw data product
- Peak shaking measures
 - RotD50 and RotD100 at dozens of periods, PGA, PGV
 - Vertical response spectra for 3-component seismograms
 - Subset (~25%) stored in relational database for quick access with the CyberShake Data Access Tool
- Durations
 - 5-75%, 5-95%, others
 - Period-dependent durations
 - ~25% stored in relational database
- Smoothed EAS (Effective Amplitude Spectra)
- Aggregate products: disaggregations, hazard curves, hazard maps
- You'll get to try data retrieval later in the workshop!

Example California Results

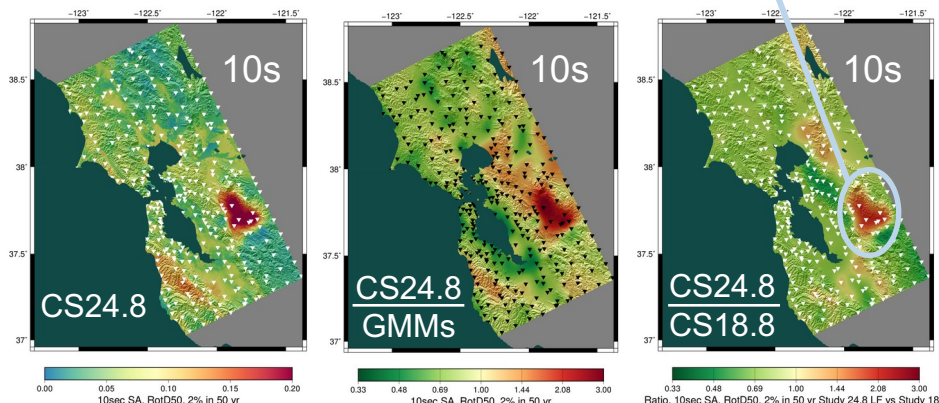
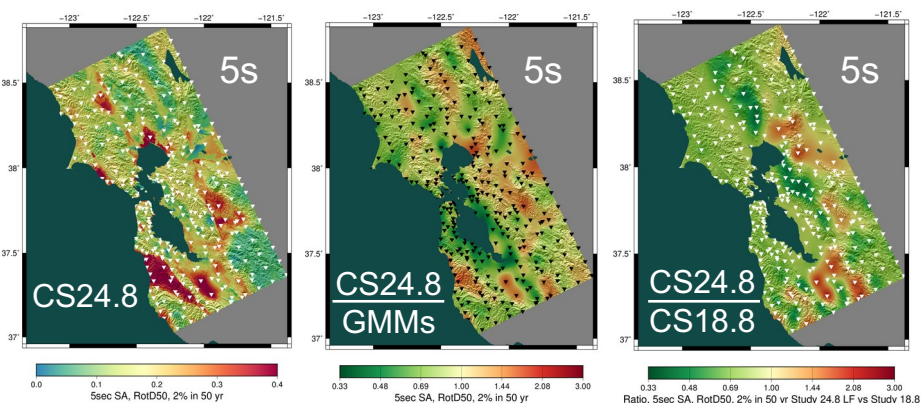
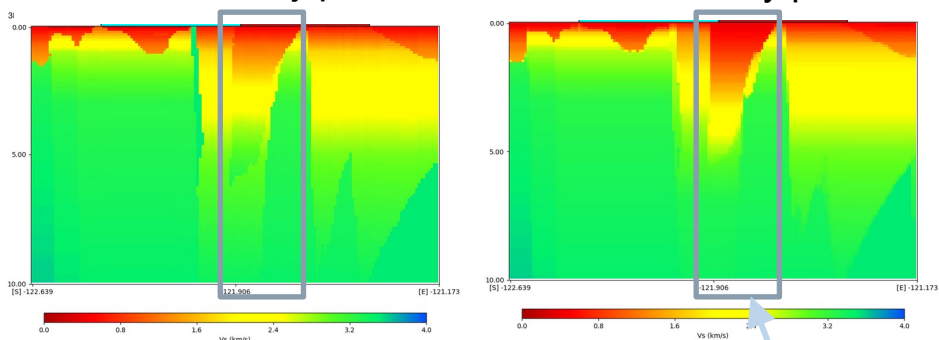


