Solving Prediction Problems in Earthquake System Science on Blue Waters

Thomas H. Jordan [1]

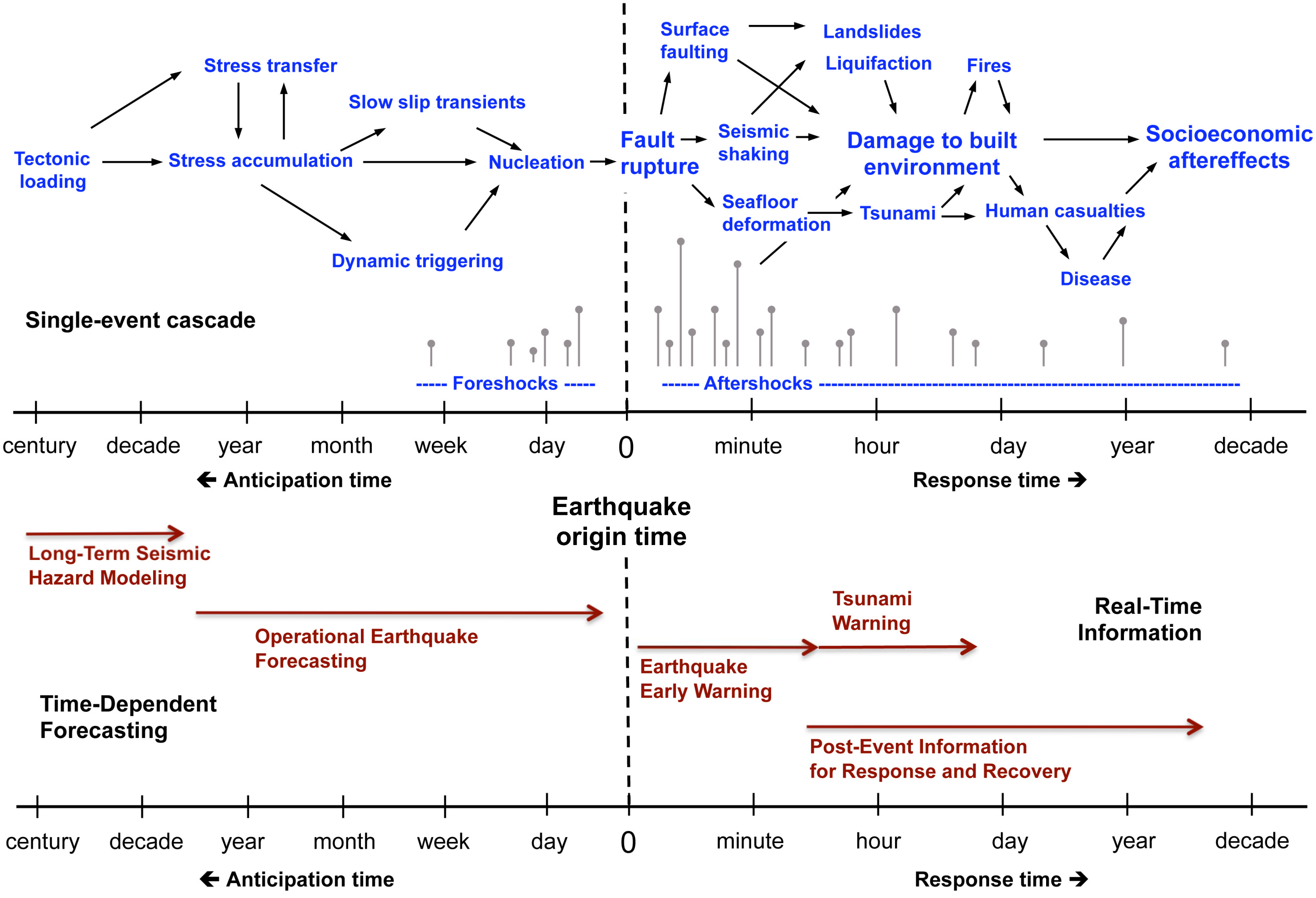
Scott Callaghan [1], **Robert Graves** [2], **Kim Olsen** [3], **Yifeng Cui** [4], **Jun Zhou** [4], **Efecan Poyraz** [4], **Philip J Maechling** [1], **David Gill** [1], **Kevin Milner** [1], **Omar Padron, Jr.** [5], **Gregory H. Bauer** [5], **Timothy Bouvet** [5], **William T. Kramer** [5], **Gideon Juve** [6], **Karan Vahi** [6] **& Ewa Deelman** [6]

[1] **Southern California Earthquake Center,** [2] **USGS,** [3] **San Diego State University,** [4] **San Diego Supercomputer Center,** [5] **National Center for Supercomputing Applications,** [6] **Information Sciences Institute**

Blue Waters Symposium

14 May 2014

Prediction Problems of Earthquake System Science



Single-event cascade

Tectonic loading → Stress accumulation → Nucleation → Fault rupture

Stress transfer

Slow slip transients

Dynamic triggering

Surface faulting

Seismic shaking

Seafloor deformation

Landslides

Liquifaction

Fires

Damage to built environment

Tsunami

Human casualties

Disease

Socioeconomic aftereffects

----- Foreshocks -----

------ Aftershocks --

century decade year month week day 0 minute hour day year decade

← Anticipation time

Response time →

Earthquake origin time

Long-Term Seismic Hazard Modeling

Operational Earthquake Forecasting

Earthquake Early Warning

Tsunami Warning

Post-Event Information for Response and Recovery

Time-Dependent Forecasting

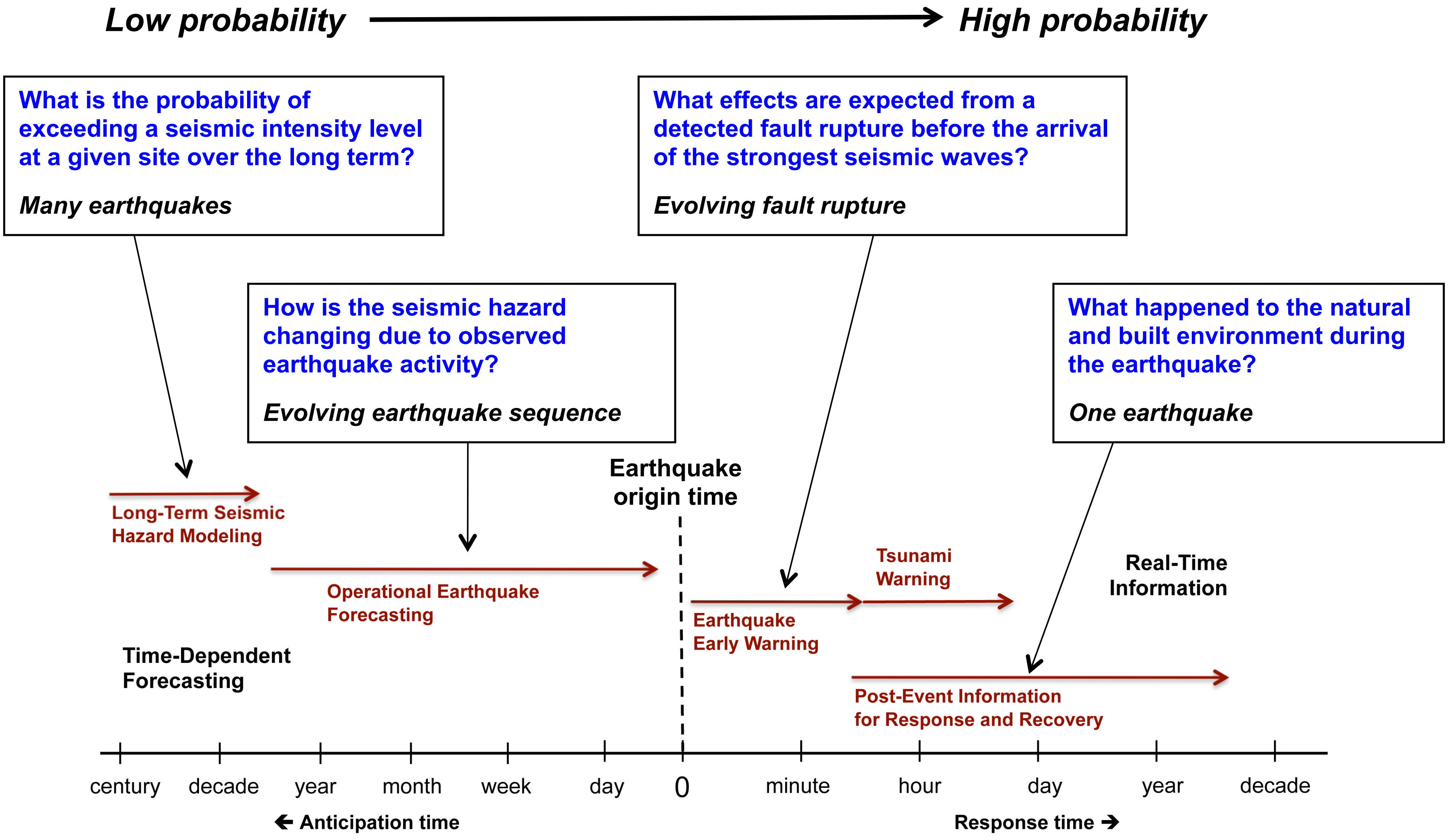
Real-Time Information

century decade year month week day 0 minute hour day year decade

← Anticipation time

Response time →

Prediction Problems of Earthquake System Science



Low probability → ***High probability***

What is the probability of exceeding a seismic intensity level at a given site over the long term?

Many earthquakes

How is the seismic hazard changing due to observed earthquake activity?

Evolving earthquake sequence

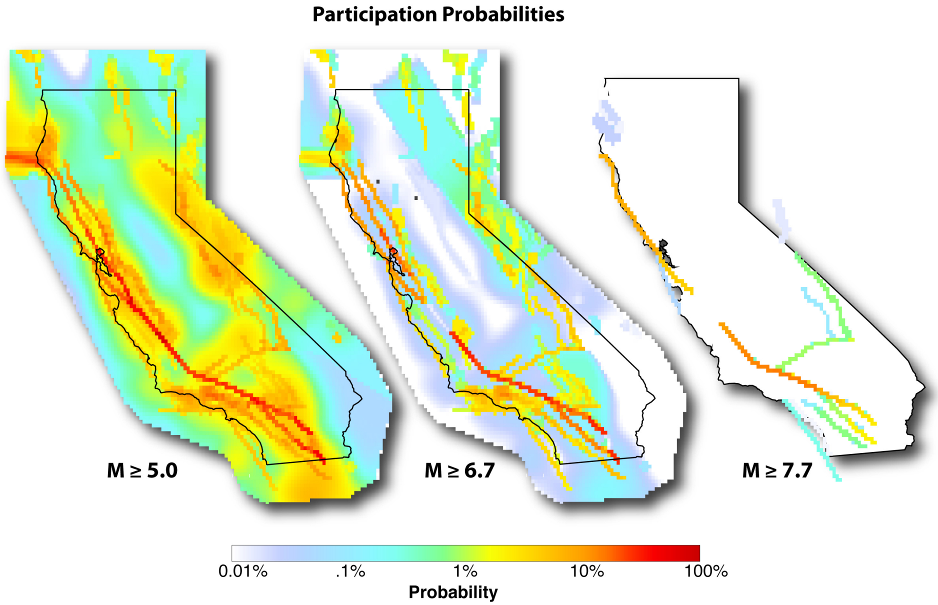
What effects are expected from a detected fault rupture before the arrival of the strongest seismic waves?

Evolving fault rupture

What happened to the natural and built environment during the earthquake?

One earthquake

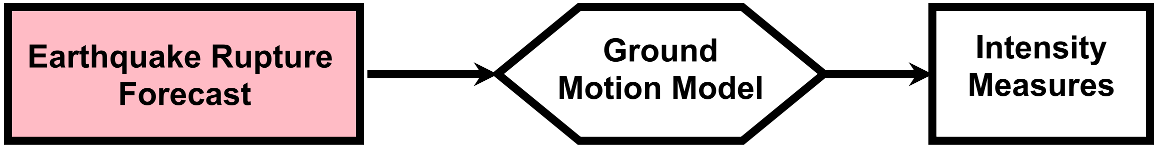
Probabilistic Seismic Hazard Model



Participation Probabilities

SCEC-USGS-CGS Working Group on California Earthquake Probabilities (2008)

Uniform California Earthquake Rupture Forecast (UCERF2)



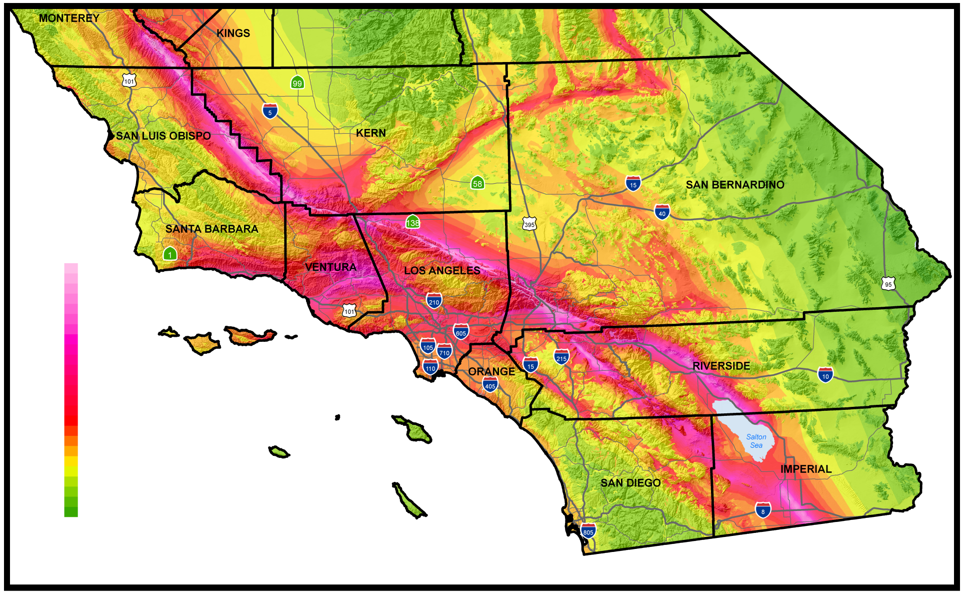
Earthquake Rupture Forecast → **Ground Motion Model** → **Intensity Measures**

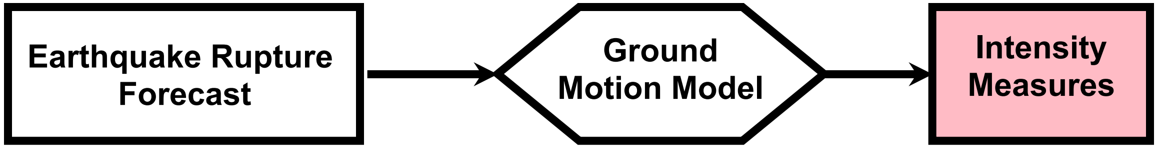
$P(S_k)$ $P(Y_n \mid S_k)$ $P(Y_n)$

Probabilistic Seismic Hazard Model

National Seismic Hazard Map

PGA (%g) with 2% Probability of Exceedance in 50 years

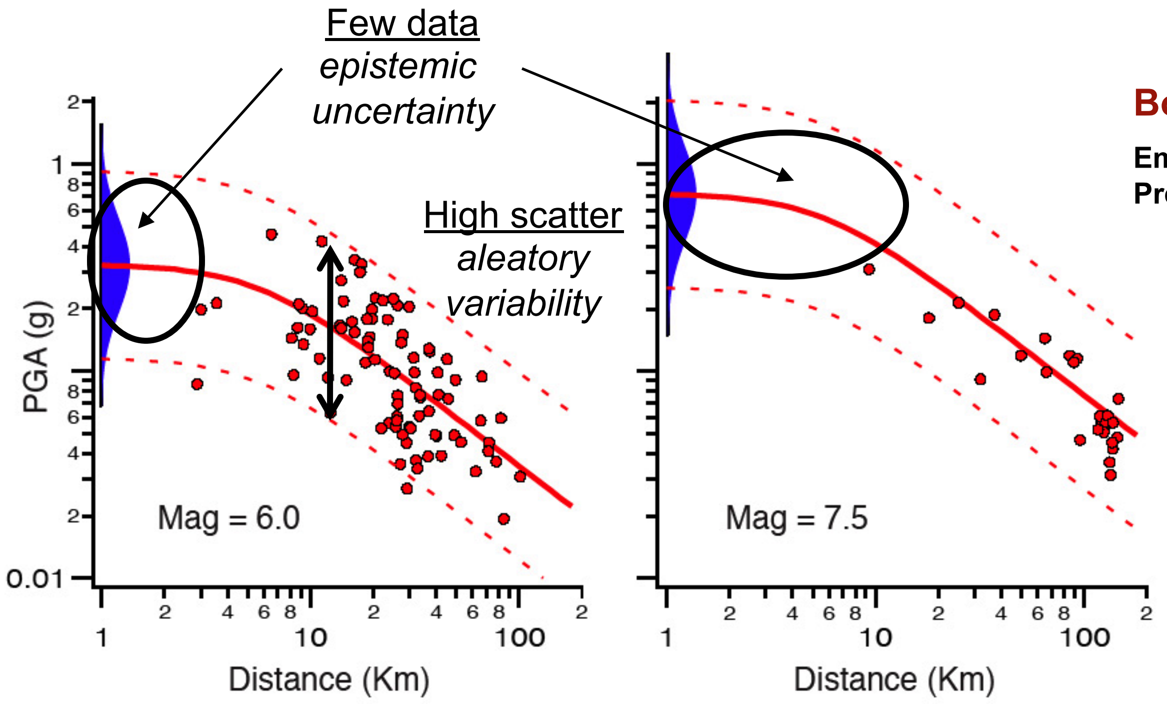




Earthquake Rupture Forecast → **Ground Motion Model** → **Intensity Measures**

$P(S_k)$ $P(Y_n \mid S_k)$ $P(Y_n)$

Probabilistic Seismic Hazard Model

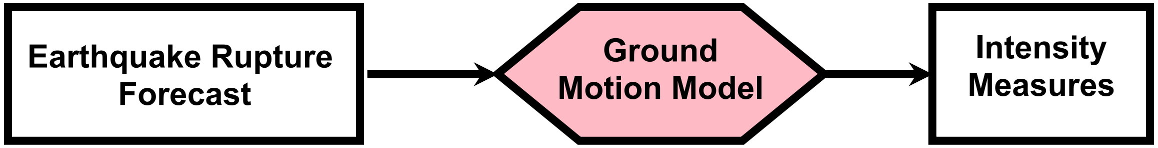


Few data
epistemic
uncertainty

High scatter
aleatory
variability

Boore et al. (1997)

Empirical Ground Motion Prediction Equations (GMPEs)



