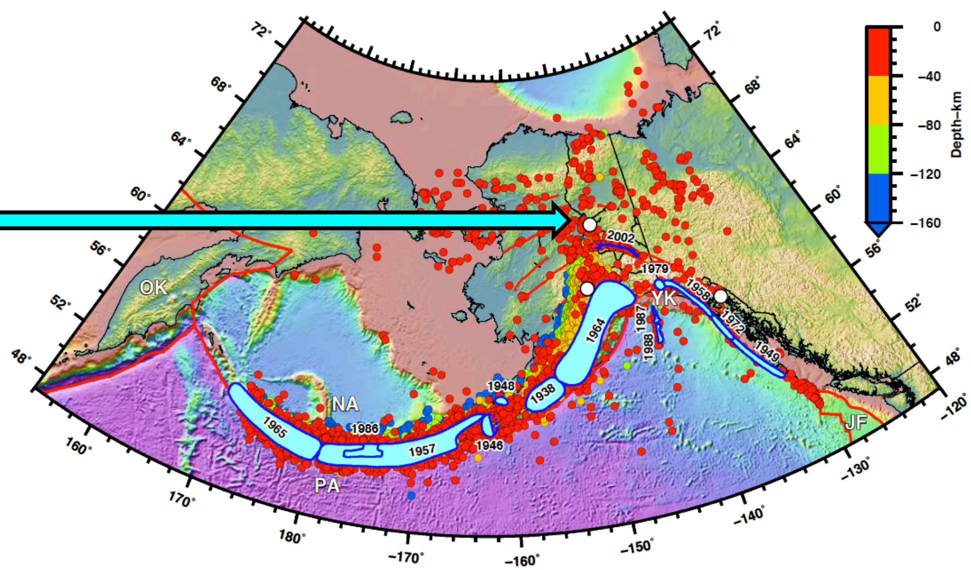
Navigating a sea of choices within a large-scale tomographic inversion

Carl Tape
carltape@gi.alaska.edu
Geophysical Institute
University of Alaska Fairbanks



QUEST Annual Workshop
Iceland
July 16, 2011

A quick point about names

- (full) waveform inversion
- (full) waveform tomography
- (full) wavefield tomography
- traveltime tomography
- adjoint tomography

A particular study might be described by any one of these.

Better questions to ask:

1. What are the data?
2. What is the misfit function?
3. What is the forward modeling tool?
4. What is the inverse modeling tool?

Background, papers, and seismograms for southern California:

- see Carl Tape webpage, then follow link

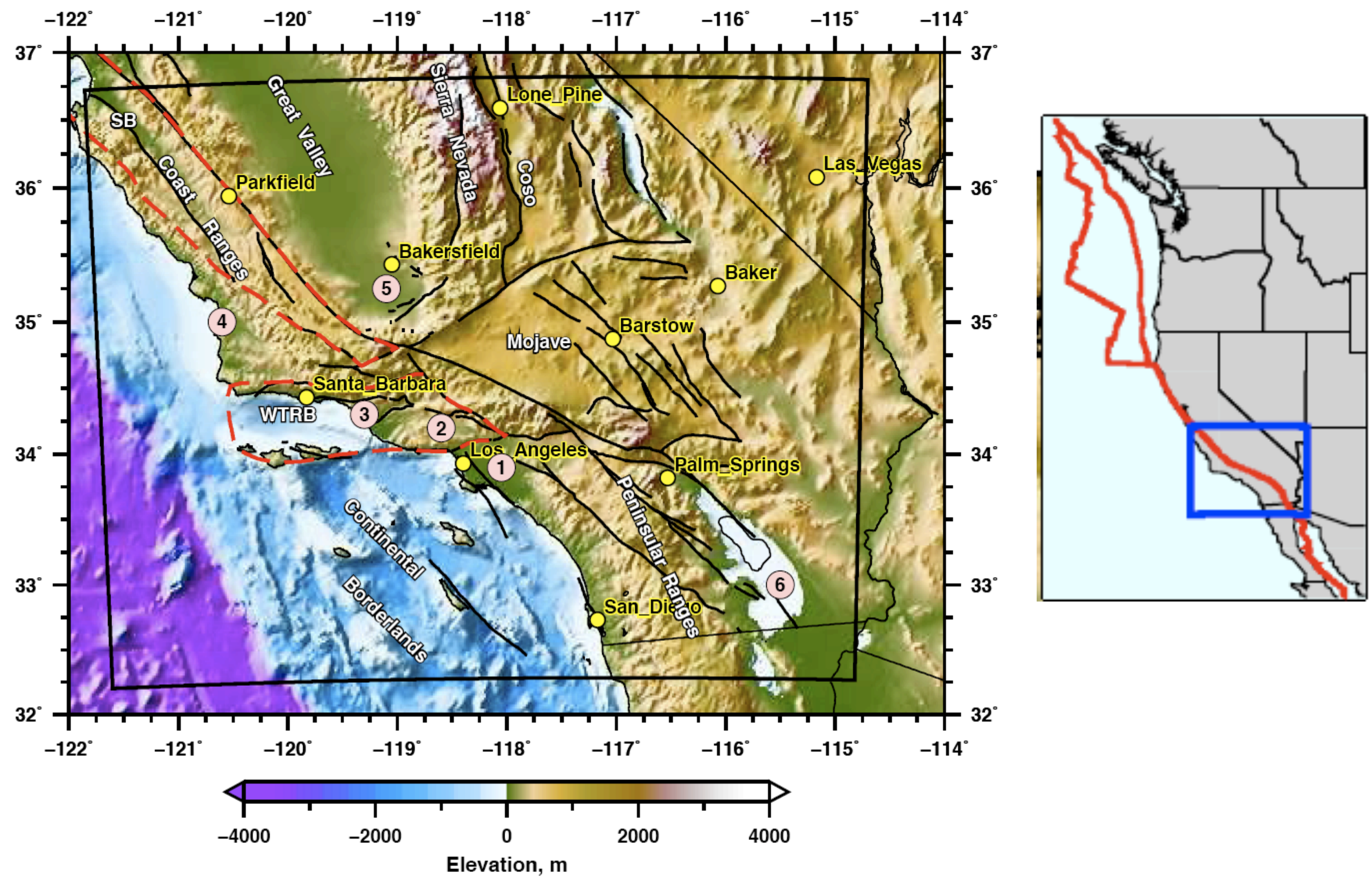
Seismic tomography using 3D reference models, wavefield simulations, and iterative methods should – and does – work.