Low-frequency Hazard Map Comparisons



CS18.8 velocity profile

CS24.8 velocity profile



Validation of CyberShake

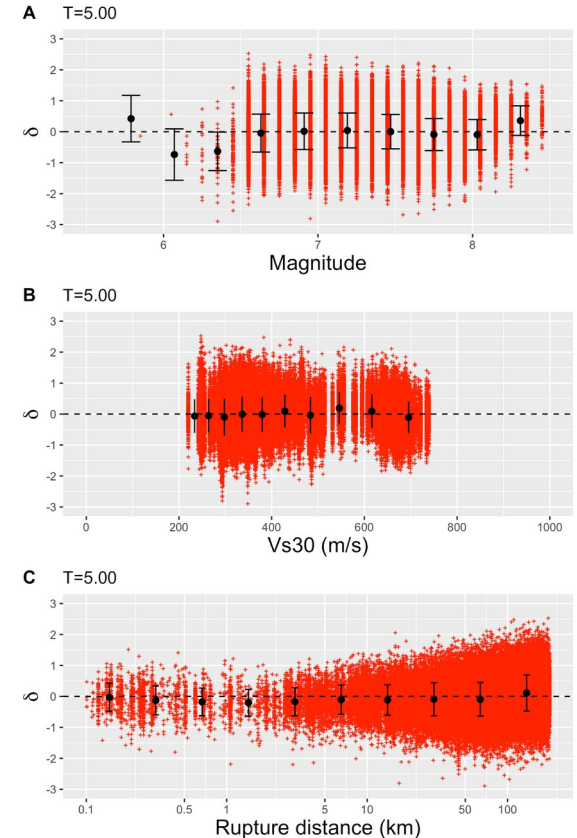
CyberShake-based Ergodic GMM

$$\ln(Y) = b_1 + b_2M + b_3M^2 + (b_4 + b_5M) \ln\left(\sqrt{R_{rup}^2 + h^2}\right) + b_6 \ln\left(\frac{V_{S30}^*}{800}\right) + b_7 \ln\left(\frac{Z_1 + 0.01}{Z_{1,ref} + 0.01}\right) + b_8 R_{rup} + \delta B_e + \delta W_{es}$$

$$V_{S30}^* = \begin{cases} V_{S30} & V_{S30} \leq 800 \text{ m/s} \\ 800 & V_{S30} > 800 \text{ m/s} \end{cases}$$

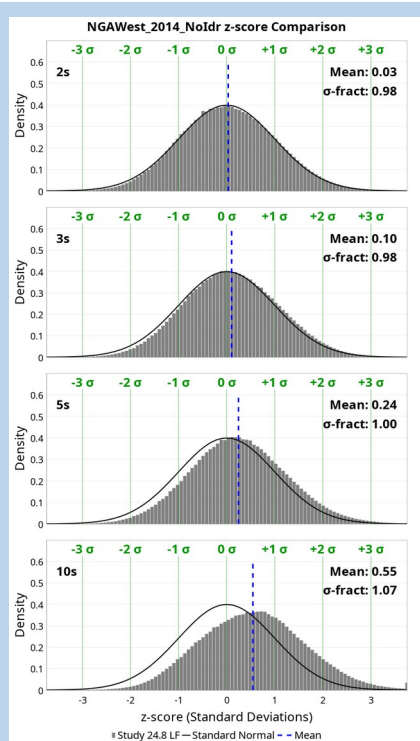
δB_e : the between-event residual (event term)