Earthquake Rupture Forecast

Ground Motion Model

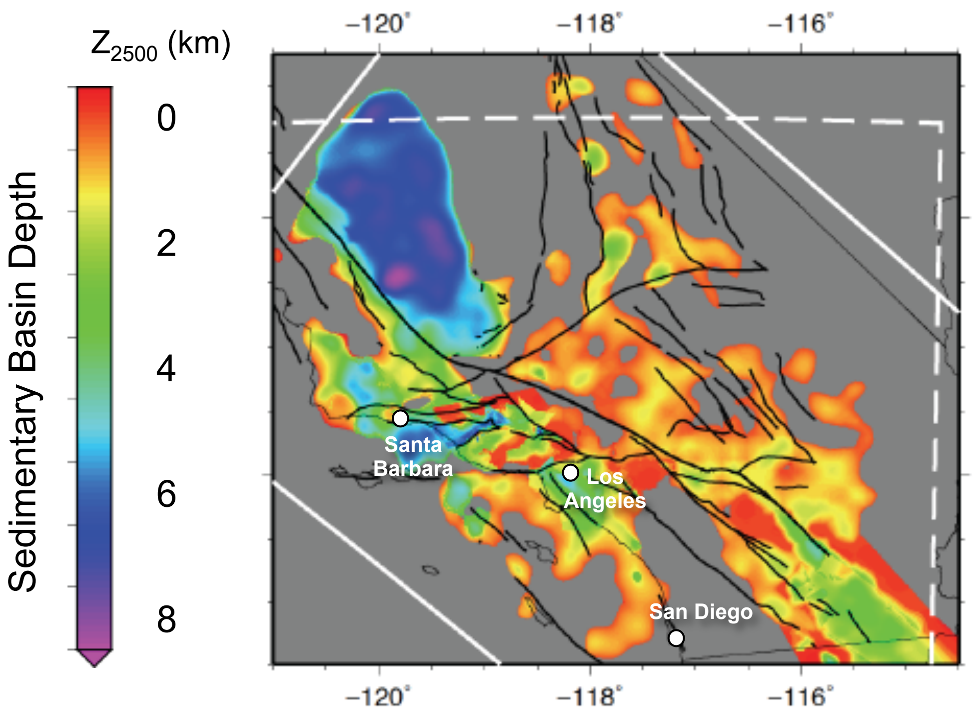
Intensity Measures

$P(S_k)$

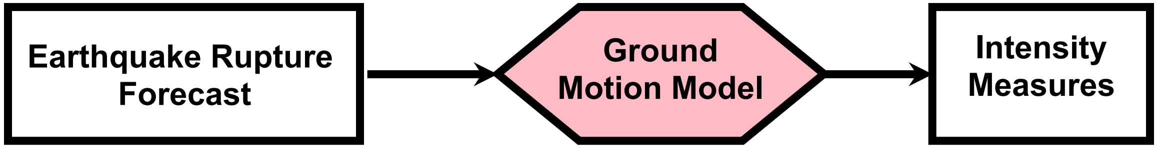
$P(Y_n \mid S_k)$

$P(Y_n)$

Probabilistic Seismic Hazard Model



Much of the aleatory variability and epistemic uncertainty in the GMPEs comes from 3D heterogeneity in crustal structure

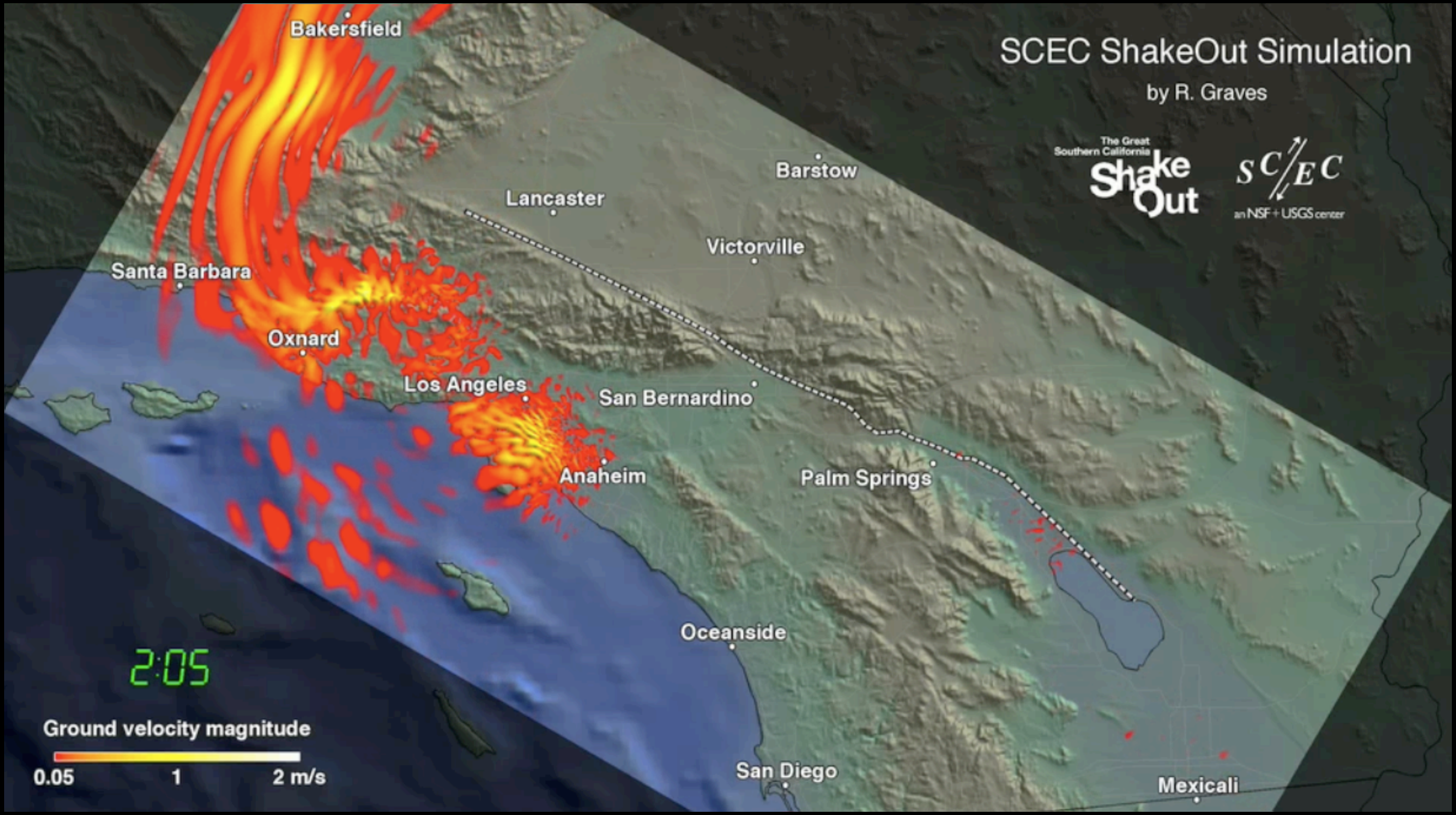


Earthquake Rupture Forecast → **Ground Motion Model** → **Intensity Measures**

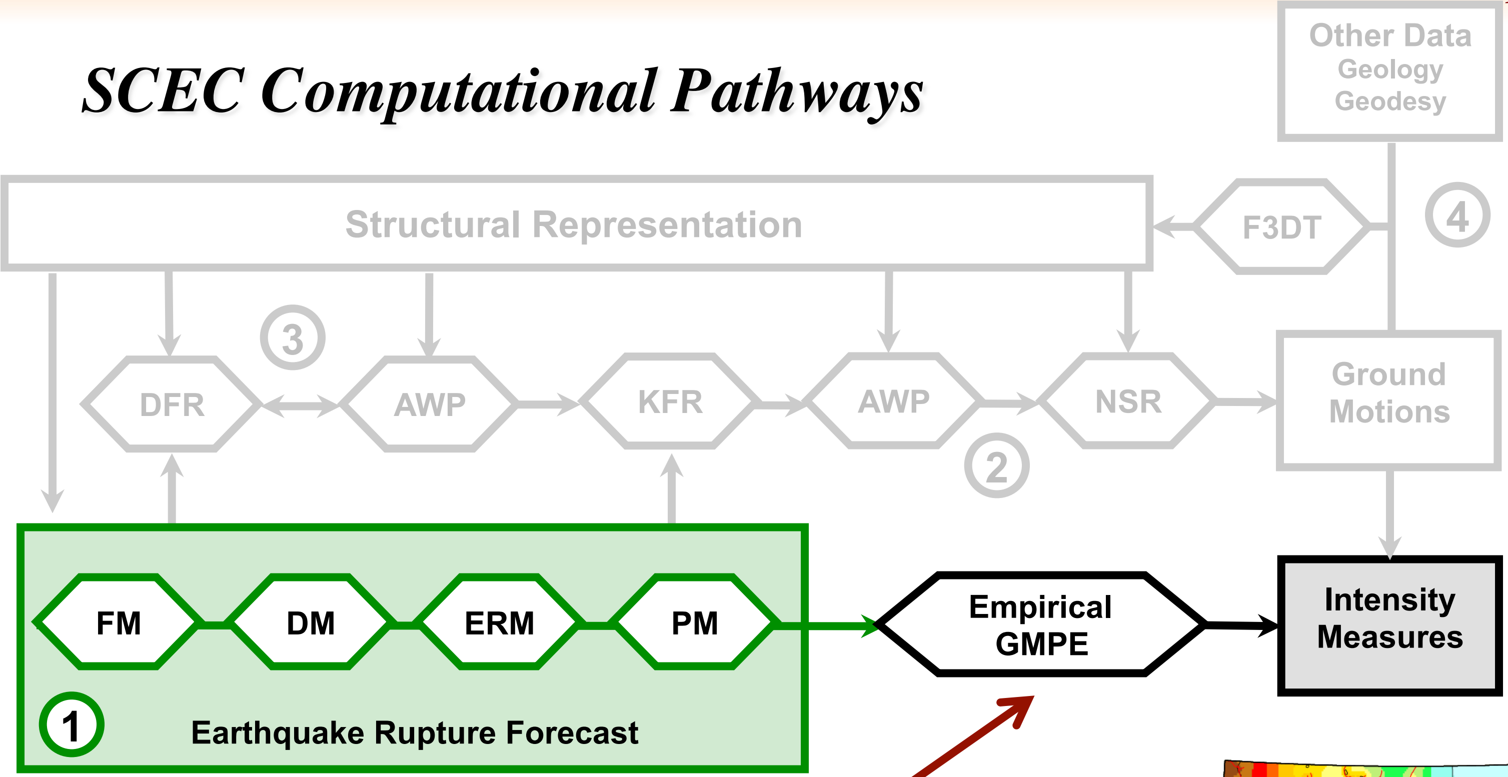
$P(S_k)$ $P(Y_n \mid S_k)$ $P(Y_n)$

ShakeOut Scenario

M7.8 Earthquake on Southern San Andreas Fault



SCEC Computational Pathways



Other Data
Geology
Geodesy

Structural Representation

F3DT

④

③

DFR

AWP

KFR

AWP

NSR

Ground Motions

②

FM

DM

ERM

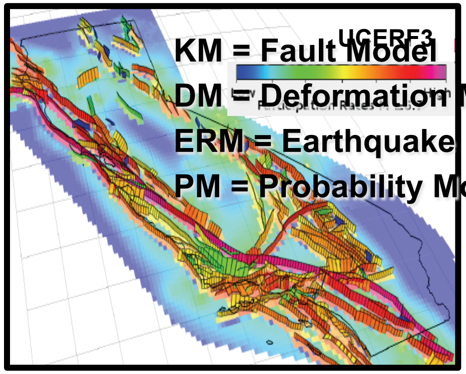
PM

Empirical GMPE

Intensity Measures

① Earthquake Rupture Forecast

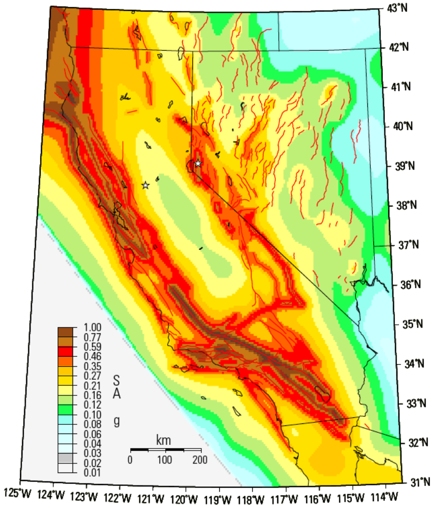
TACC *Stampede*



KM = Fault Model
DM = Deformation Model
ERM = Earthquake Rate Model
PM = Probability Model

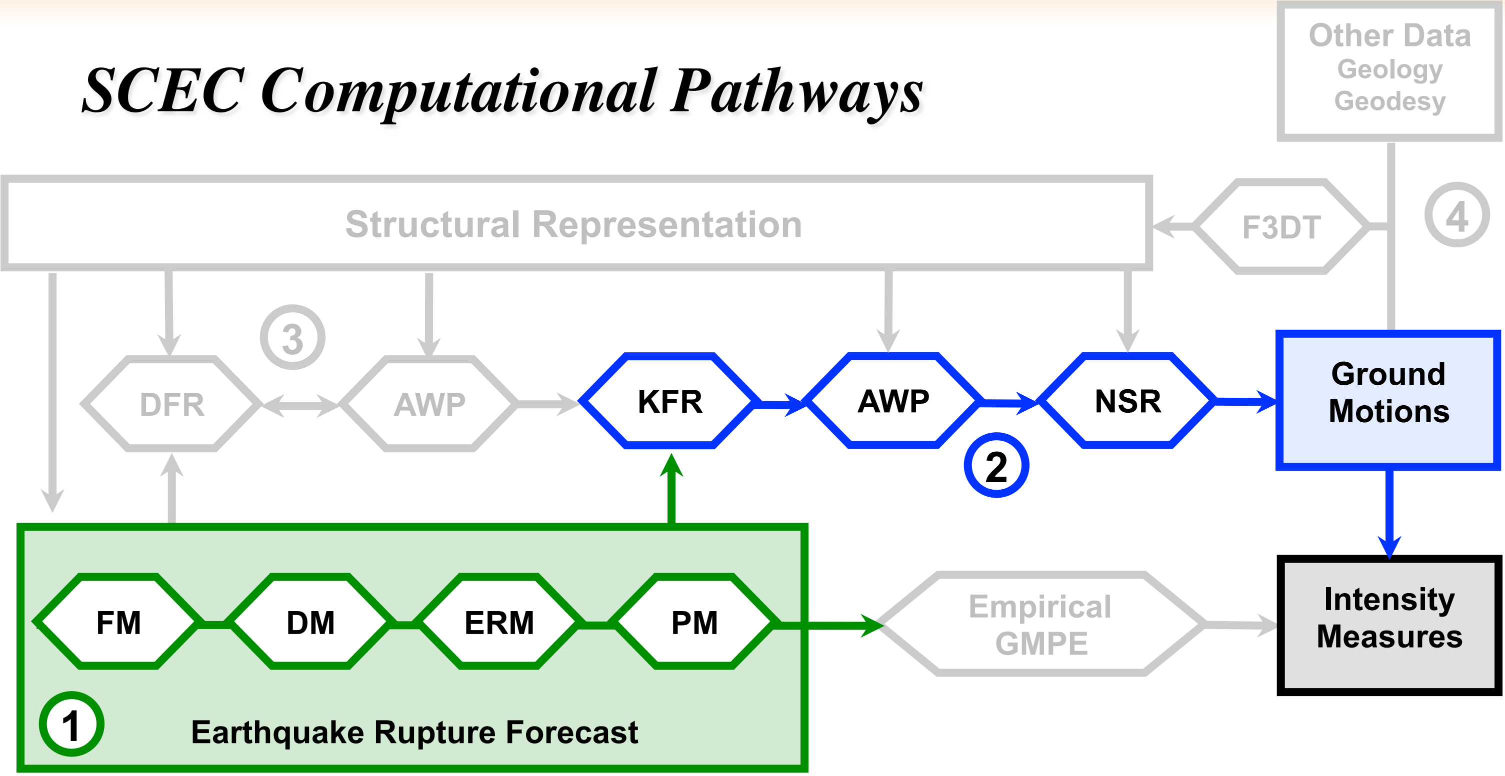
① Uniform California Earthquake Rupture Forecast (UCERF3)

Main goal of our *Blue Waters* research is to replace the empirical GMPEs with physics-based ground motion models