Time-domain examples:

- industry scale
- crustal scale
  - southern California: Tape et al. (2009, 2010), Po Chen
- regional/continental
  - Australia: Fichtner et al. 2009, 2010
  - Europe: Fichtner; Zhu/Tromp

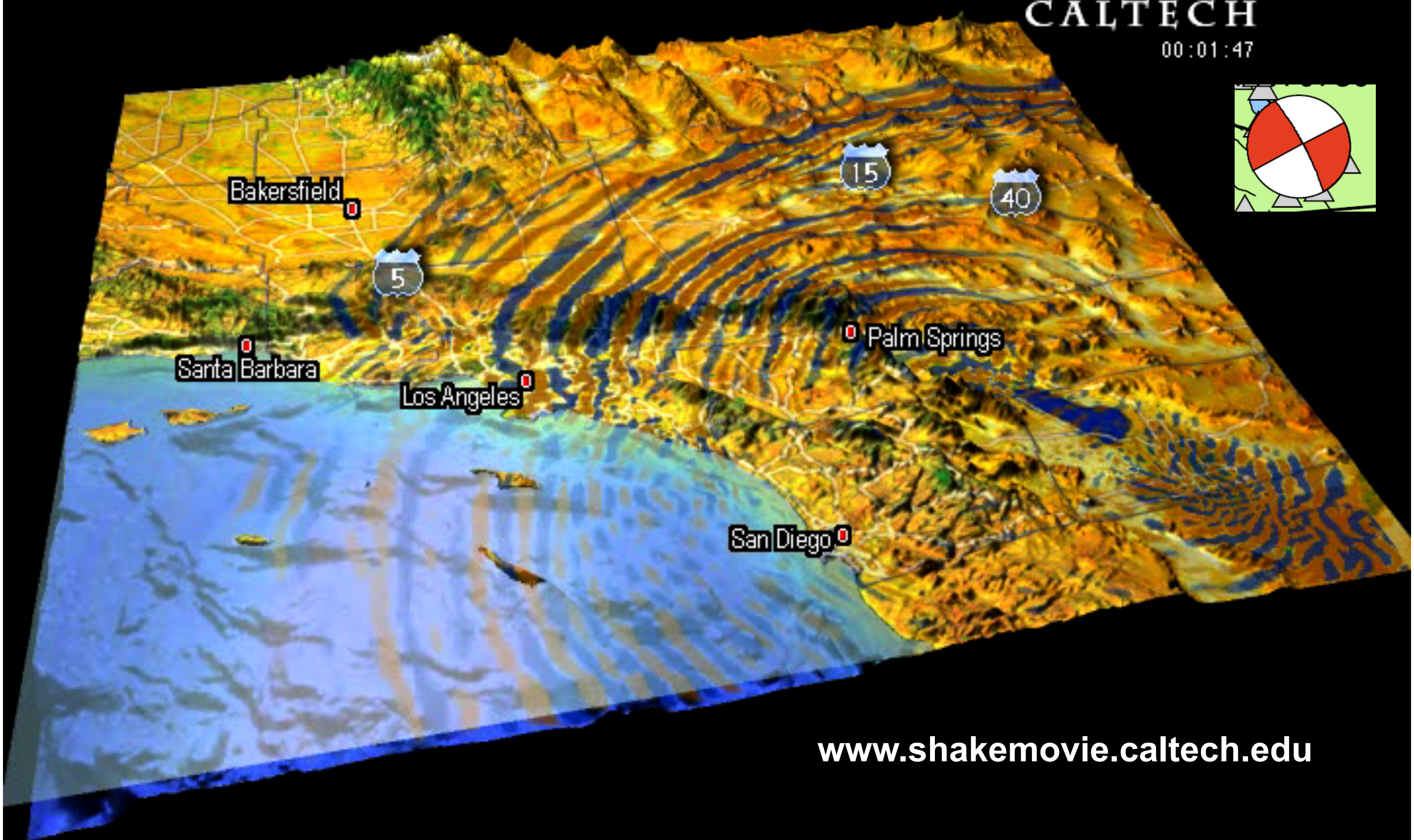
Frequency-domain examples from industry:

- Plessix, Sirgue, and more

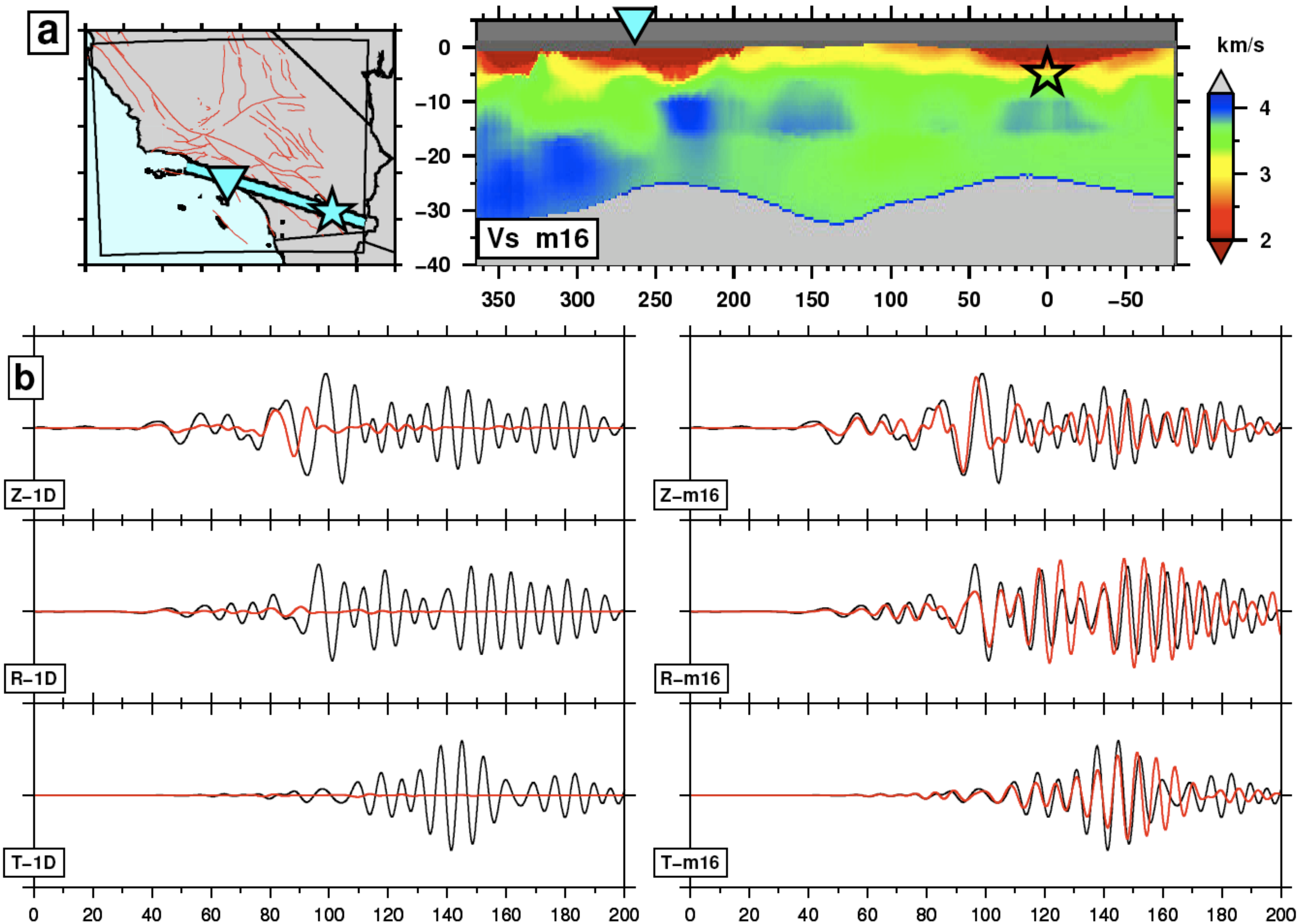


–122˚
–121˚
–120˚
–119˚
–118˚
–117˚
–116˚
–115˚
–114˚
37˚
36˚
35˚
34˚
33˚
32˚
SB
Great Valley
Sierra Nevada
Coso
Lone_Pine
Las_Vegas
Parkfield
Coast Ranges
Bakersfield
Baker
Barstow
Mojave
Santa_Barbara
WTRB
Los_Angeles
Palm_Springs
Peninsular Ranges
Continental Borderlands
San_Diego
1
2
3
4
5
6
–4000
–2000
0
2000
4000
Elevation, m