δW_{es} : the within-event residual



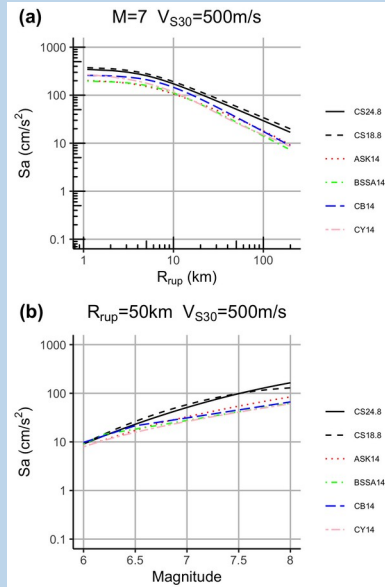


Low-frequency Aggregate Analyses

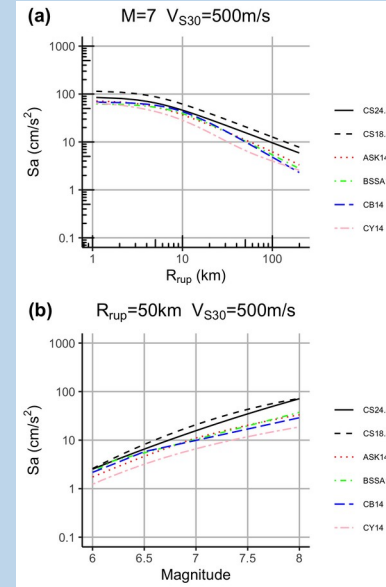


Z-score distribution

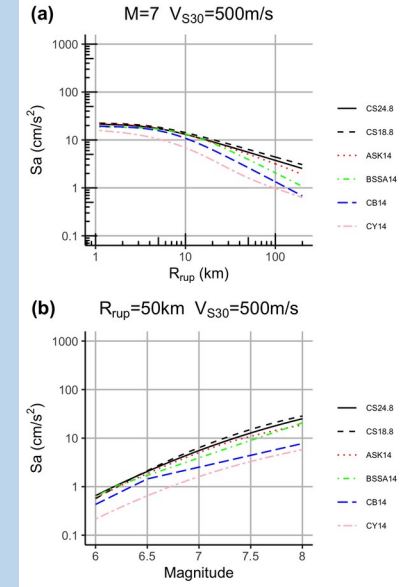
2 sec



5 sec

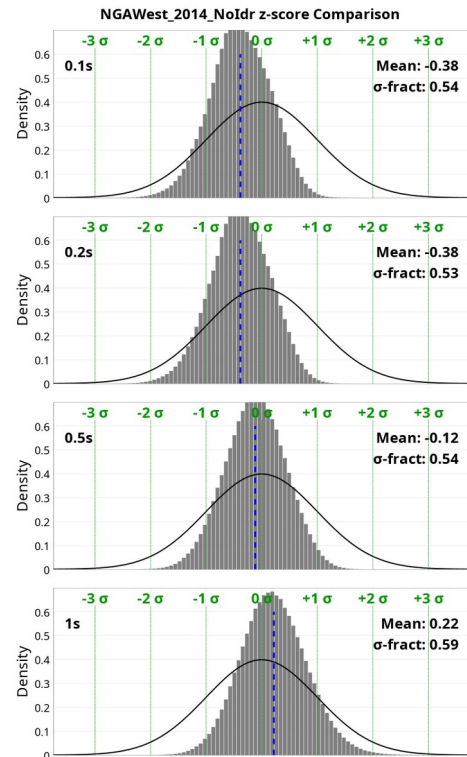
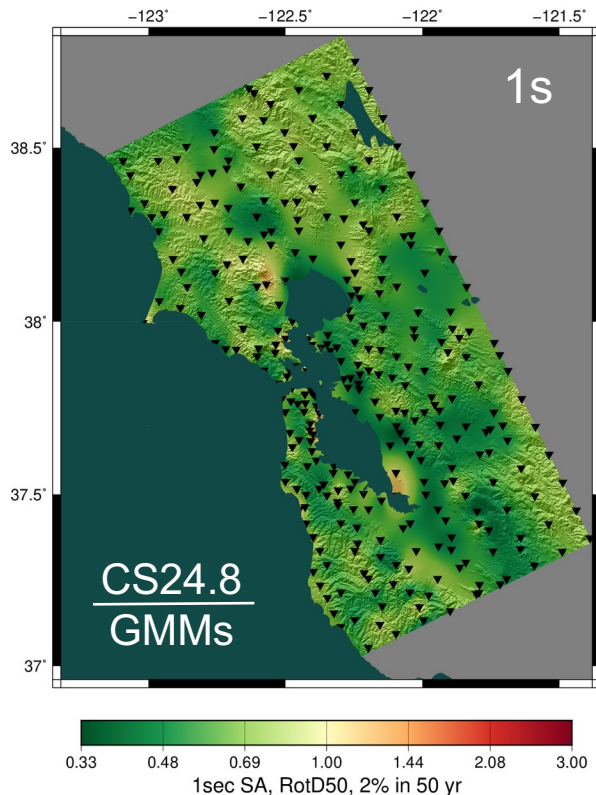
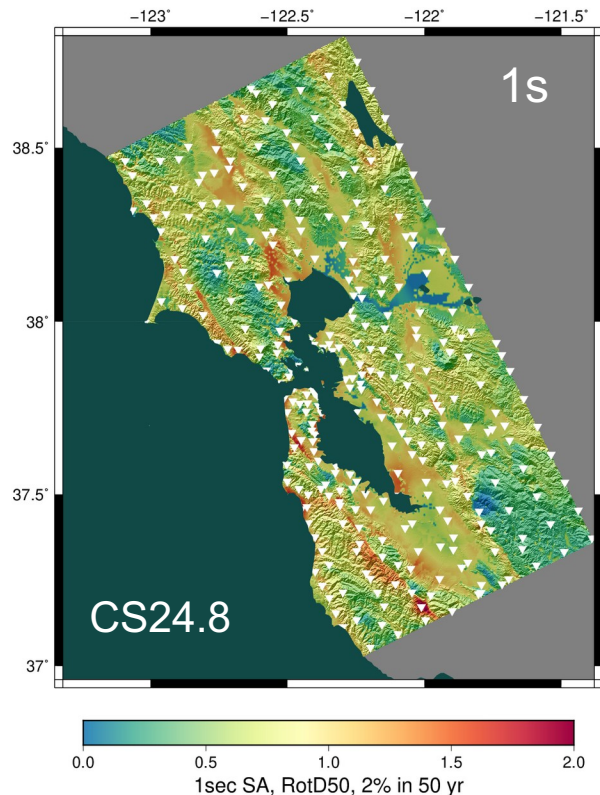