2014 National Seismic Hazard Maps

SCEC Computational Pathways



Other Data
Geology
Geodesy

Structural Representation

F3DT

④

③

DFR

AWP

KFR

AWP

NSR

②

Ground Motions

FM

DM

ERM

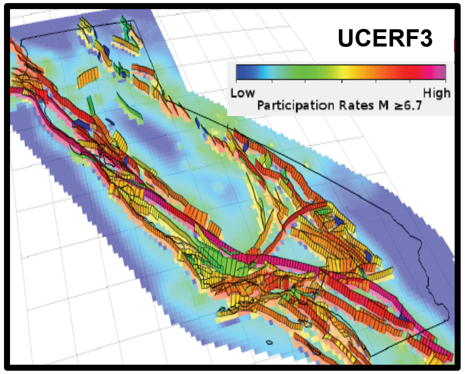
PM

Empirical GMPE

Intensity Measures

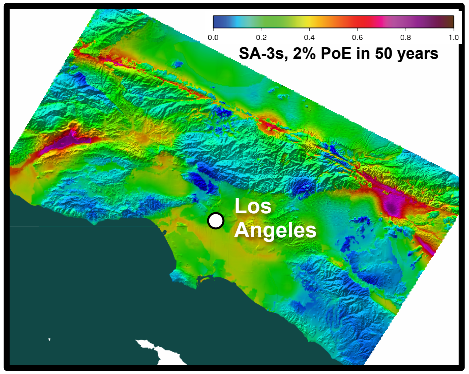
① Earthquake Rupture Forecast

TACC *Stampede*



① **Uniform California Earthquake Rupture Forecast (UCERF3)**

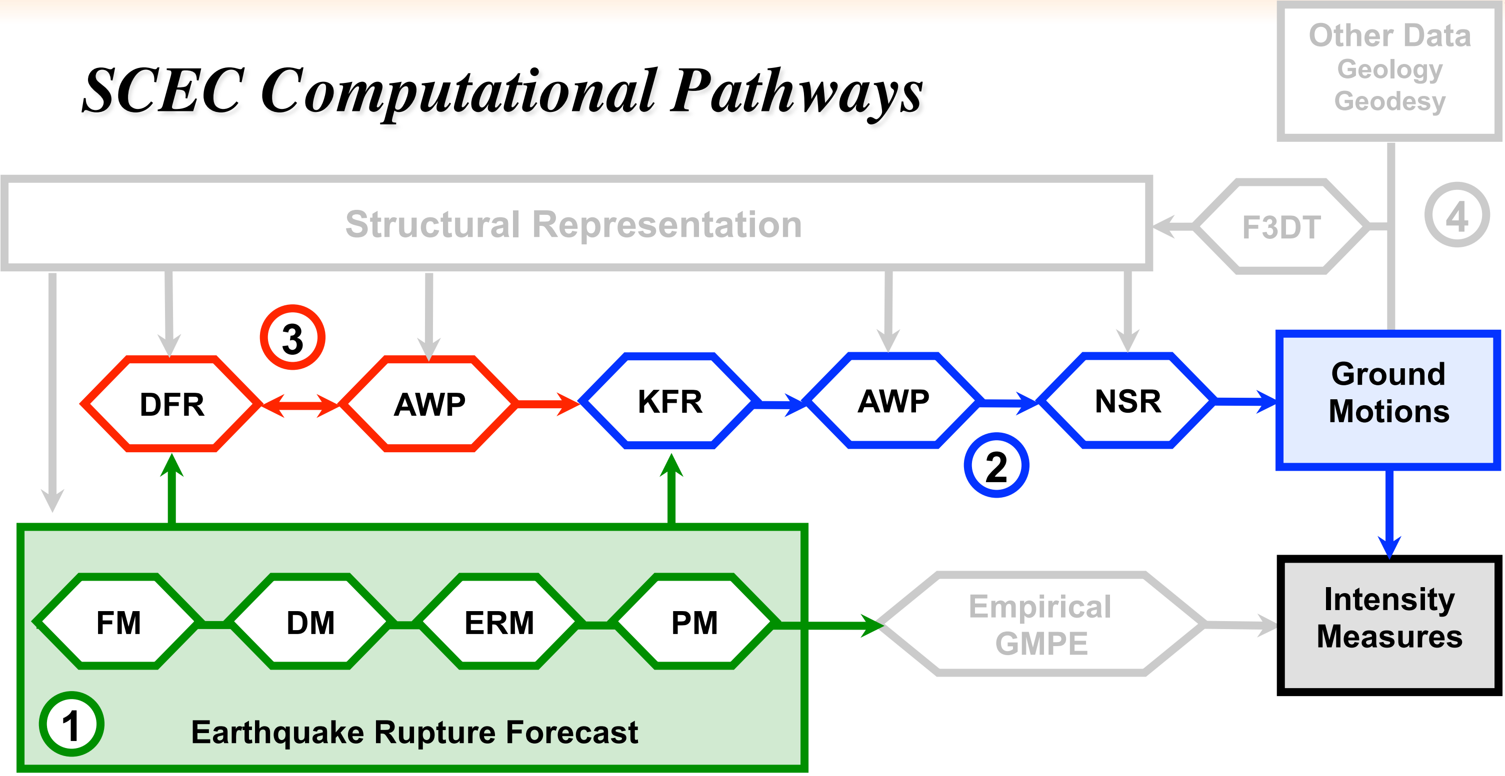
NCSA *Blue Waters*



② **CyberShake 14.2 seismic hazard model for LA region**

KFR = Kinematic Fault Rupture
AWP = Anelastic Wave Propagation
NSR = Nonlinear Site Response
DFR = Dynamic Fault Rupture
F3DT = Full-3D Tomography

SCEC Computational Pathways



Other Data
Geology
Geodesy

④

Structural Representation

F3DT

③

DFR

AWP

KFR

AWP

NSR

②

Ground Motions

FM

DM

ERM

PM

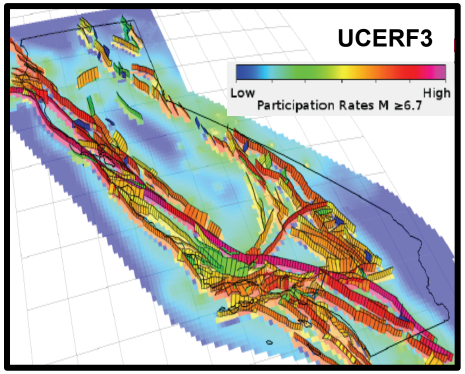
Empirical GMPE

Intensity Measures

①

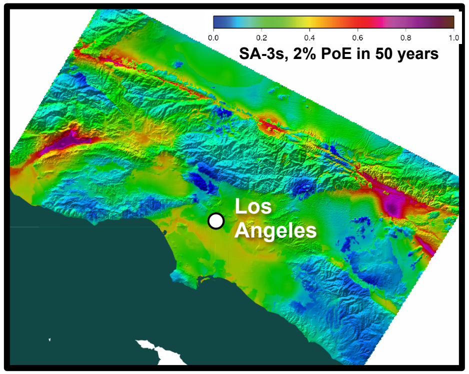
Earthquake Rupture Forecast

TACC *Stampede*



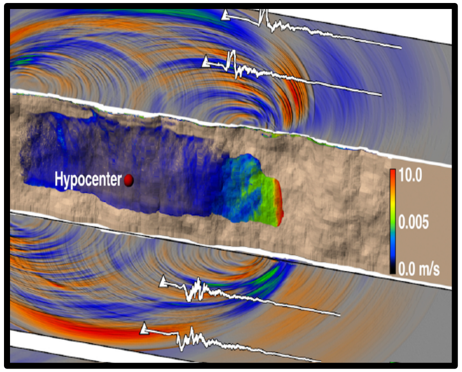
① Uniform California Earthquake Rupture Forecast (UCERF3)

NCSA *Blue Waters*



② CyberShake 14.2 seismic hazard model for LA region

OLCF *Titan*



③ Dynamic rupture model of fractal roughness on SAF

KFR = Kinematic Fault Rupture

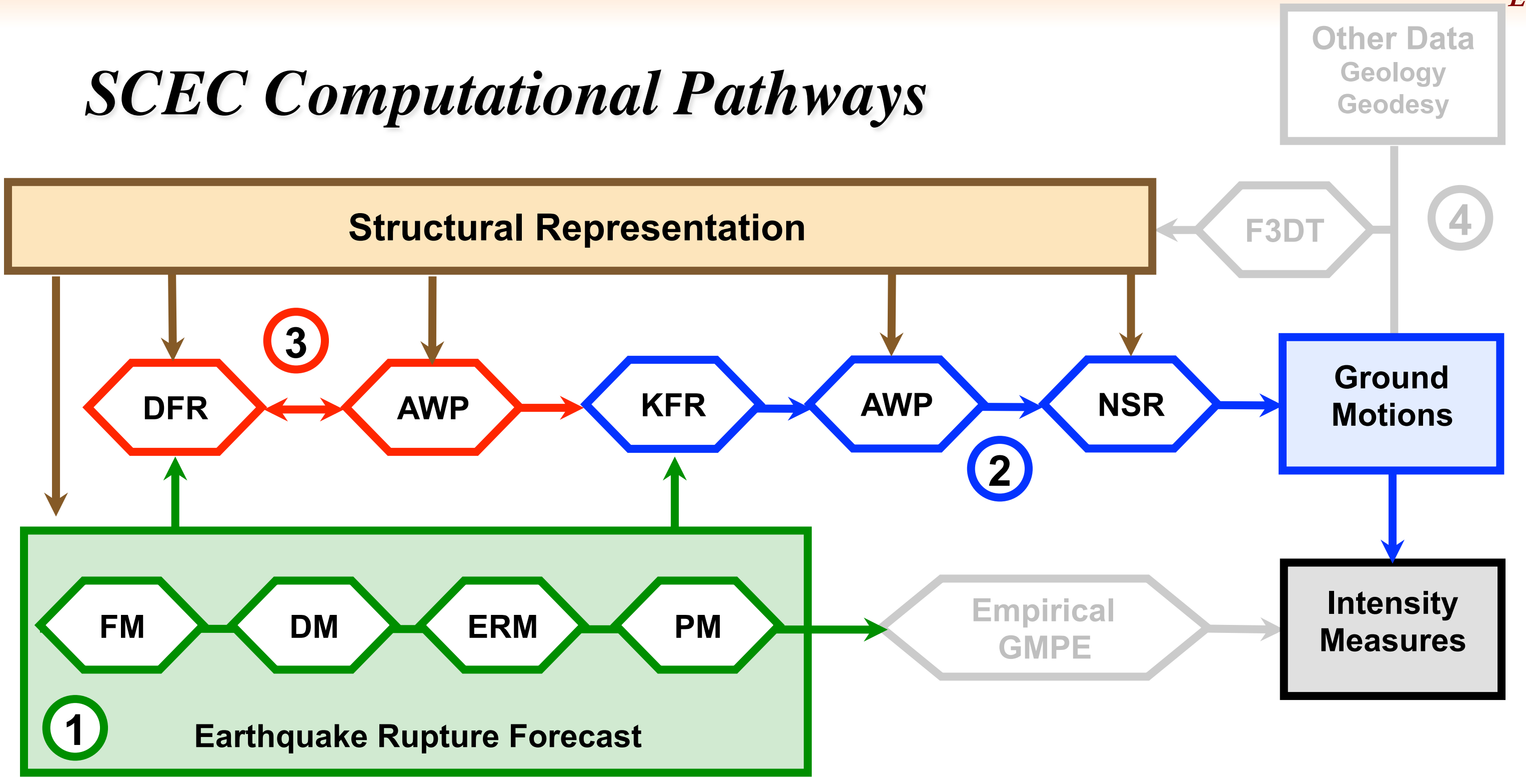
AWP = Anelastic Wave Propagation

NSR = Nonlinear Site Response

DFR = Dynamic Fault Rupture

F3DT = Full-3D Tomography

SCEC Computational Pathways



Other Data
Geology
Geodesy

Structural Representation

F3DT

④

③

DFR

AWP

KFR

AWP

NSR

Ground Motions

②

FM

DM

ERM

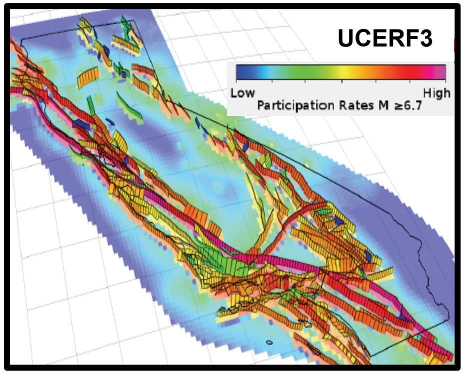
PM

① Earthquake Rupture Forecast

Empirical GMPE

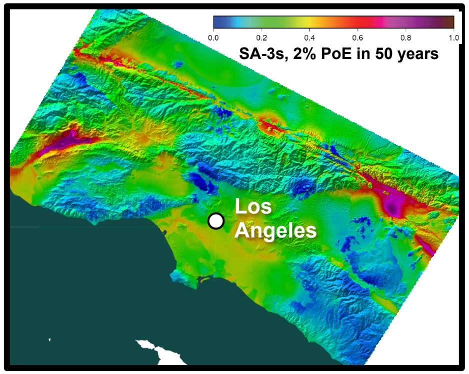
Intensity Measures

TACC *Stampede*



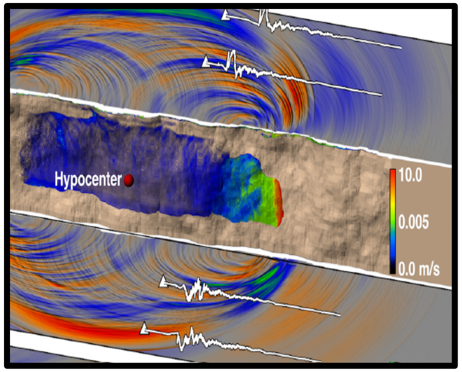
① **Uniform California Earthquake Rupture Forecast (UCERF3)**

NCSA *Blue Waters*



② **CyberShake 14.2 seismic hazard model for LA region**

OLCF *Titan*



③ **Dynamic rupture model of fractal roughness on SAF**

KFR = Kinematic Fault Rupture

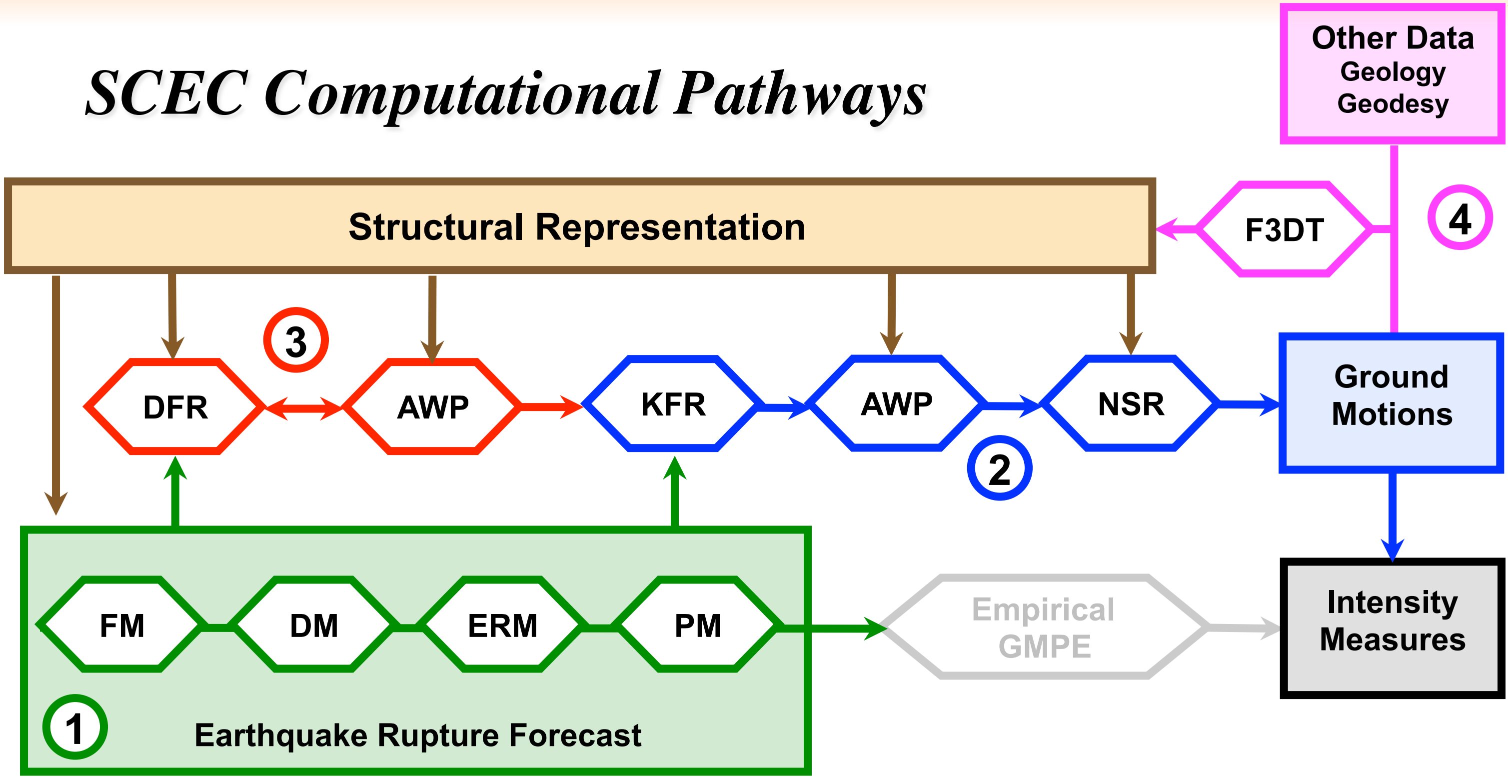
AWP = Anelastic Wave Propagation

NSR = Nonlinear Site Response

DFR = Dynamic Fault Rupture

F3DT = Full-3D Tomography

SCEC Computational Pathways



Other Data
Geology
Geodesy

④ F3DT

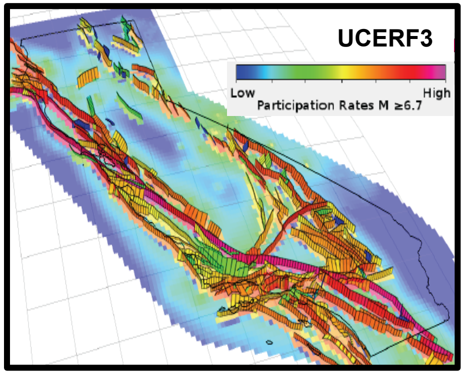
Structural Representation

③ DFR ↔ AWP → KFR → AWP → NSR ② → Ground Motions → Intensity Measures

FM – DM – ERM – PM → Empirical GMPE → Intensity Measures

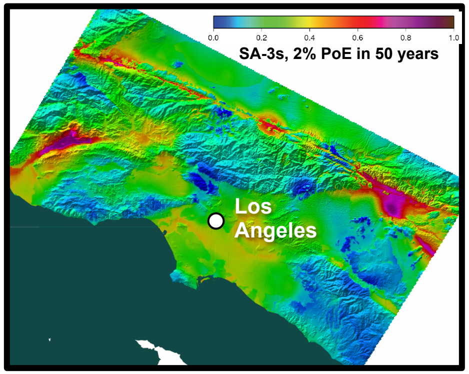
① Earthquake Rupture Forecast

TACC *Stampede*



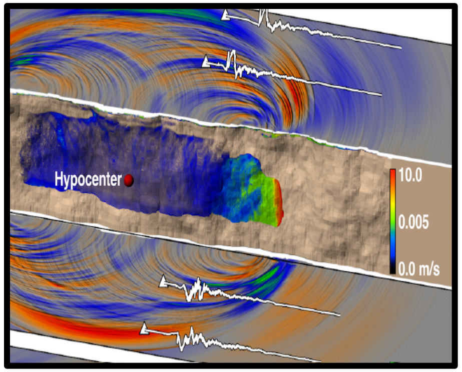
① Uniform California Earthquake Rupture Forecast (UCERF3)

NCSA *Blue Waters*



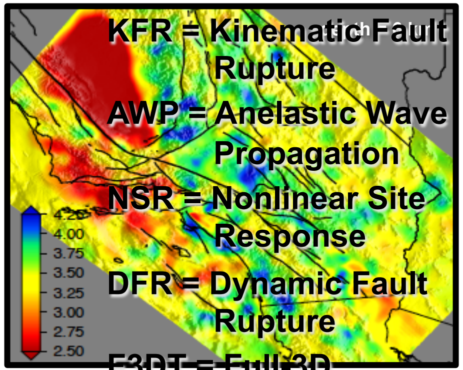
② CyberShake 14.2 seismic hazard model for LA region