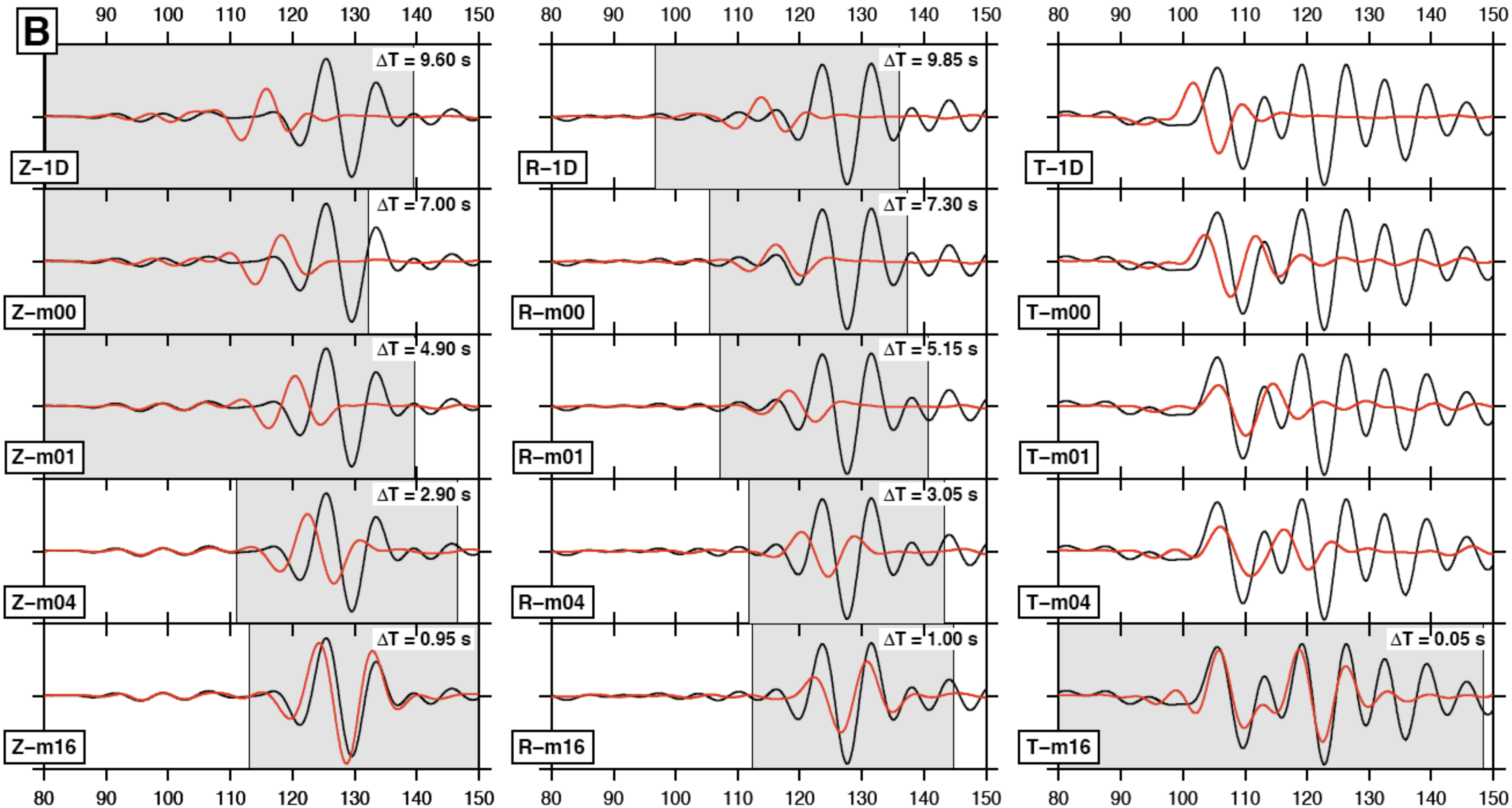
Sept 2, 2005 Mw 5.0 Salton trough



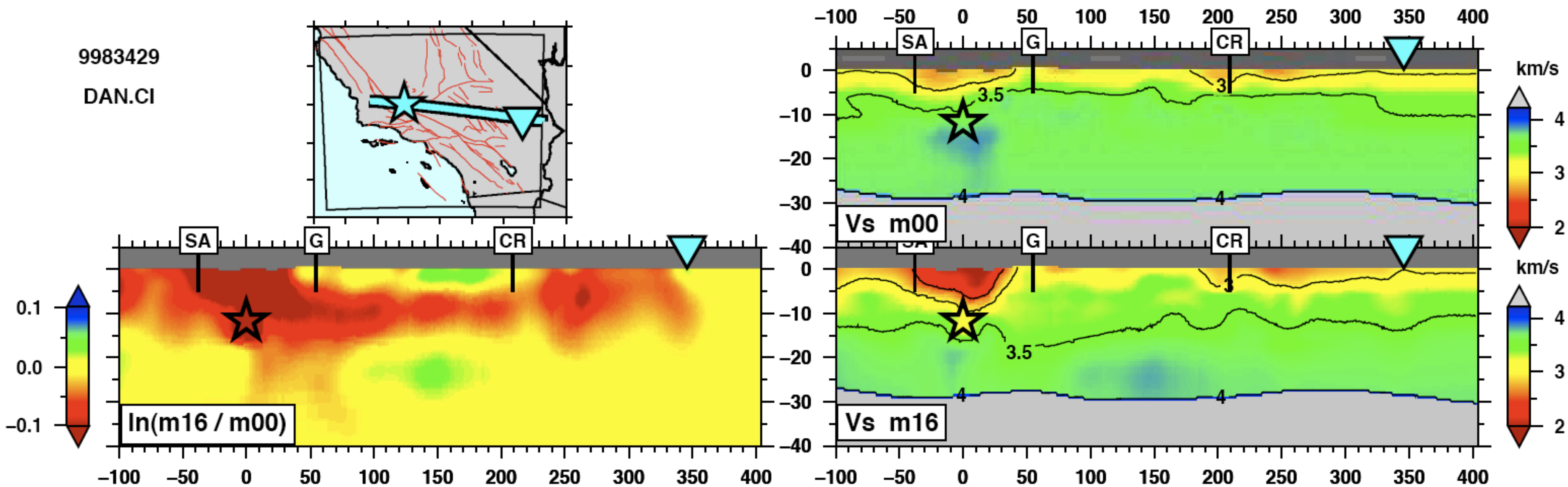
www.shakemovie.caltech.edu



a
km/s
4
3
2
0
−10
−20
−30
−40
350
300
250
200
150
100
50
0
−50
Vs m16
b
Z−1D
R−1D
T−1D
Z−m16
R−m16
T−m16
0
20
40
60
80
100
120
140
160
180
200