10 sec



Comparison of distance and magnitude scaling between CyberShake-derived GMM and other GMMs

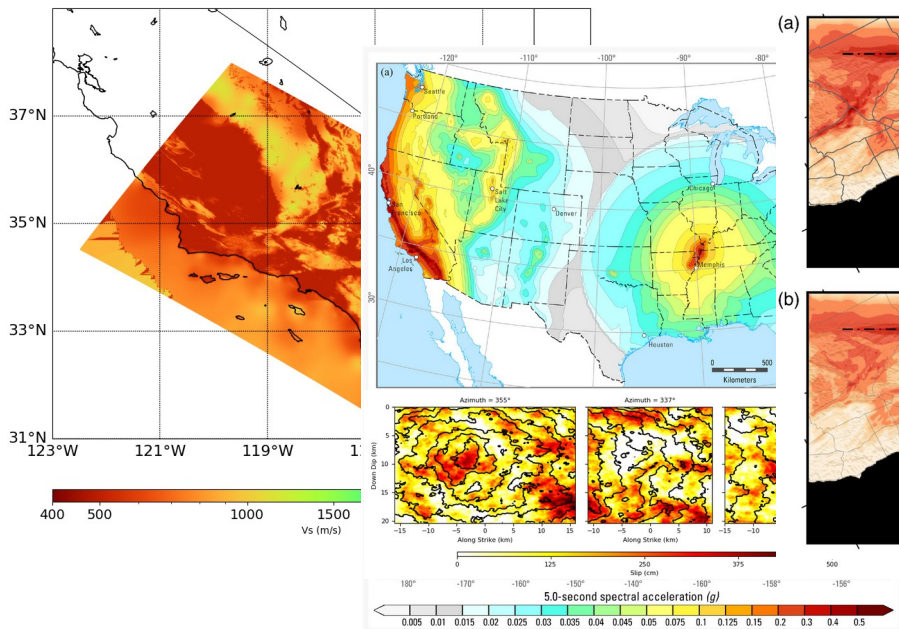
Broadband Data Products



Z-score distribution

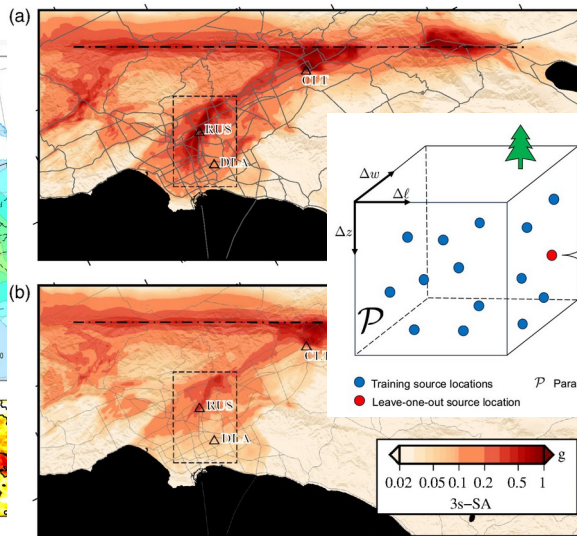
Next Steps

Updated velocity models



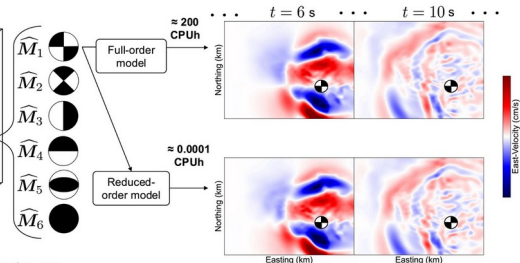
New ERFs with multi-fault ruptures

Nonlinearity



Roten et al., 2023

Surrogate models



Rekoske et al., 2025

...as well as technical improvements to support all this!

CyberShake Applications

- Seismic hazard analysis
- Earthquake early warning
- Examining spatial correlations
- Performance of distributed infrastructure
- Basin amplifications for the US national seismic hazard model
- Response history analysis
- Any application needing a large dataset of ground motions

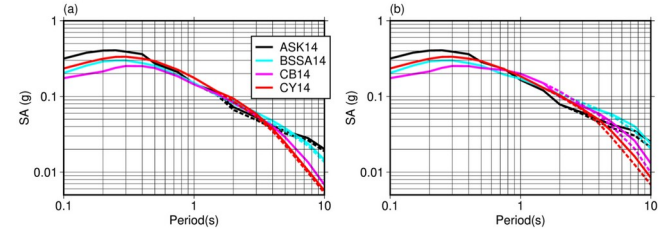


Figure 16. Comparison of median response spectra from the empirical GMMs (dashed lines) and CyberShake-adjusted GMMs (solid lines), for example, site conditions corresponding to (a) shallow