OLCF *Titan*



③ Dynamic rupture model of fractal roughness on SAF

ALCF *Mira*

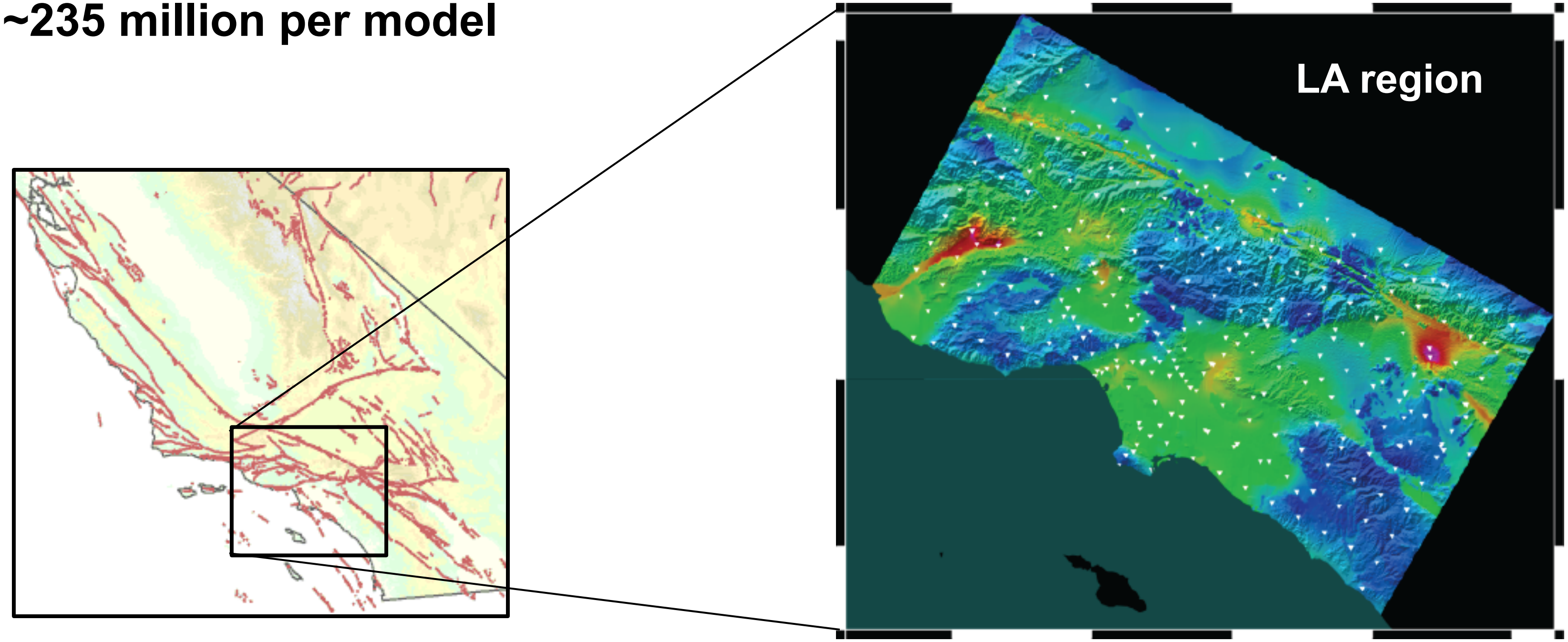


KFR = Kinematic Fault Rupture
AWP = Anelastic Wave Propagation
NSR = Nonlinear Site Response
DFR = Dynamic Fault Rupture
F3DT = Full-3D Tomography

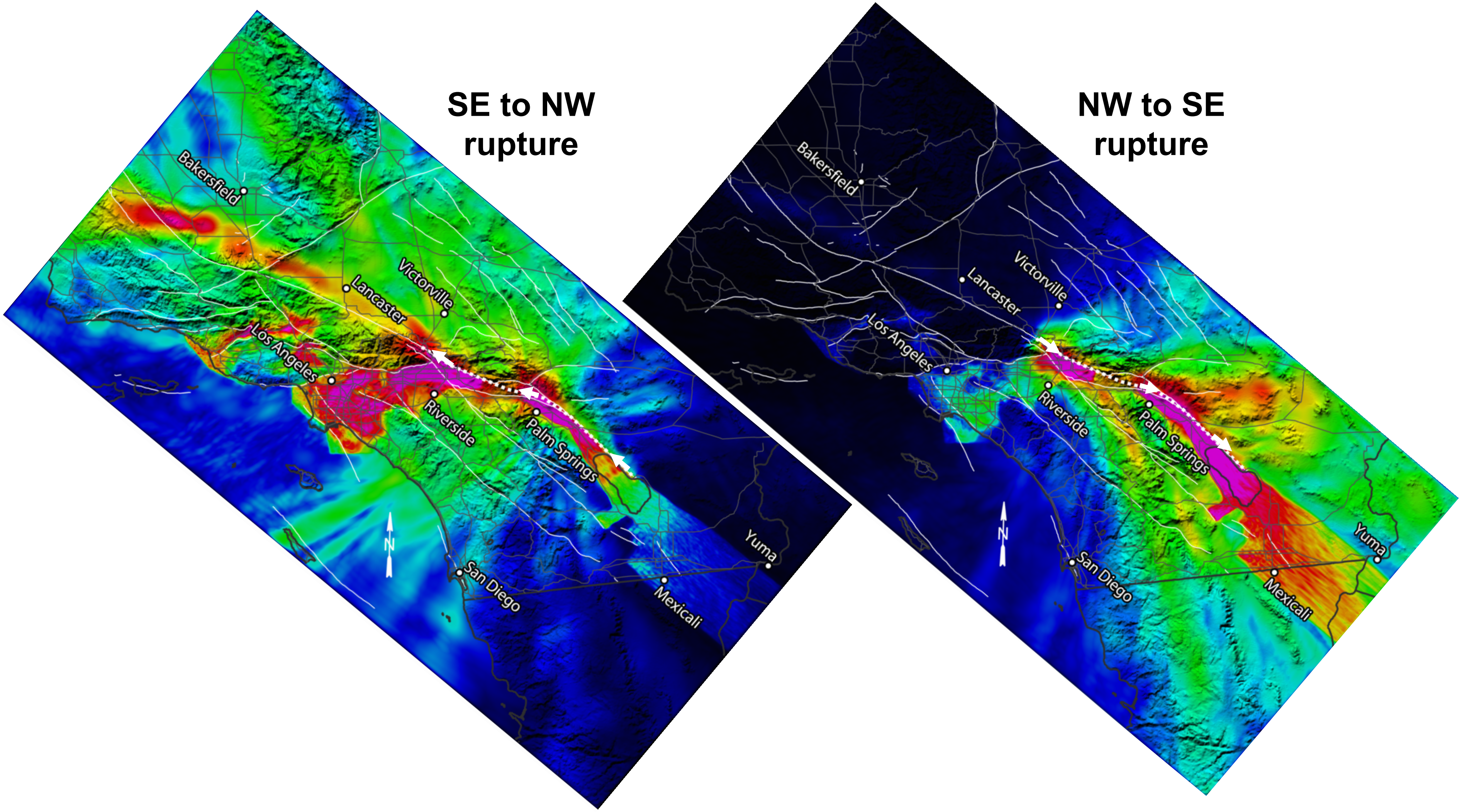
④ Full-3D tomographic model CVM-S4.26 of S. California

CyberShake Hazard Model

- **Sites:**
  - 283 sites in the greater Los Angeles region
- **Ruptures:**
  - All UCERF2 ruptures within 200 km of site (~14,900)
- **Rupture variations:**
  - ~415,000 per site using Graves-Pitarka pseudo-dynamic rupture model
- **Seismograms:**
  - ~235 million per model



Coupling of Directivity and Basin Effects

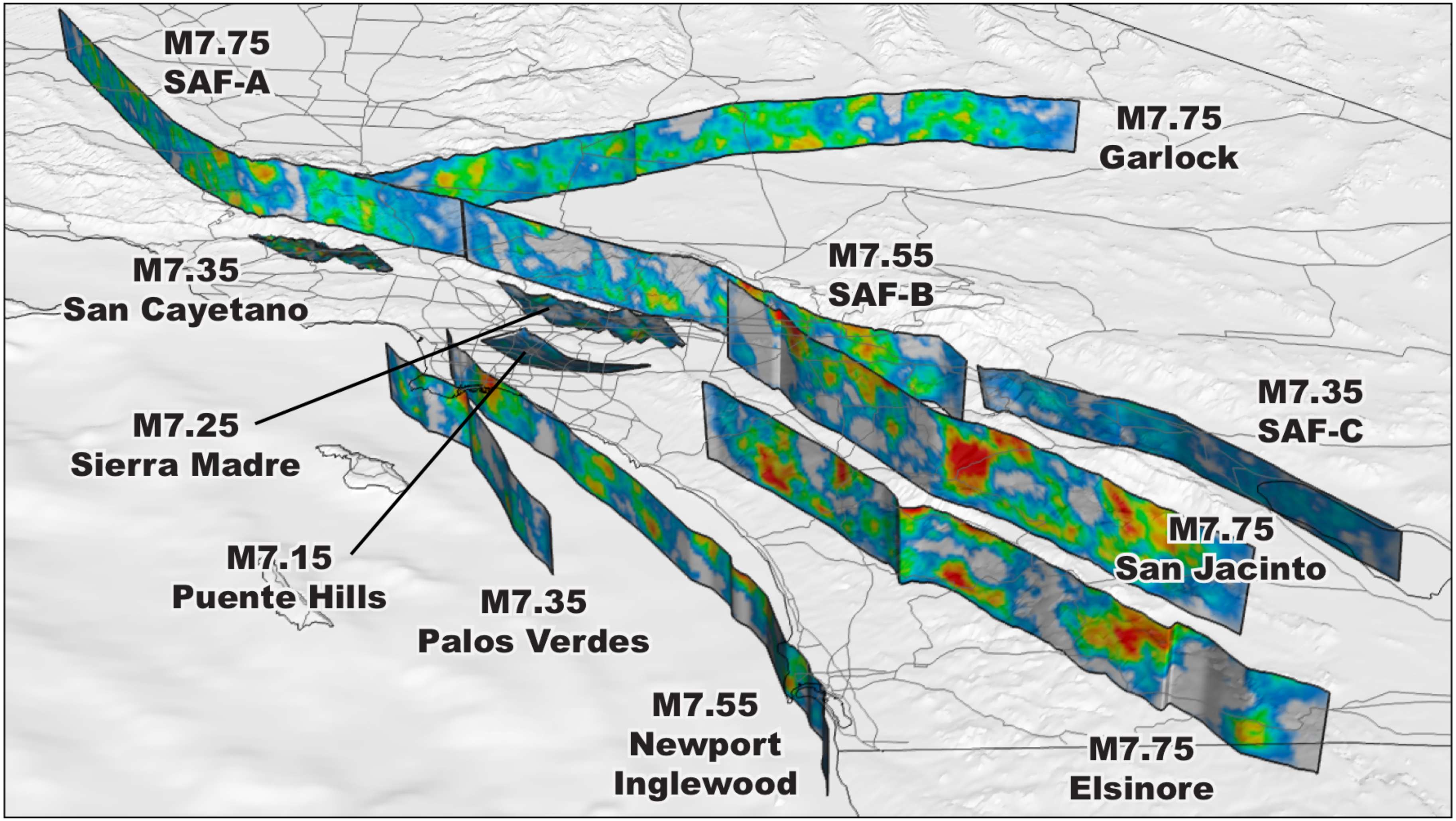


SE to NW rupture

NW to SE rupture

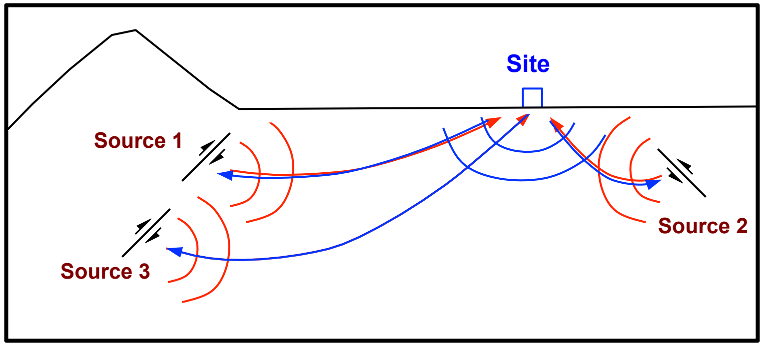
TeraShake simulations of M7.7 earthquake on Southernmost San Andreas

Examples of CyberShake Rupture Models



Rapid Simulation of Large Rupture Ensembles Using Seismic Reciprocity

- **To account for source variability requires very large sets of simulations**
  - **14,900 ruptures from UCERF2; 415,000 rupture variations**
- **Ground motions need only be calculated at much smaller number of surface sites to produce hazard map**
  - **283 in LA region, interpolated using empirical attenuation relations**
- ***Use of reciprocity reduces CPU time by a factor of ~1,000***

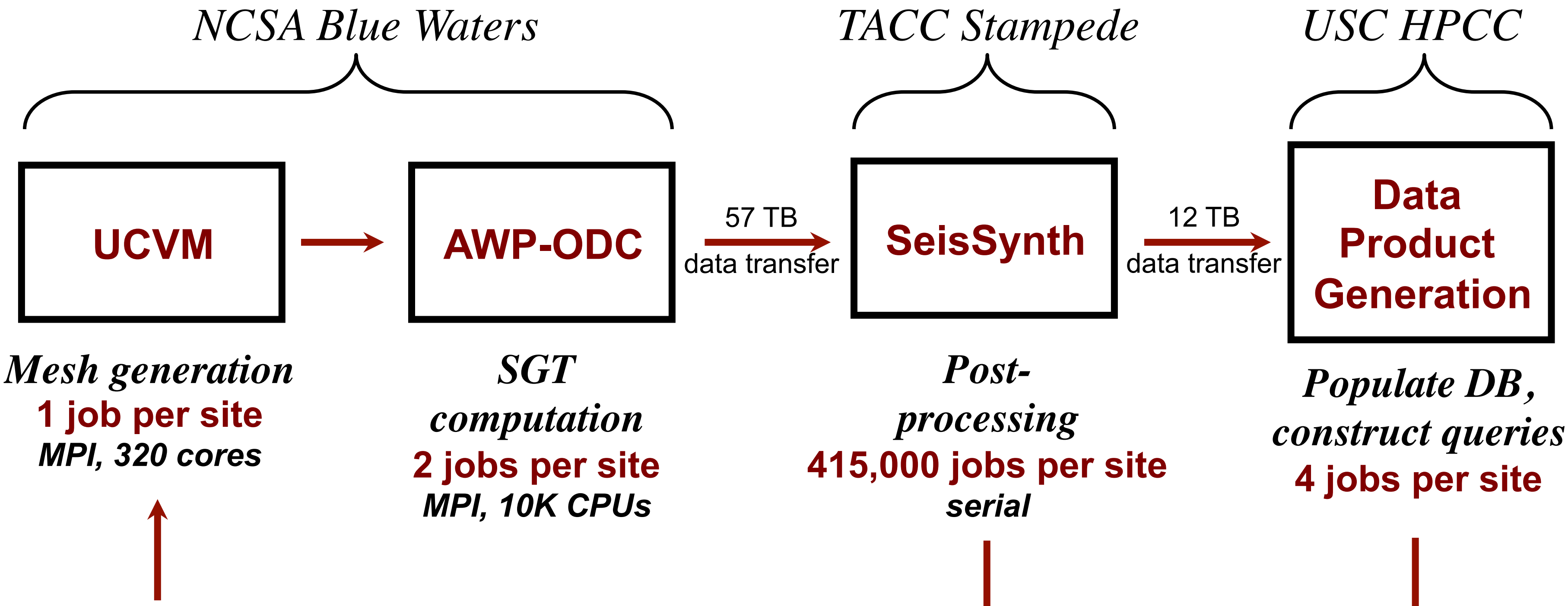


Strain Green Tensor (SGT)

M **sources to** ***N*** **sites requires** ***M*** **simulations**

M **sources to** ***N*** **sites requires 2*N* or 3*N* simulations**

CyberShake CS13.4 Workflow



NCSA Blue Waters

TACC Stampede

USC HPCC

UCVM

AWP-ODC

57 TB
data transfer

SeisSynth

12 TB
data transfer

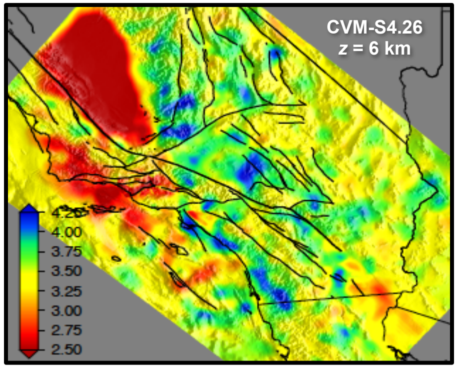
Data Product Generation

Mesh generation
1 job per site
MPI, 320 cores

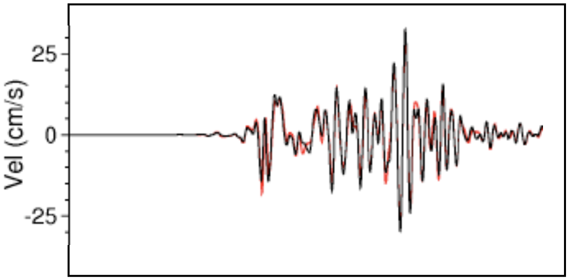
SGT computation
2 jobs per site
MPI, 10K CPUs

Post-processing
415,000 jobs per site
serial

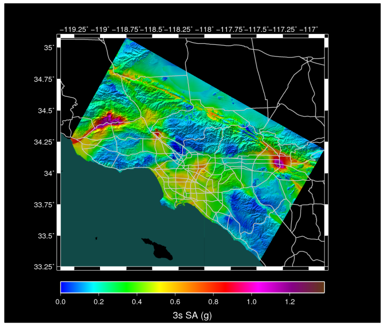
Populate DB, construct queries
4 jobs per site