B
Z–1D
ΔT = 9.60 s
Z–m00
ΔT = 7.00 s
Z–m01
ΔT = 4.90 s
Z–m04
ΔT = 2.90 s
Z–m16
ΔT = 0.95 s
R–1D
ΔT = 9.85 s
R–m00
ΔT = 7.30 s
R–m01
ΔT = 5.15 s
R–m04
ΔT = 3.05 s
R–m16
ΔT = 1.00 s
T–1D
T–m00
T–m01
T–m04
T–m16
ΔT = 0.05 s
80
90
100
110
120
130
140
150



9983429
DAN.CI
SA
G
CR
ln(m16 / m00)
Vs m00
Vs m16
km/s
3.5
3
4

Question: What region? What data? What scale?

1. Where are the seismic stations?

2. What level of “discovery” are you aiming for? You should always be able to better fit seismic waveforms. You will refine structure, but not necessarily discover new structure.
   - Suggestion: Tackle structure that cannot be imaged well with less accurate techniques. Active tectonics!

3. You can always find waveforms that do not fit by changing the target periods. **We do not have the case of many models fitting the data equally well.**

4. Is your objective to fit the seismic wavefield or to resolve the full domain of the physical model? (Both, of course!)

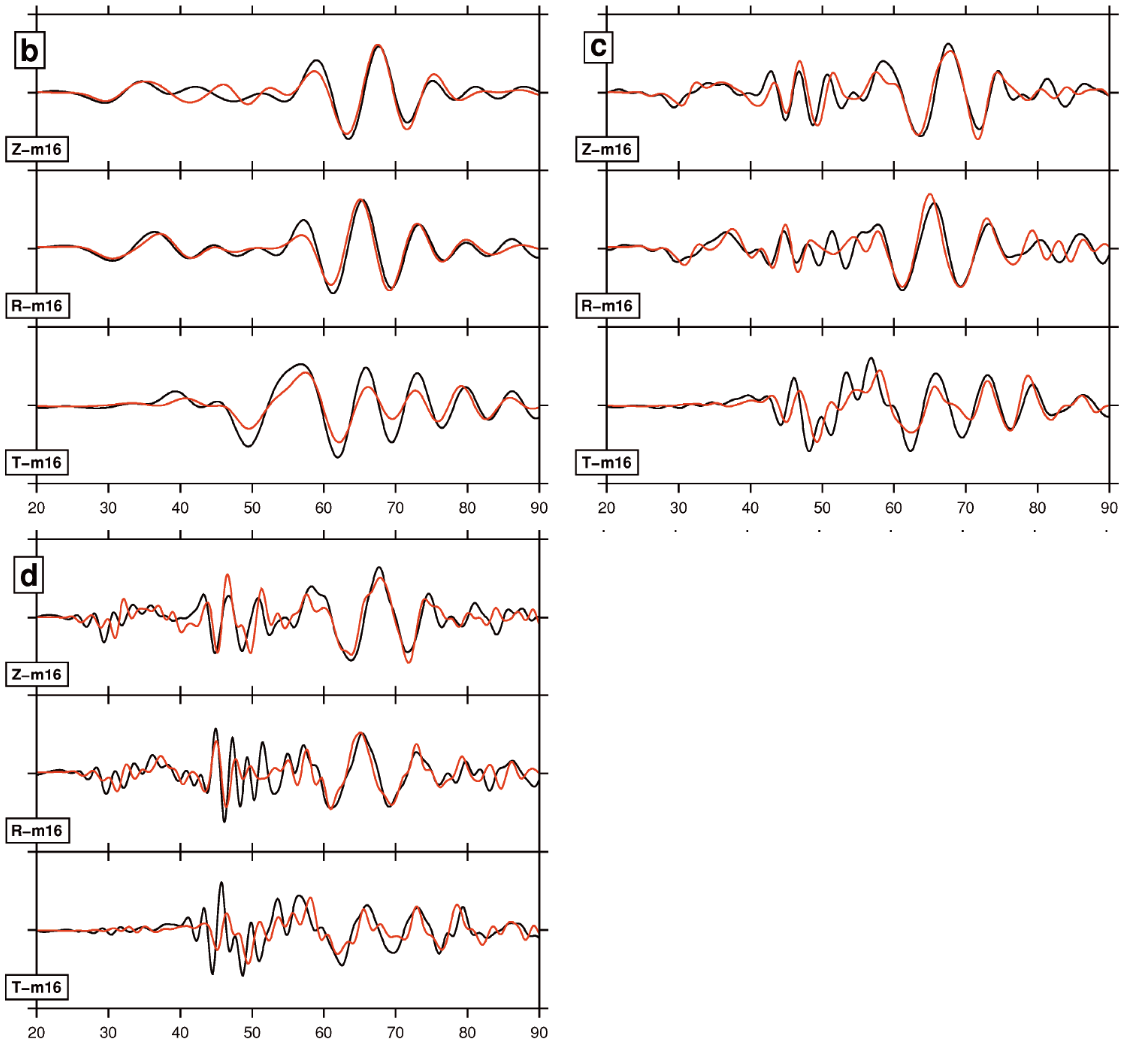
5. Earthquakes? Ambient noise? Receiver functions? Gravity?

6. Spend considerable time getting all available waveforms and station responses. These don’t change with each iteration!

What scale? = What bandpass? = What minimum Vs?