basin depths, $Z_1 = 0.45$ km, $Z_{2.5} = 3$ km (i.e. $\delta Z_1 \approx 0$ km) and a (b) deeper basin site, $Z_1 = 0.8$ km, $Z_{2.5} = 5.5$ km. Scenario ground motions were computed for $M_W = 7.5$, $R_{rup} = 50$ km, $R_B = 50$ km, $R_X = 50$ km, width = 15 km, dip = 90° , $Z_{hyp} = 7.5$ km, $Z_{TOR} = 0$ km. Reference Z_1 values from NGA-West-2 GMMs are $Z_{1,ref}^{ASK14}(V_{S30} = 260 \text{ m/s}) = 0.475$ km, $Z_{1,ref}^{BSSA14}(V_{S30} = 260 \text{ m/s}) = 0.485$ km, and $Z_{1,ref}^{CY14}(V_{S30} = 260 \text{ m/s}) = 0.485$ km.

Moschetti et al. 2024

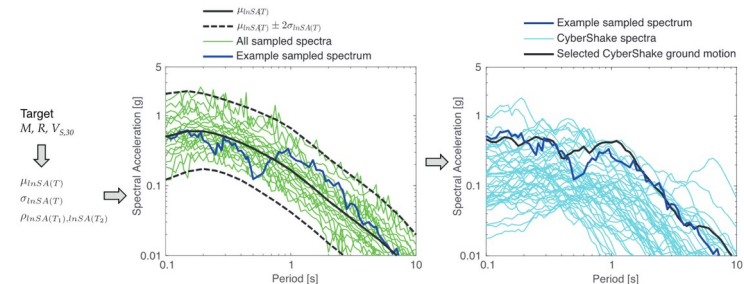


Figure 5. Schematic illustration of the ground-motion selection process. We first specify a target scenario and compute the means, standard deviations, and correlations of $\ln SA$ (left); we then sample response spectra from the resulting distribution (middle); finally, we compare each sampled response spectrum to the CyberShake spectra and select the closest match (right).

Baker et al. 2021



Questions?