Community Velocity Model



Seismograms



Hazard Products

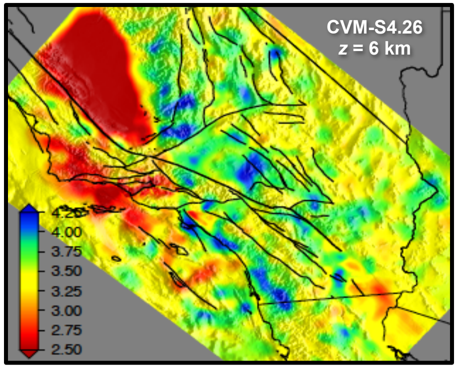
CyberShake CS14.2 Workflow

NCSA Blue Waters

USC HPCC

UCVM

Mesh generation
1 job per site
MPI, 320 cores



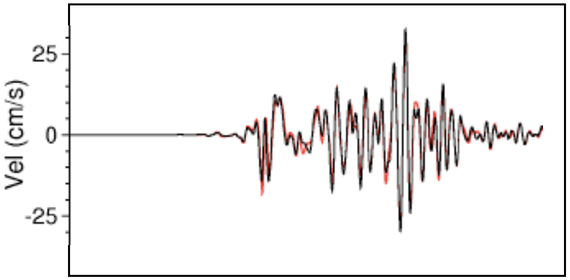
Community Velocity Model

AWP-ODC

SGT computation
2 jobs per site
MPI, 10K CPUs & 100 GPUs

SeisSynth

Post-processing
415,000 jobs per site
serial

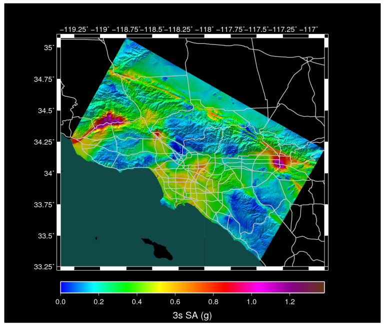


Seismograms

12 TB
data transfer

Data Product Generation

Populate DB, construct queries
4 jobs per site



Hazard Products

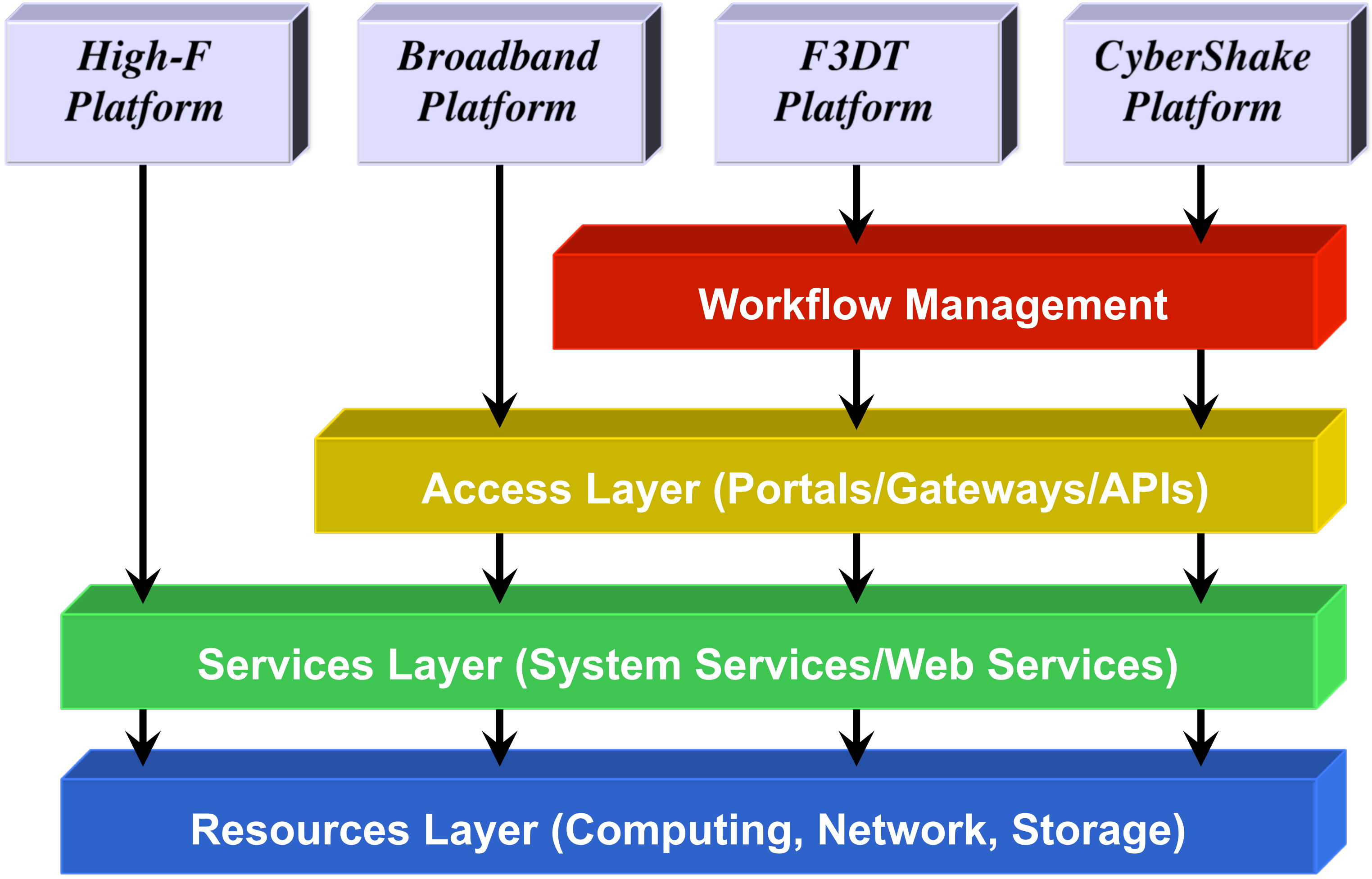
CyberShake Time-to-Solution Comparison

Los Angeles Region Hazard Models (1144 sites)

CyberShake Application Metrics (Hours)	**2008 (Mercury, normalized)**	**2009 (Ranger, normalized)**	**2013 (Blue Waters / Stampede)**	**2014 (Blue Waters)**
Application Core Hours:	19,488,000 (CPU)	16,130,400 (CPU)	12,200,000 (CPU)	15,800,000 (CPU +GPU)
Application Makespan:	70,165	6,191	1,467	342

Metric	**2013 (Study 13.4)**	**2014 (Study 14.2)**
Simultaneous processors	21,100 (CPU)	46,720 (CPU) + 160 (GPU)
Concurrent Workflows	5.8	26.2
Job Failure Rate	2.6%	1.3%
Data transferred	57 TB	12 TB

Vertical Integration of CI Layers



High-F Platform

Broadband Platform

F3DT Platform

CyberShake Platform

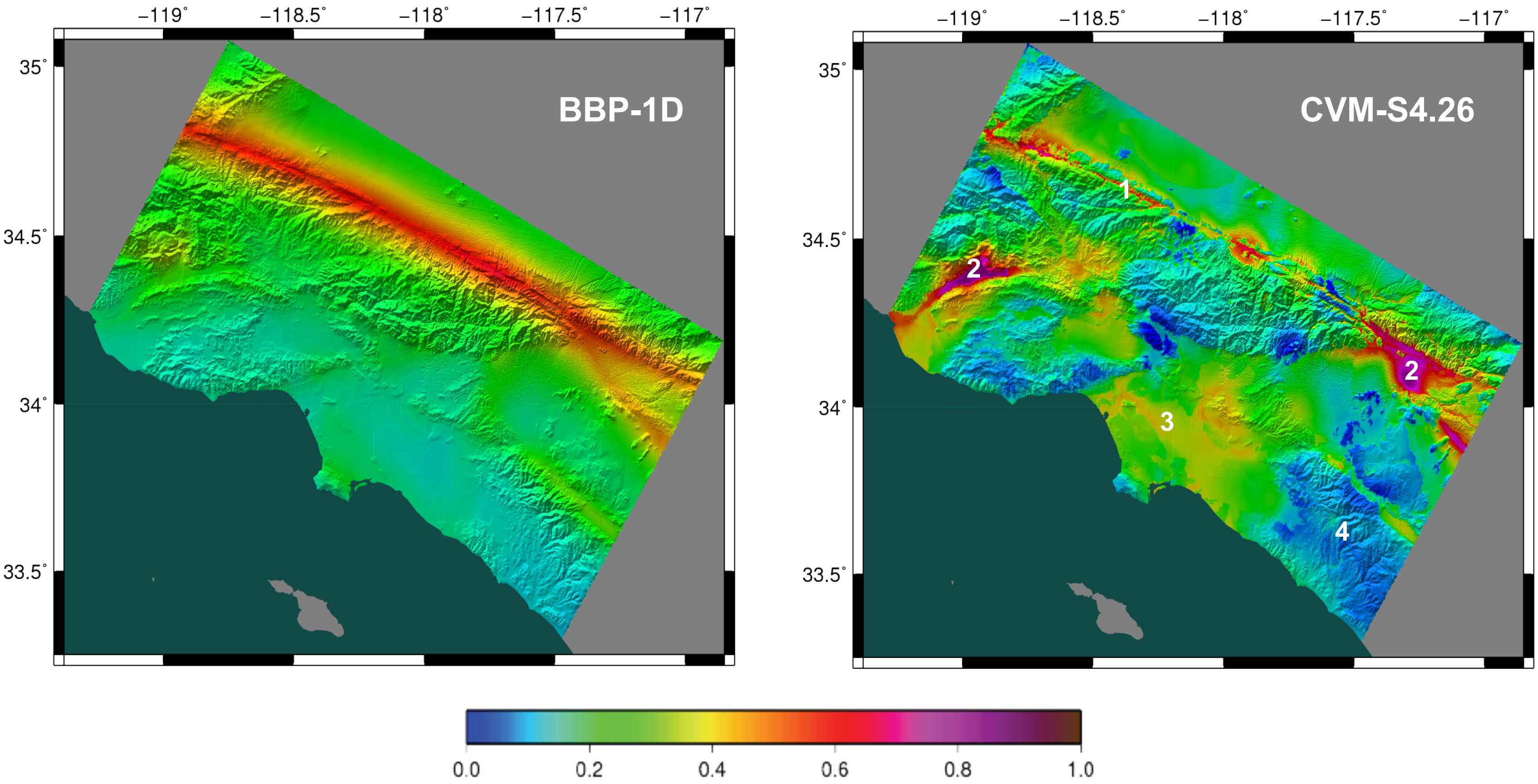
Workflow Management

Access Layer (Portals/Gateways/APIs)

Services Layer (System Services/Web Services)

Resources Layer (Computing, Network, Storage)

Comparison of 1D and 3D CyberShake Models for the Los Angeles Region



CyberShake Hazard Map, 3sec SA, 2% in 50 yrs

1. lower near-fault intensities due to 3D scattering
2. much higher intensities in near-fault basins
3. higher intensities in the Los Angeles basins
4. lower intensities in hard-rock areas

Averaging-Based Factorization

- **Representation of excitation functionals**

  Expected shaking intensities constructed by averaging over slip variations (s), hypocenters (x), sources (k), and sites (r)

$$\underset{\substack{\uparrow \\ \text{ln }(Y)}}{G(r,k,x,s)} \;=\; \underset{\substack{\uparrow \\ \text{level}}}{A} + \underset{\substack{\uparrow \\ \text{site} \\ \text{effect}}}{B(r)} + \underset{\substack{\uparrow \\ \text{path} \\ \text{effect}}}{C(r,k)} + \underset{\substack{\uparrow \\ \text{directivity} \\ \text{effect}}}{D(r,k,x)} + \underset{\substack{\uparrow \\ \text{slip complexity} \\ \text{effect}}}{E(r,k,x,s)}$$

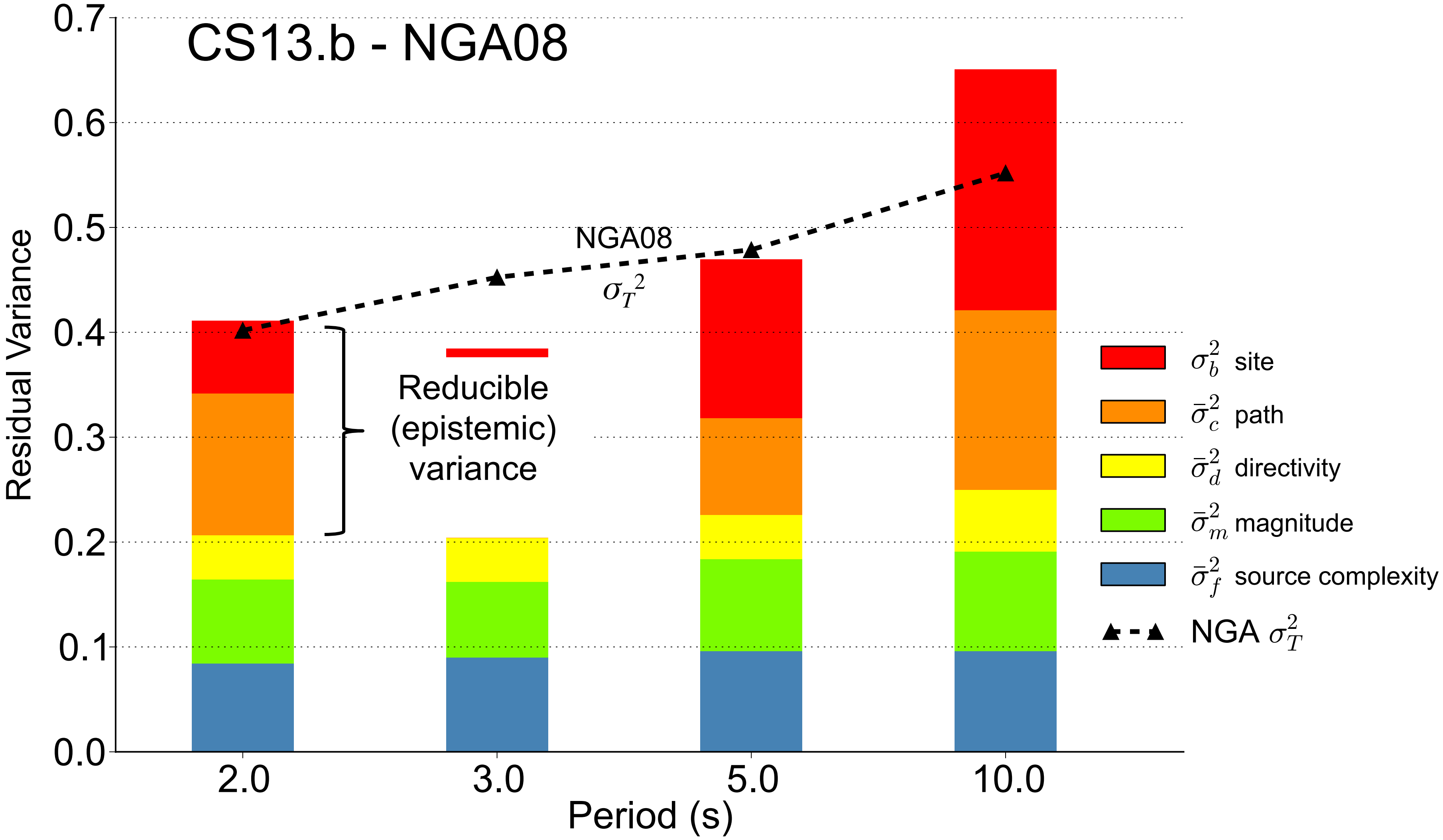
- **Representation of excitation variance**

$$\mathrm{Var}[G] \;=\; \bar{\sigma}_G^2 \equiv \left\langle [G(r,k,x,s) - A]^2 \right\rangle_{S,X,K,R}$$

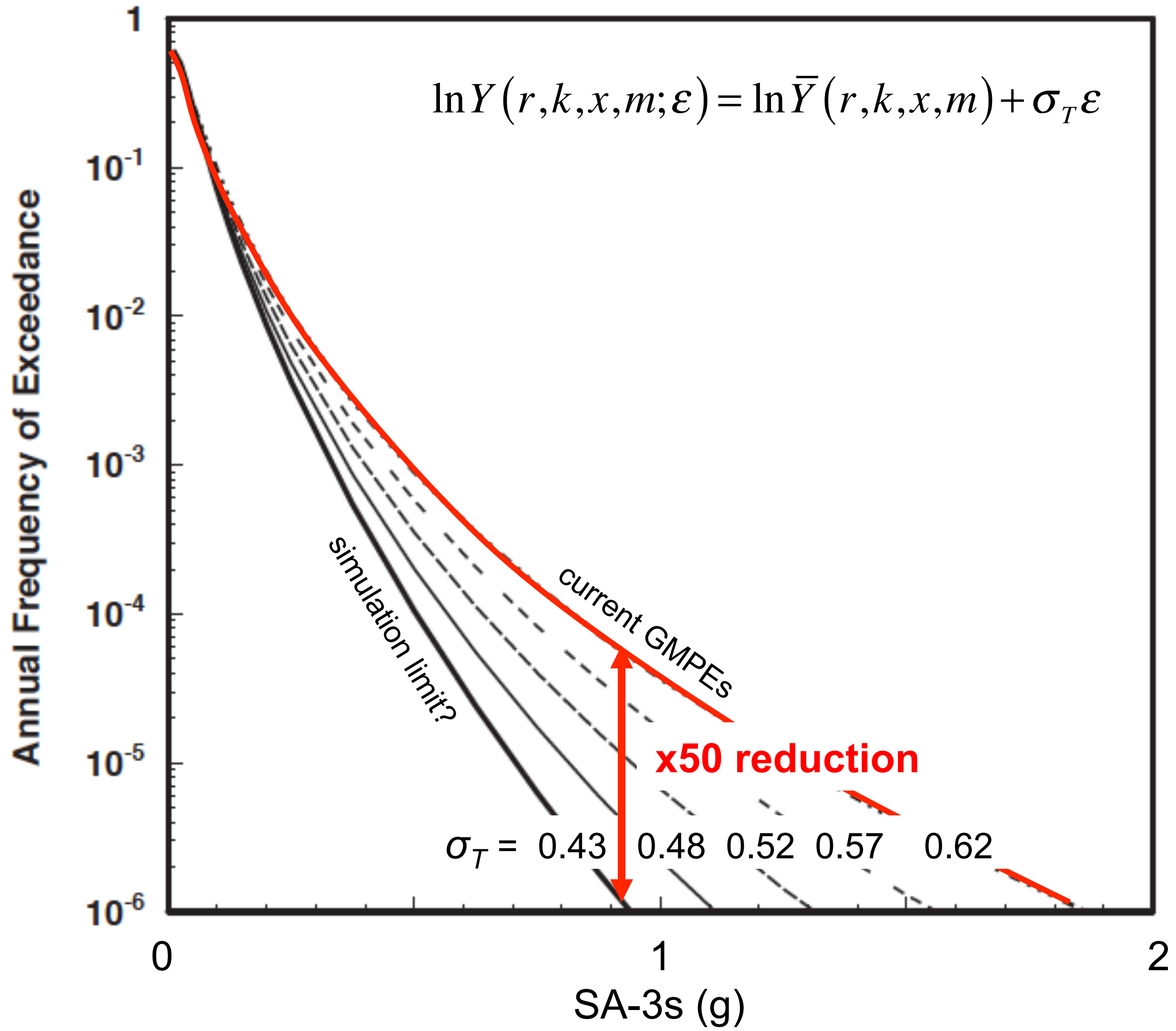
$$= \;\sigma_B^2 + \left\langle \sigma_C^2(r) \right\rangle_R + \left\langle \sigma_D^2(r,k) \right\rangle_{K,R} + \left\langle \sigma_E^2(r,k,x) \right\rangle_{X,K,R}$$

$$\equiv \;\sigma_B^2 + \quad \bar{\sigma}_C^2 \quad + \quad \bar{\sigma}_D^2 \quad + \quad \bar{\sigma}_E^2$$