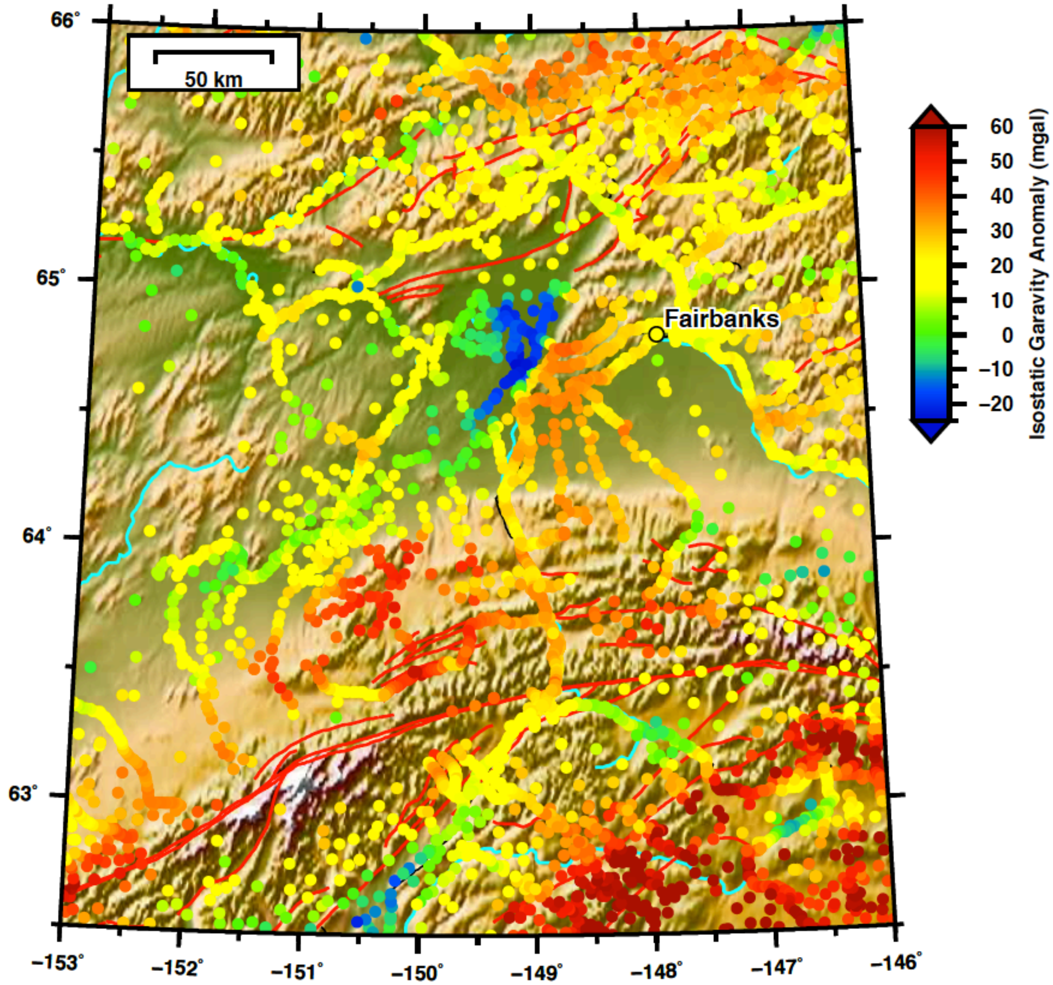
6 – 30 s

3 – 30 s



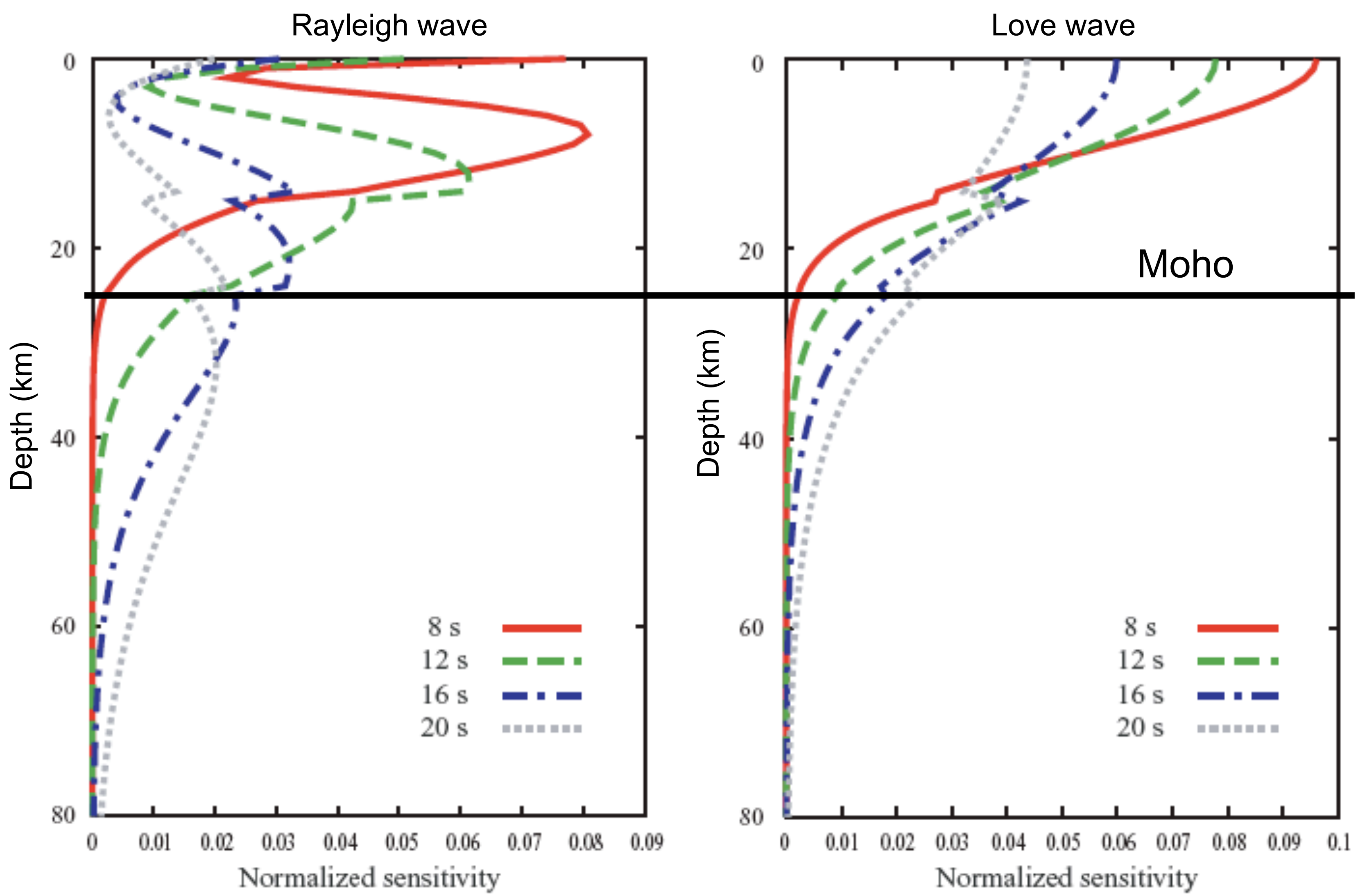
2 – 30 s

Example of central Alaska:
No seismic velocity model + very active tectonics = plenty to discover



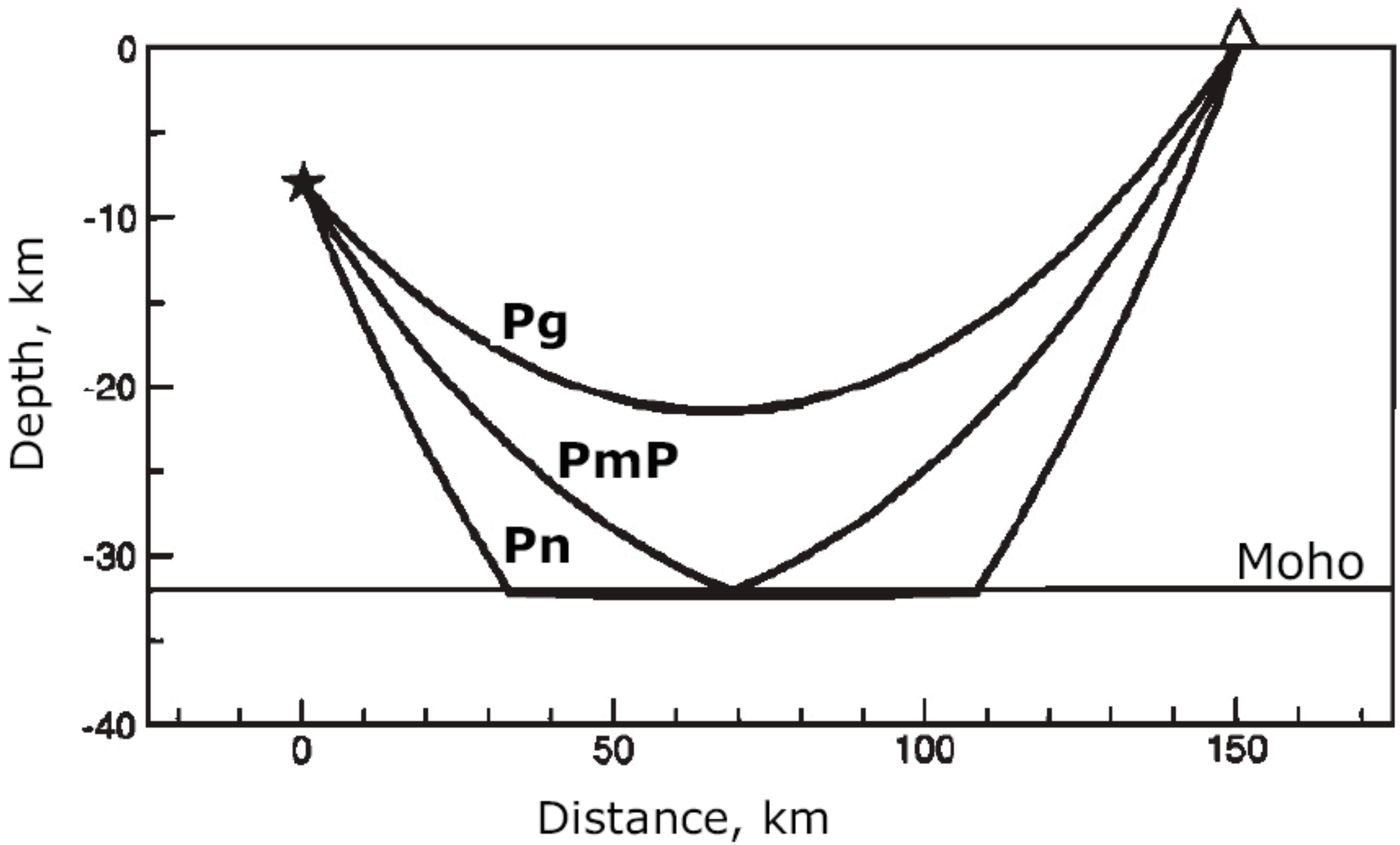
USGS gravity data set:
Saltus et al. (2006)