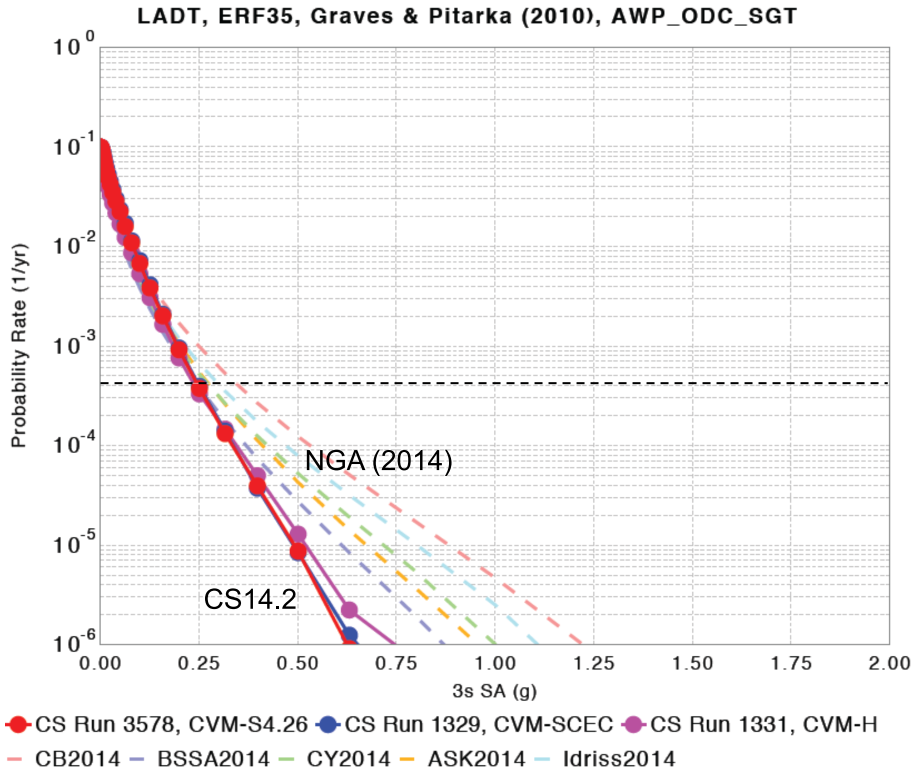
ABF Variance Analysis



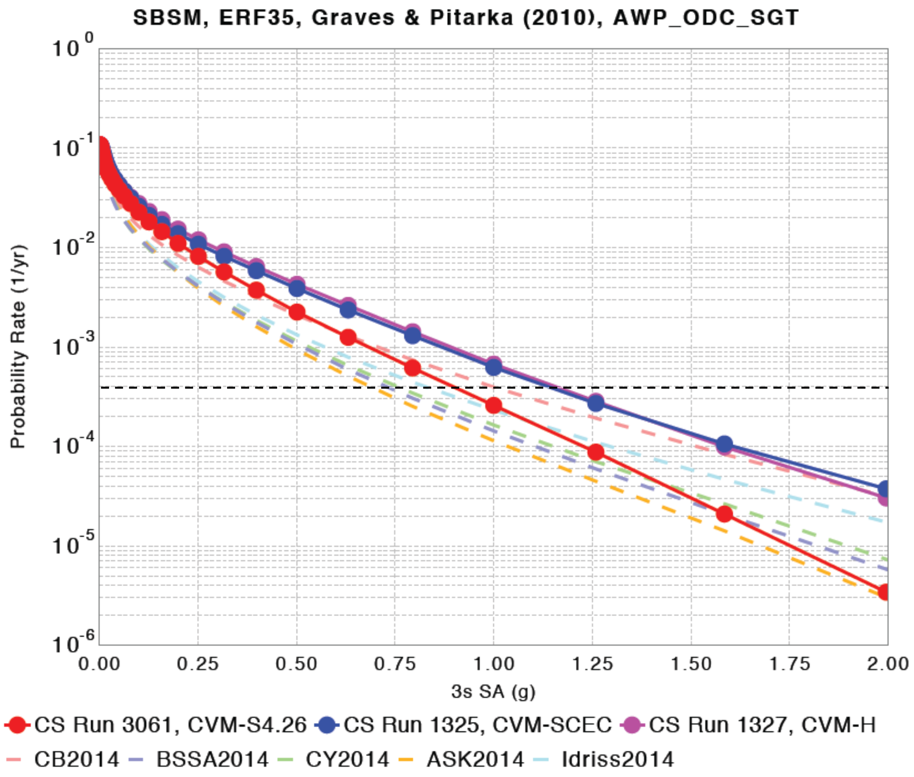
Importance of Reducing Aleatory Variability



NGA(2014)-CyberShake Hazard Curve Comparisons



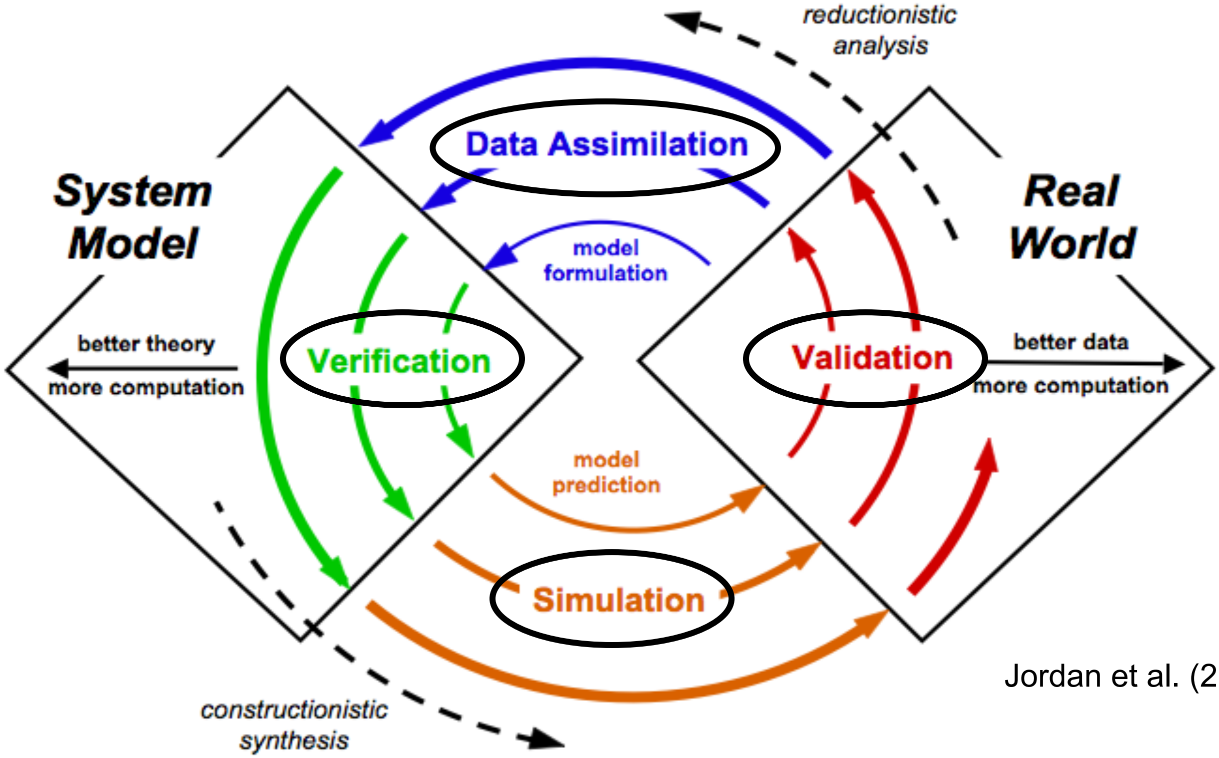
Site LADT
(Los Angeles)



Site SBSM
(San Bernardino)

Inference Spiral of System Science

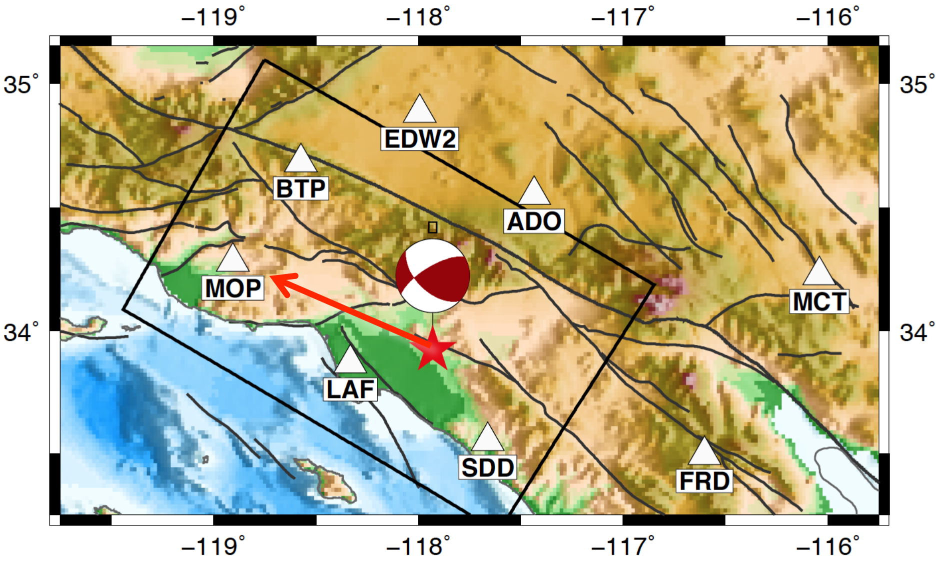
- Earthquake system science requires an iterative, computationally intense process of model formulation and verification, simulation-based predictions, validation against observations, and data assimilation to improve the model



Jordan et al. (2010)

- As models become more complex and new data bring in more information, we require ever increasing computational resources

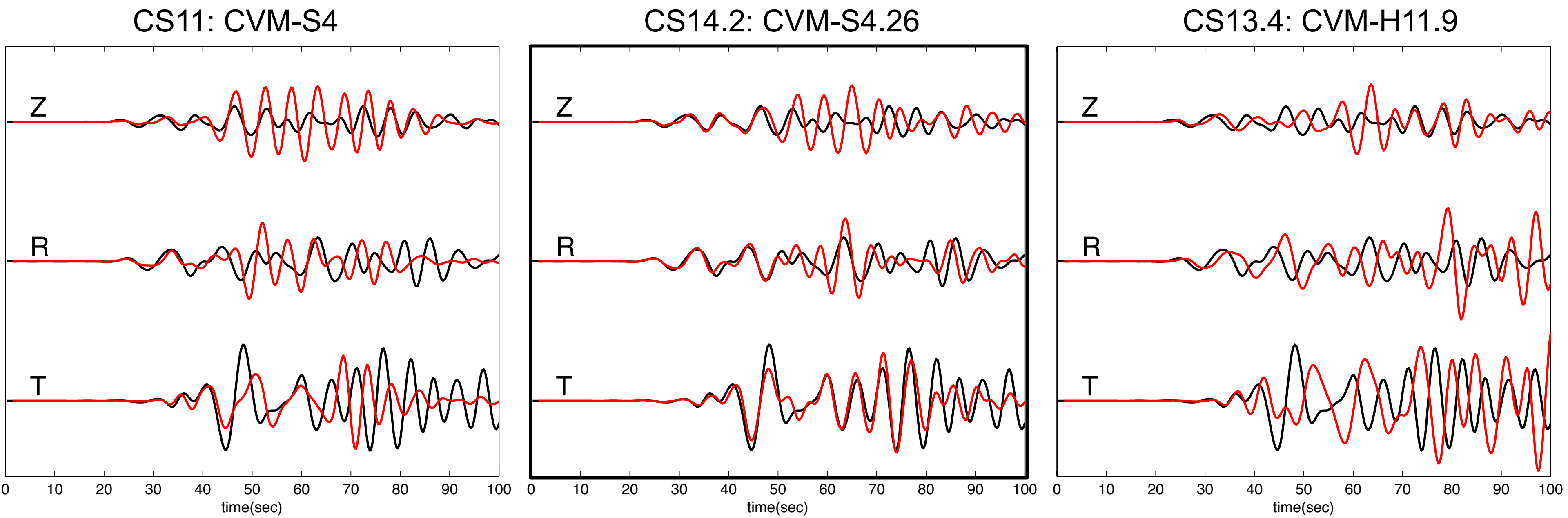
03/28/14 La Habra Earthquake (M5.1)



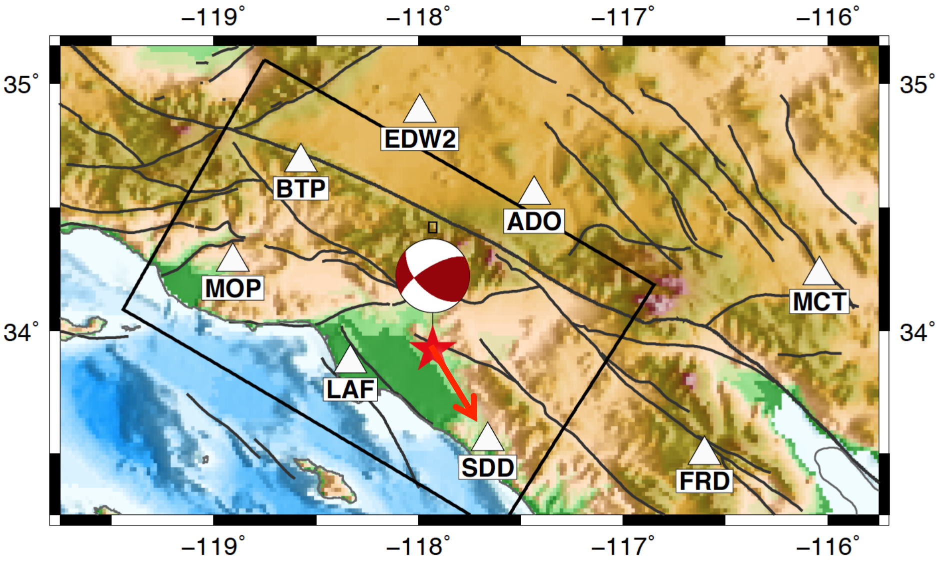
Station MOP

Observed in black

Synthetic in red

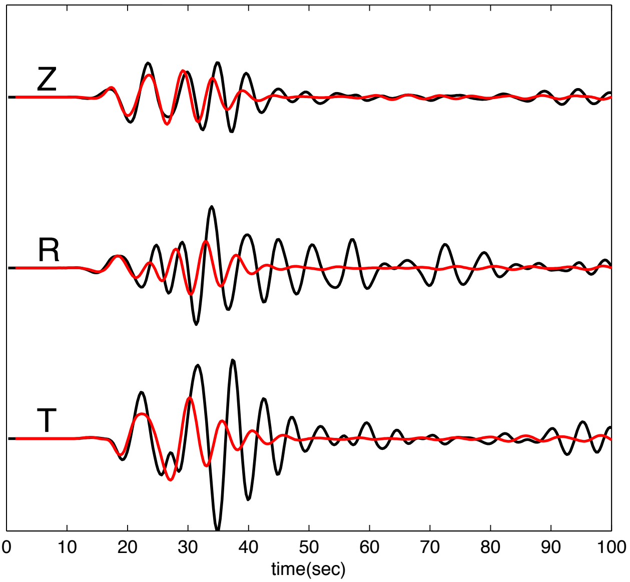


03/28/14 La Habra Earthquake (M5.1)

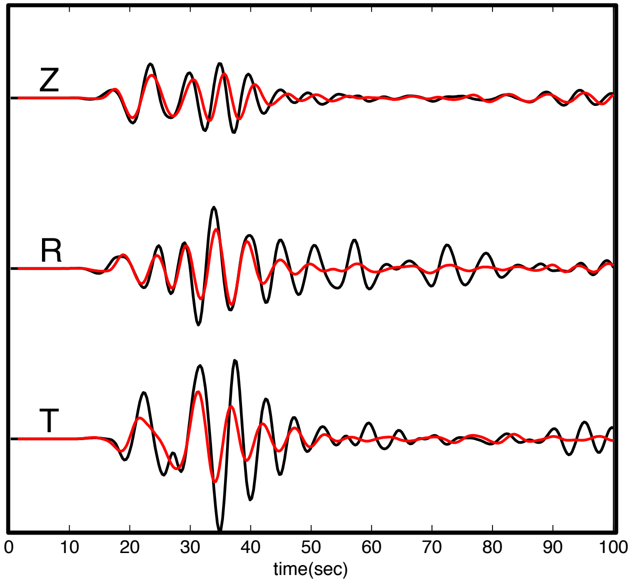


Station SDD
Observed in black
Synthetic in red

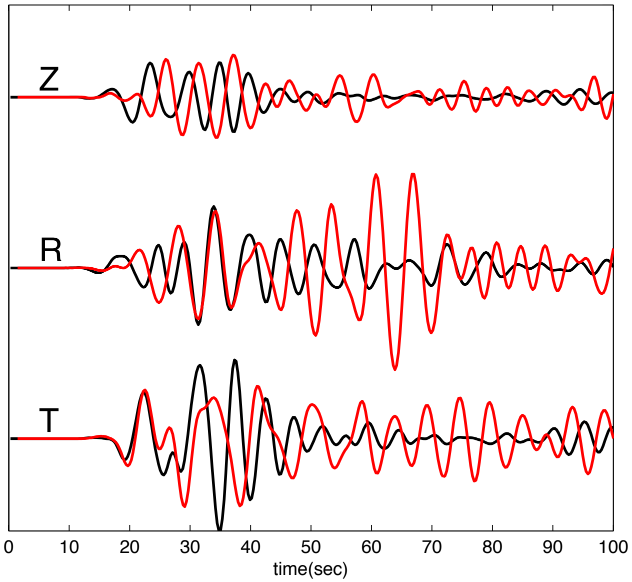
CS11: CVM-S4



CS14.2: CVM-S4.26

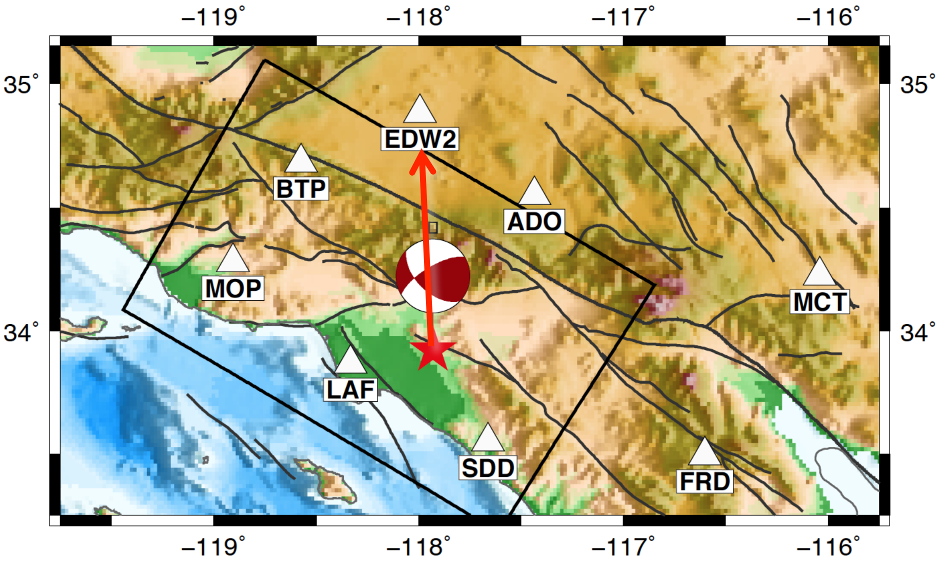


CS13.4: CVM-H11.9

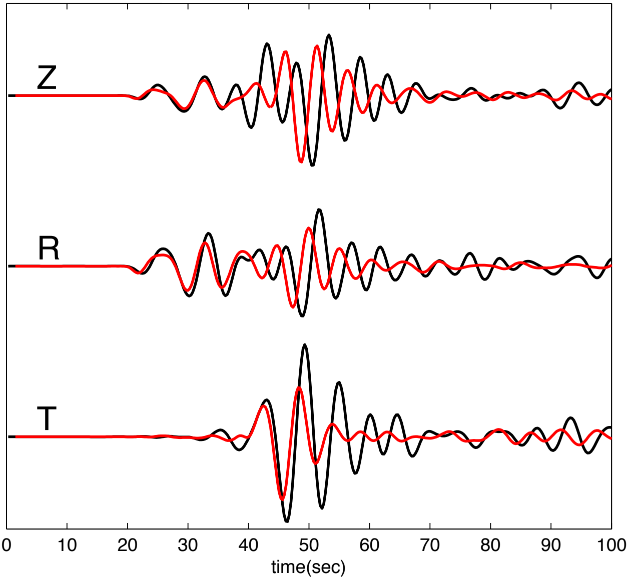


03/28/14 La Habra Earthquake (M5.1)

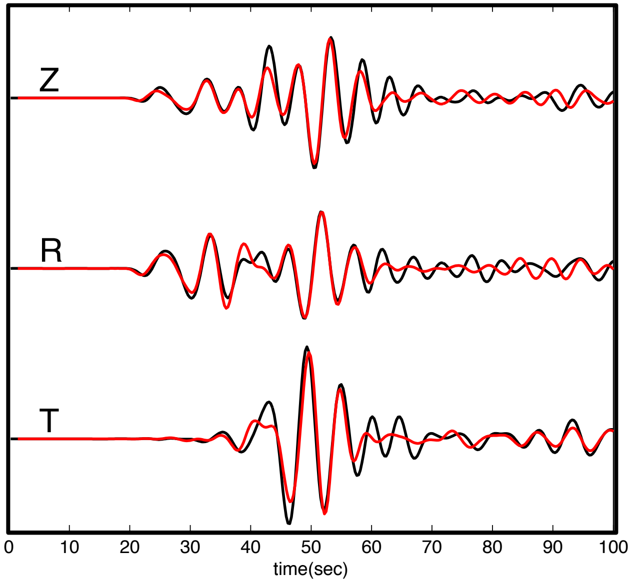
Station EDW2
Observed in black
Synthetic in red



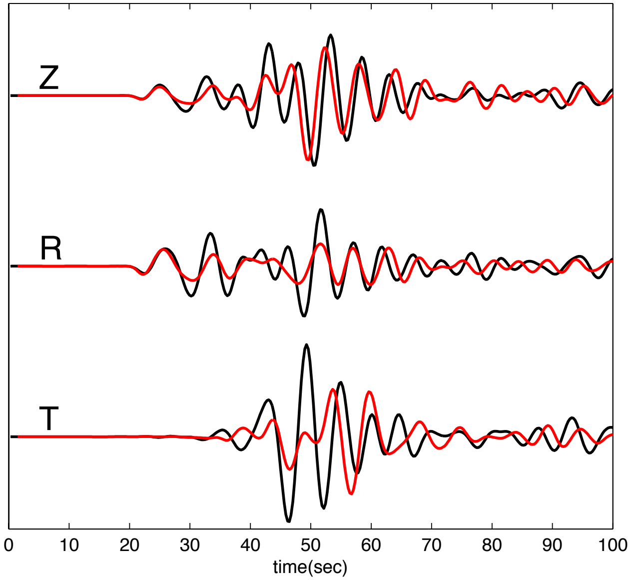
CS11: CVM-S4



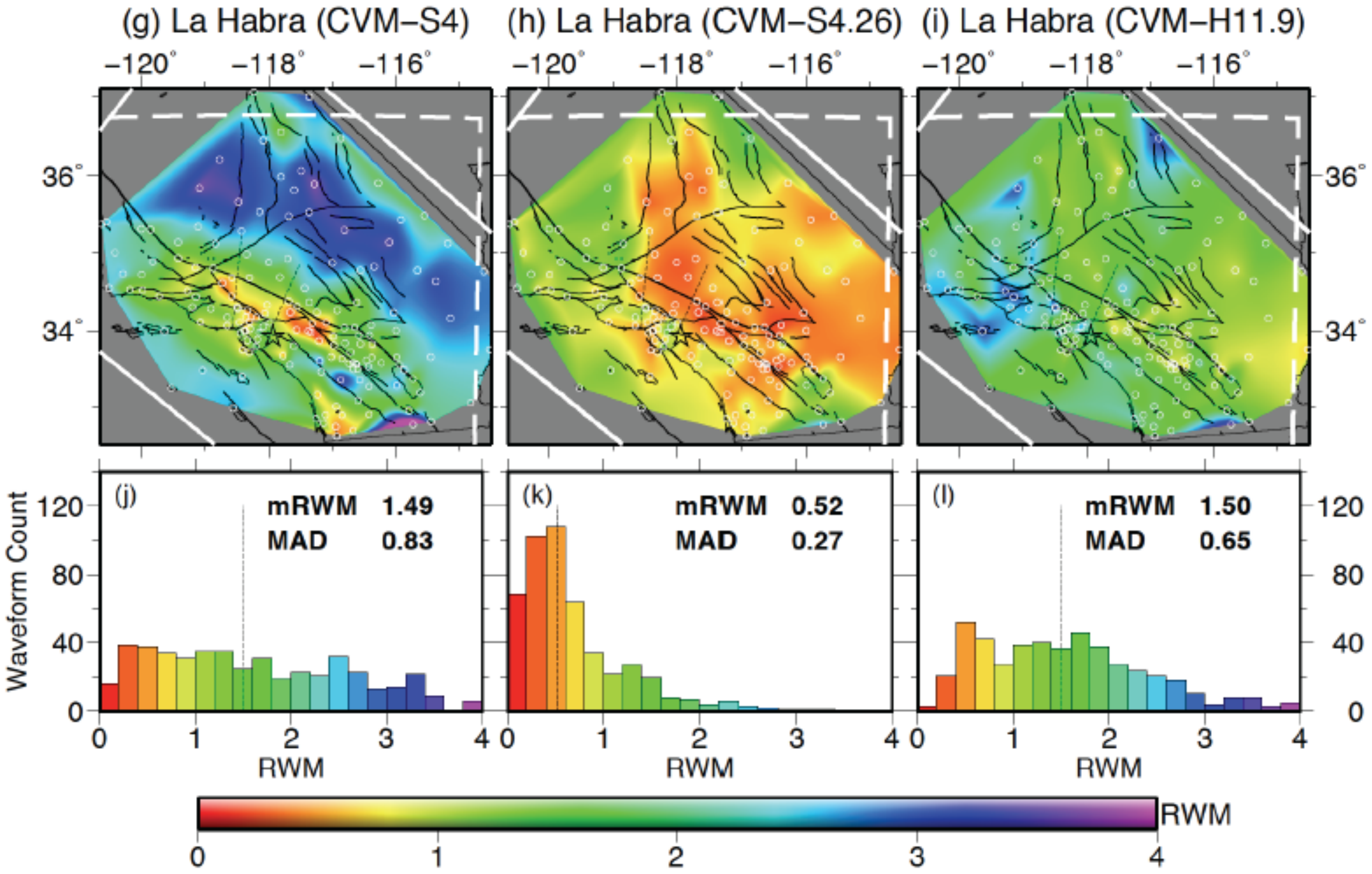
CS14.2: CVM-S4.26



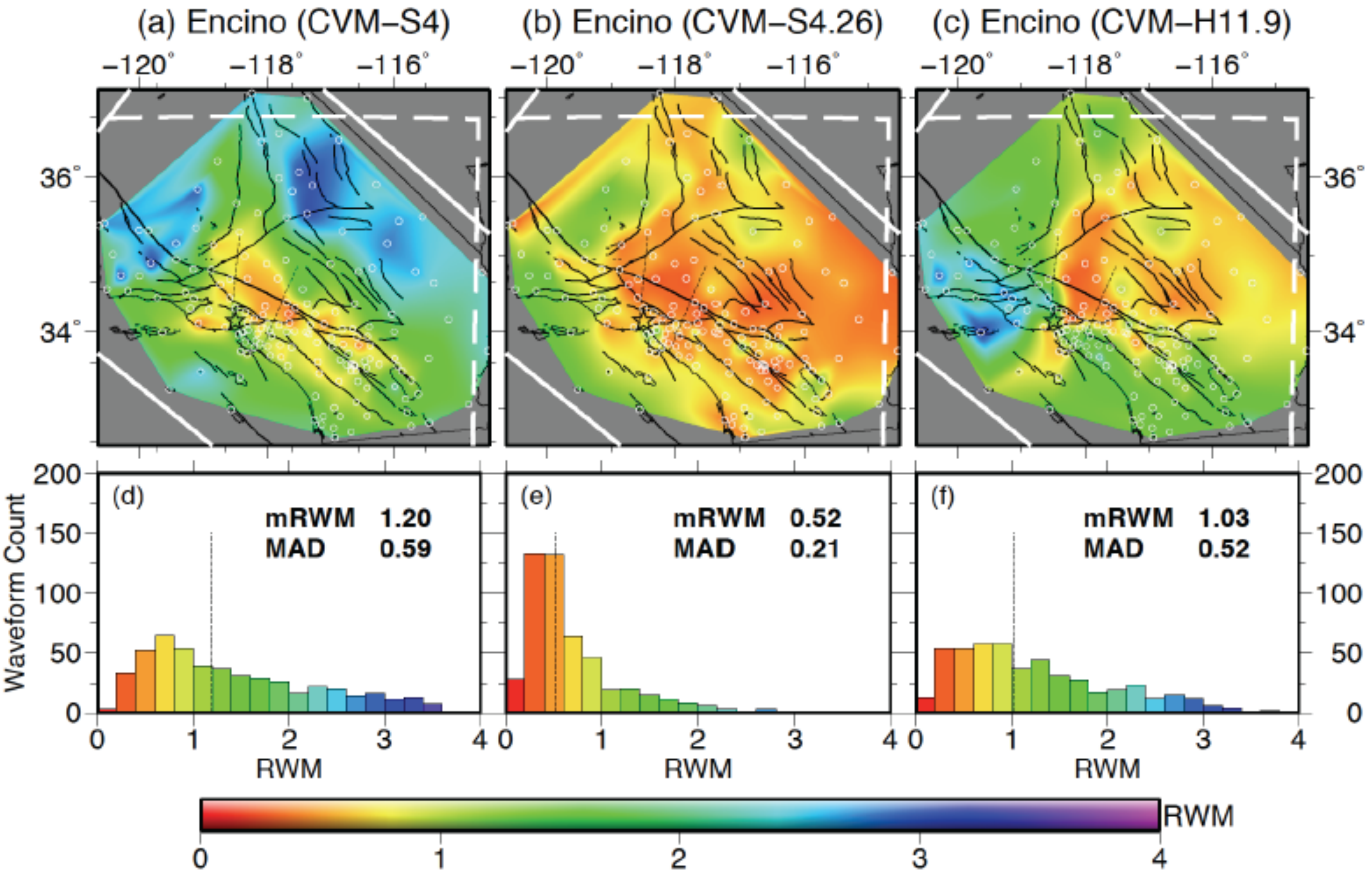
CS13.4: CVM-H11.9



03/28/14 La Habra Earthquake (M5.1)

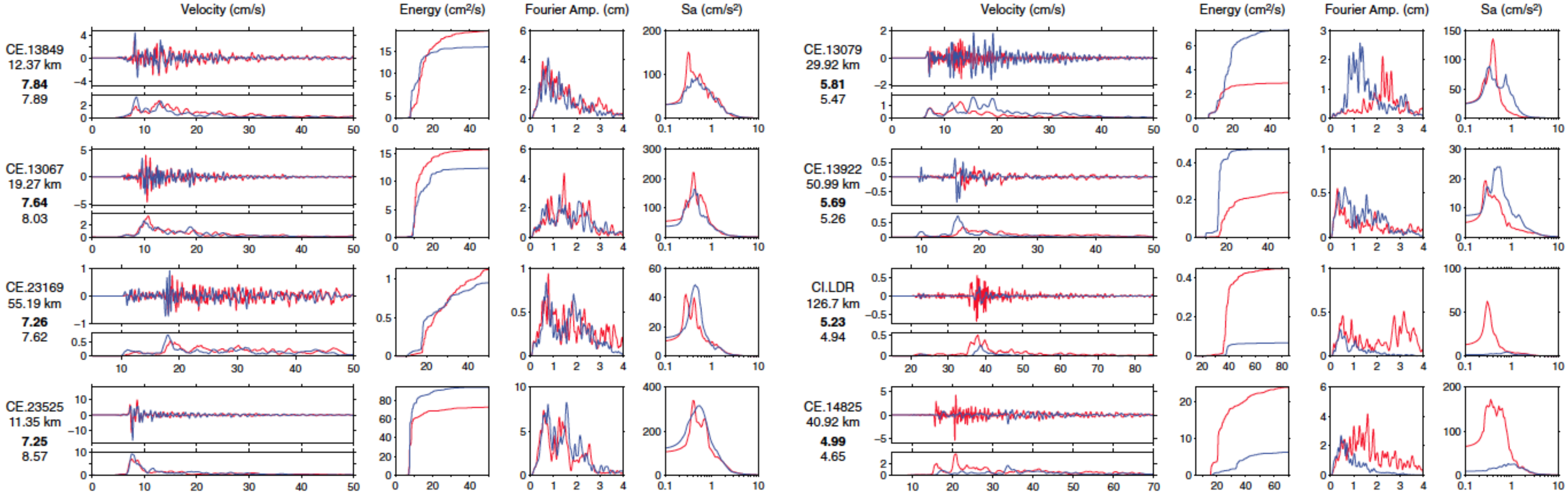


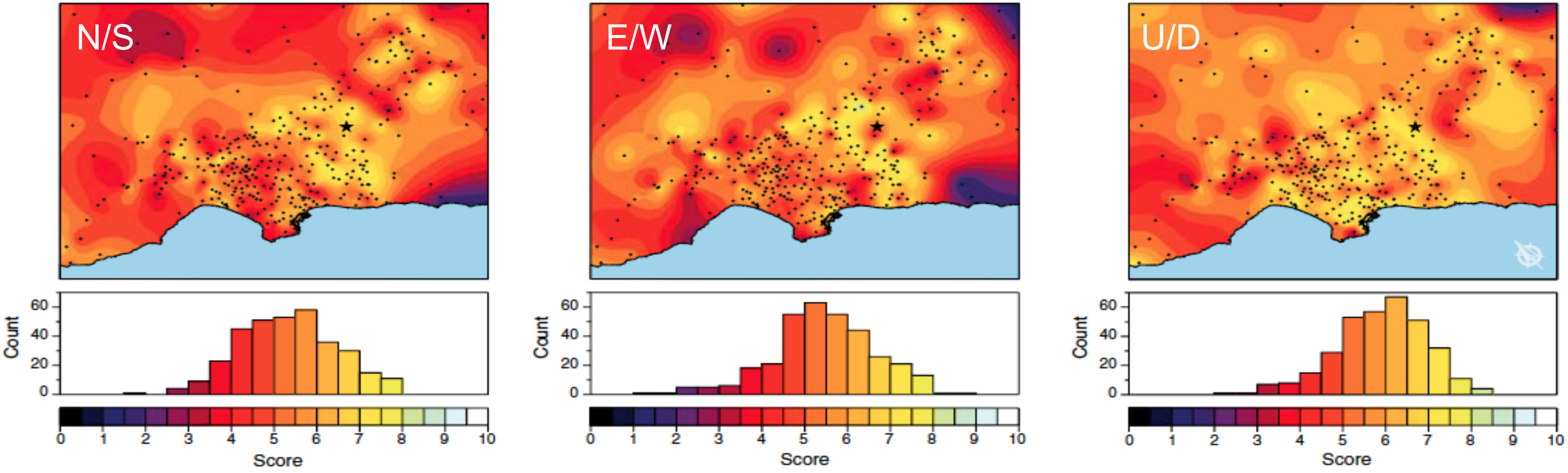
03/17/14 Encino Earthquake (M4.4)



07/28/08 Chino Hills Earthquake (M5.4)

(Taborda & Bielak, 2013)

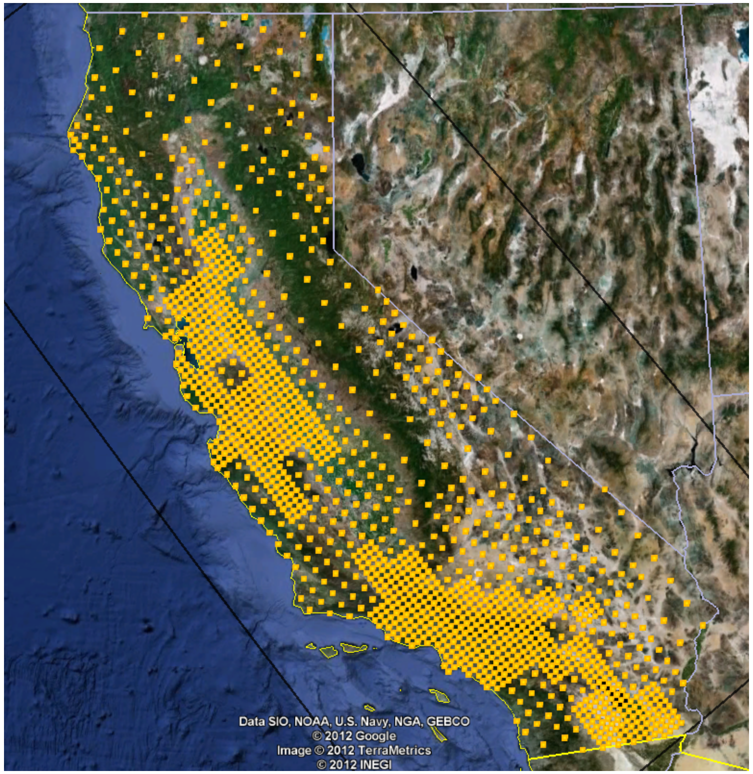




CyberShake Science Challenges

- **We plan to move towards**
  - **higher frequencies (0.5 Hz → 2 Hz)**
  - **more ruptures (UCERF3)**
  - **more sites (1440 for statewide)**
- **This will require better physics…**
  - **Frequency-dependent attenuation**
  - **Fault roughness**
  - **Near-fault plasticity**
  - **Soil non-linearities**
  - **Near-surface heterogeneities**

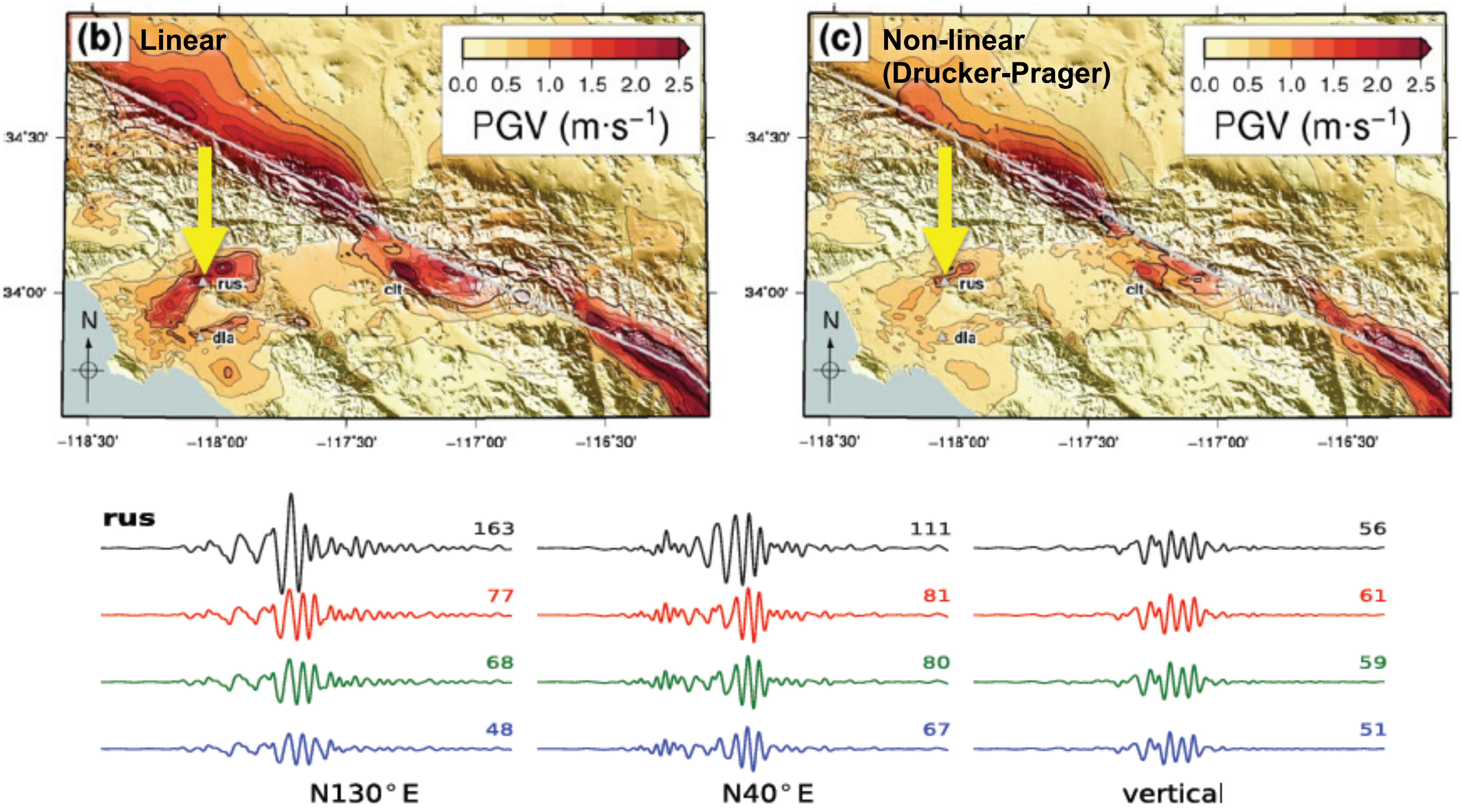
… and much more computation!


Data SIO, NOAA, U.S. Navy, NGA, GEBCO
© 2012 Google
Image © 2012 TerraMetrics
© 2012 INEGI

Statewide CyberShake

- Computational requirements for 1 Hz:
  - Number of jobs: 23.2 billion
  - Storage: 580 TB
  - CPU hrs: 253 million

Nonlinear Simulations of the ShakeOut Scenario

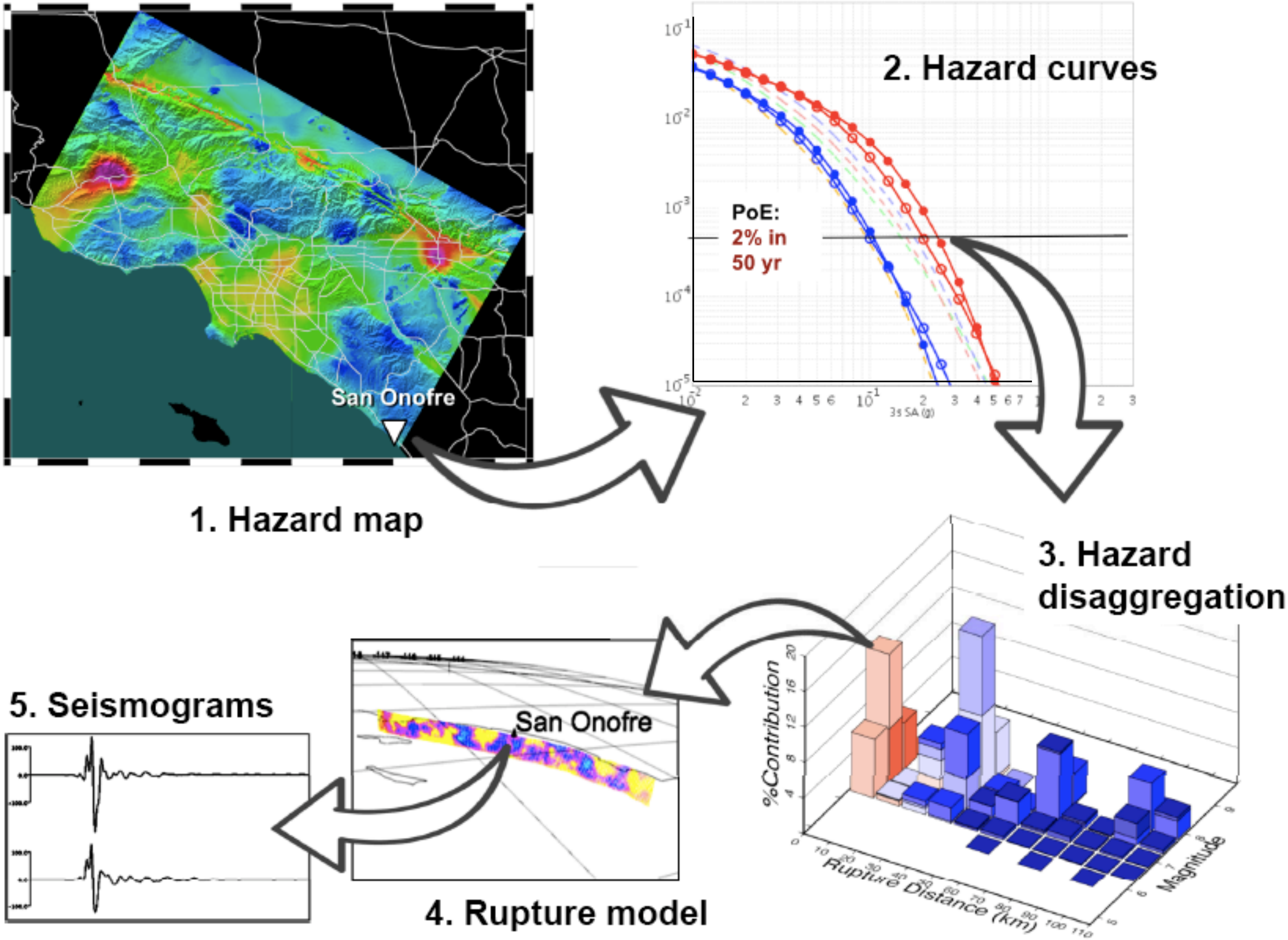


Roten et al. (2014)

Conclusions

- **Much of the aleatory variability in the conditional forecasting earthquake ground motions is due to 3D variations in crustal structure**
  - **Observed variability can be modeled by simulating seismic wave propagation through realistic 3D structures**
- **Large ensembles of simulations are needed for physics-based PSHA**
  - **Now feasible using seismic reciprocity, highly optimized anelastic wave propagation codes, and automated workflow management systems**
- **Low-frequency (< 0.5 Hz) CyberShake hazard models have been computed for the Los Angeles region on *Blue Waters***
  - **Show the importance of basin amplification and directivity-basin coupling**
  - **Predict well the low-frequency seismograms recorded from recent earthquakes**
- **More accurate earthquake simulations have the potential for reducing the residual variance of the ground motion predictions by ~2x**
  - **Will lower exceedance probabilities by >10x at high hazard levels**
  - **Practical ramifications for risk-reduction strategies are substantial**

CyberShake Platform: Physics-Based PSHA



1. Hazard map

2. Hazard curves

PoE:
2% in
50 yr

3. Hazard disaggregation

4. Rupture model

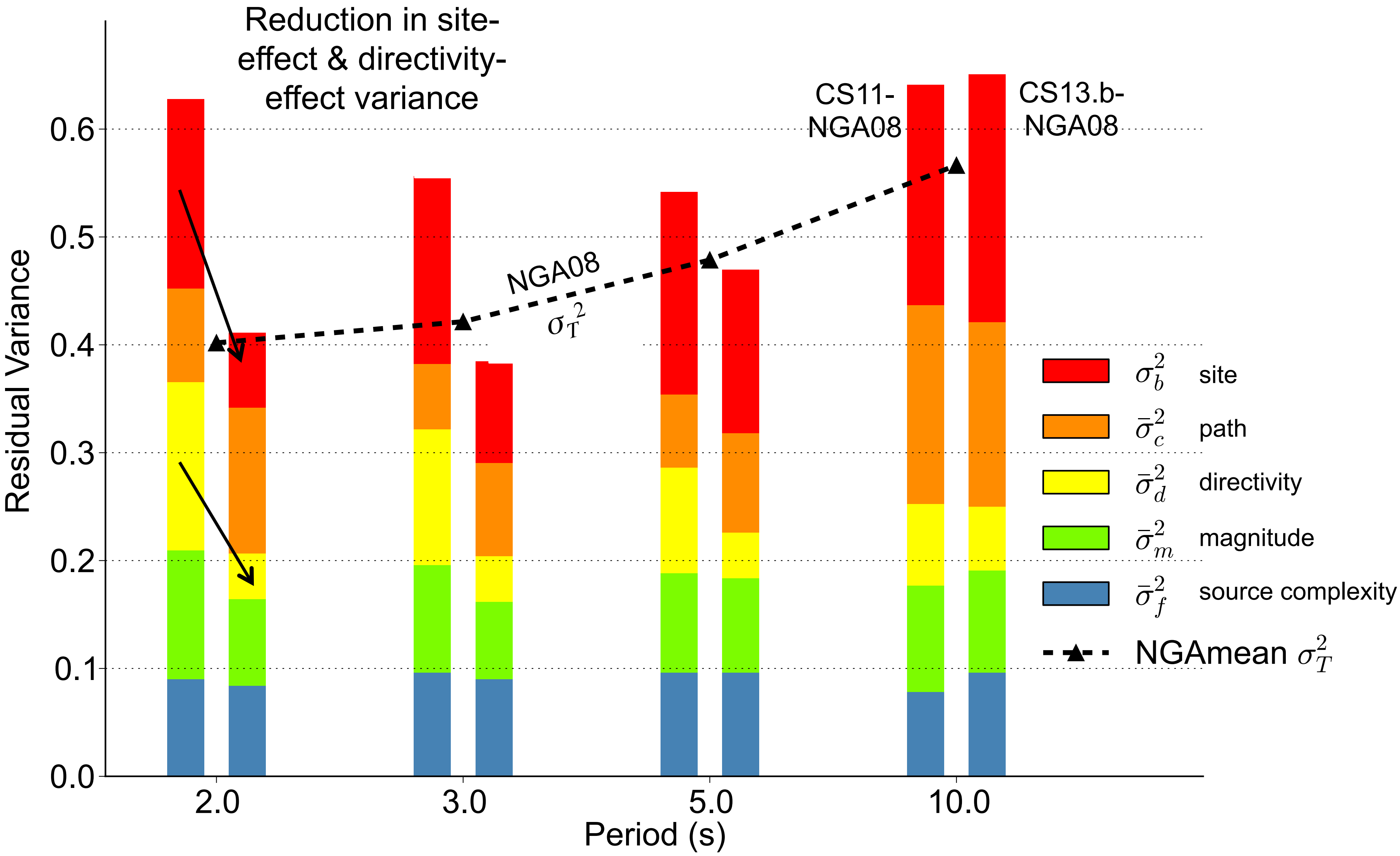
5. Seismograms

Thank you!

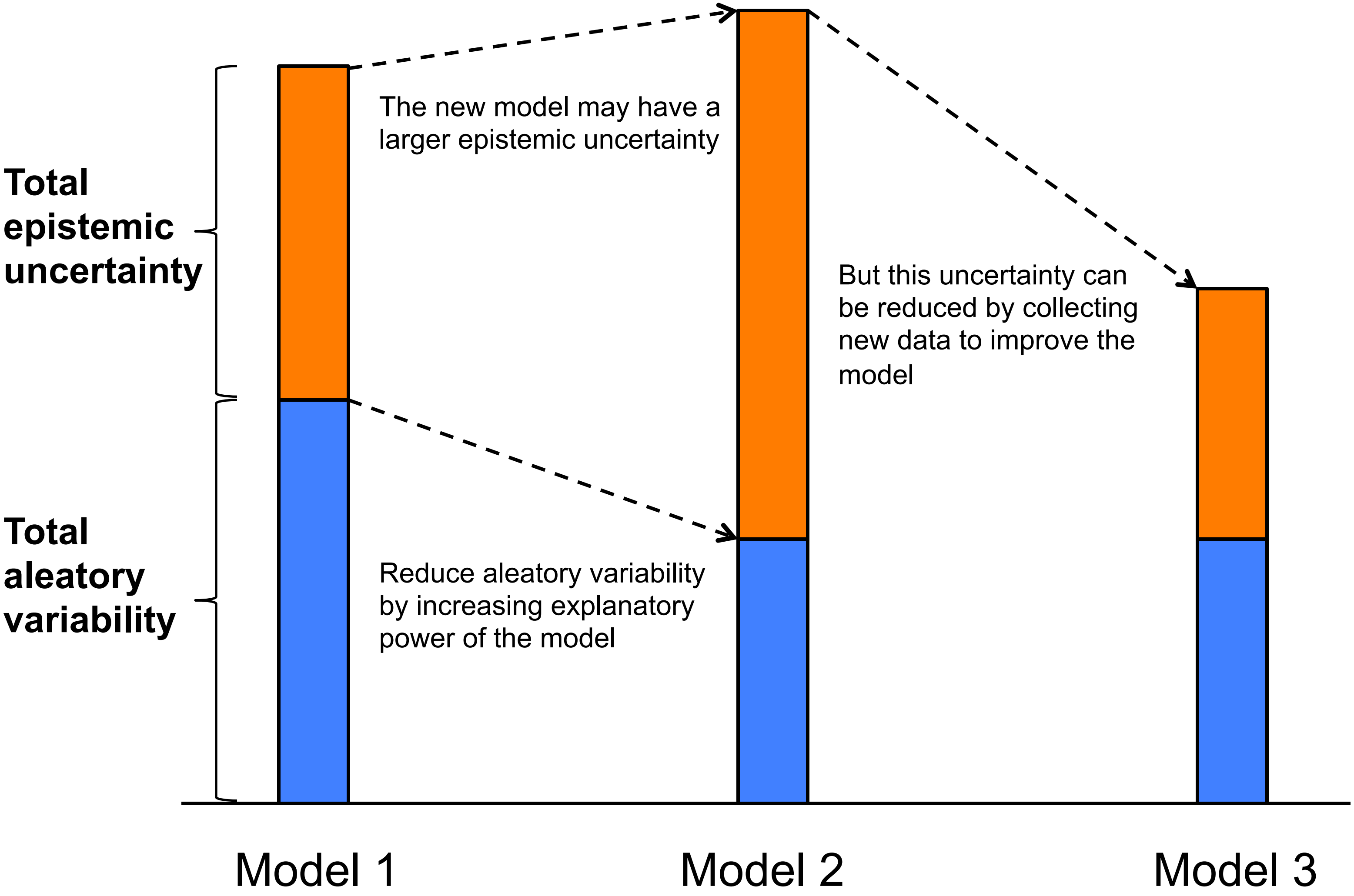
Computational Statistics for CyberShake 14.2

- Reservation for 700 XE nodes, 200 XK nodes
- 1144 CyberShake sites
  - 568 with SGT CPU
    - 2792 sec/job x 313.8 nodes = 243.4 node-hrs
    - Queue time: mean 973 sec, median 191 sec
  - 568 with SGT GPU
    - 1338 sec/job x 100 nodes = 37.2 node-hrs (6.5x efficiency improvement)
    - Queue time: mean 2889 sec, media 731 sec
- 99.8 million tasks produced 470 million seismograms
  - 81 tasks/sec
- 31,463 jobs submitted remotely to the Blue Waters queue
- 860 TB of data managed
  - 57 TB output files
  - 12 TB staged back to SCEC storage

ABF Variance Analysis



Reduction of Aleatory Variability



Total epistemic uncertainty

Total aleatory variability

The new model may have a larger epistemic uncertainty

Reduce aleatory variability by increasing explanatory power of the model

But this uncertainty can be reduced by collecting new data to improve the model