Surface wave sensitivities in a 1D model



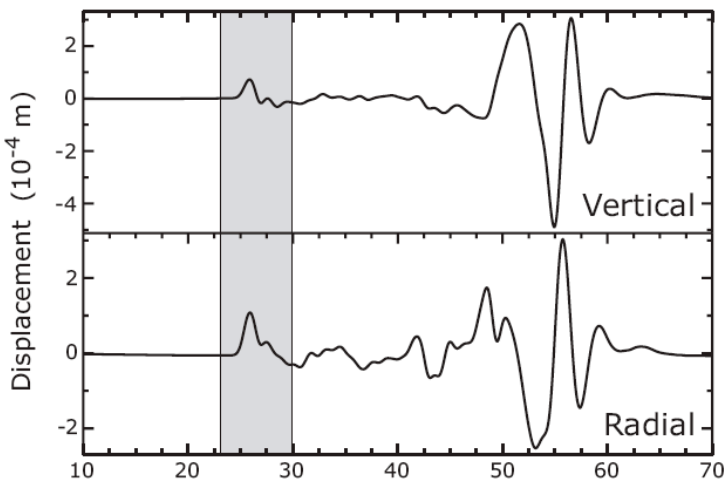
Lin, Moschetti, Ritzwoller (2008)

Body wave sensitivities in a 1D model



Richards-Dinger and Shearer (1998)

Sensitivity kernel for a P wave in a 1D model



Distance: Δ = 157 km

Bandpass: [2 s, 30 s]

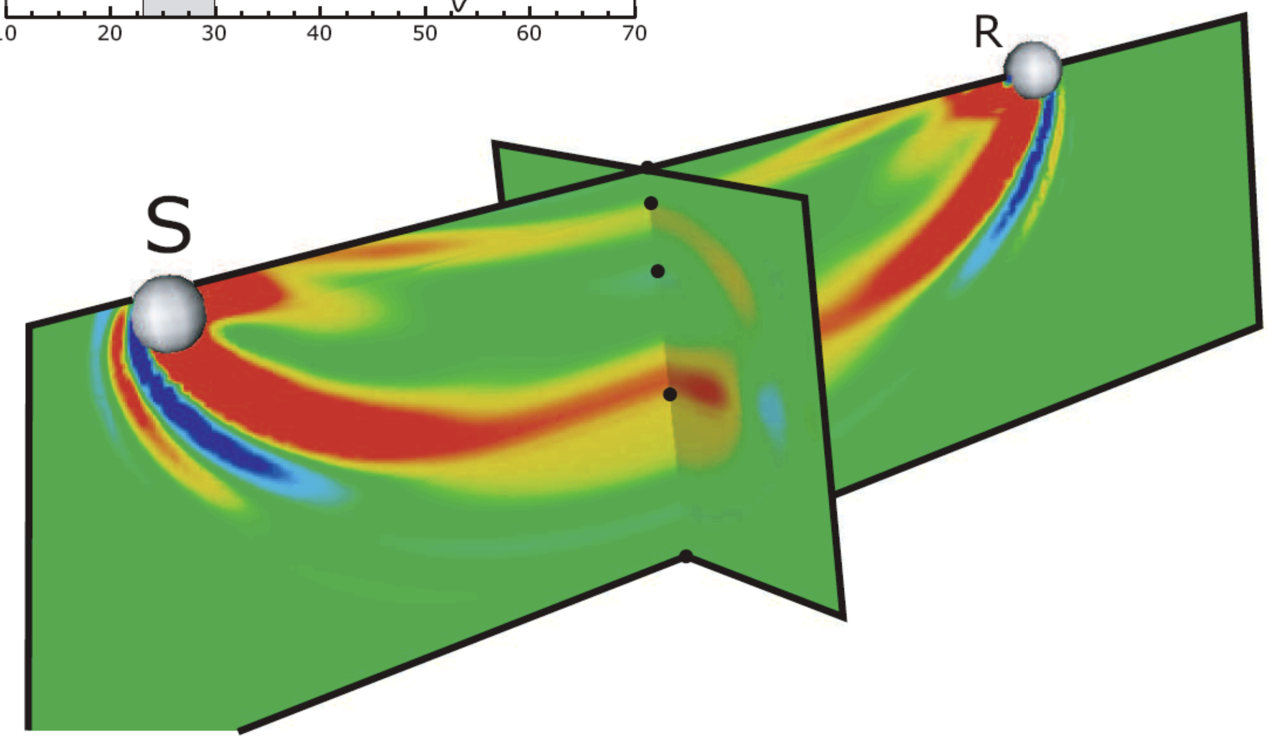
Reference model: SoCal-1D

Source: Explosion (depth 2.75 km)

Time window: P arrival (Z, R)

Measurement: CC traveltime

Model parameter: Vp



Question: What sources?

1. Spend considerable time testing many sources!

2. origin time, hypocenter, M_0, moment tensor (3-4 parameters)

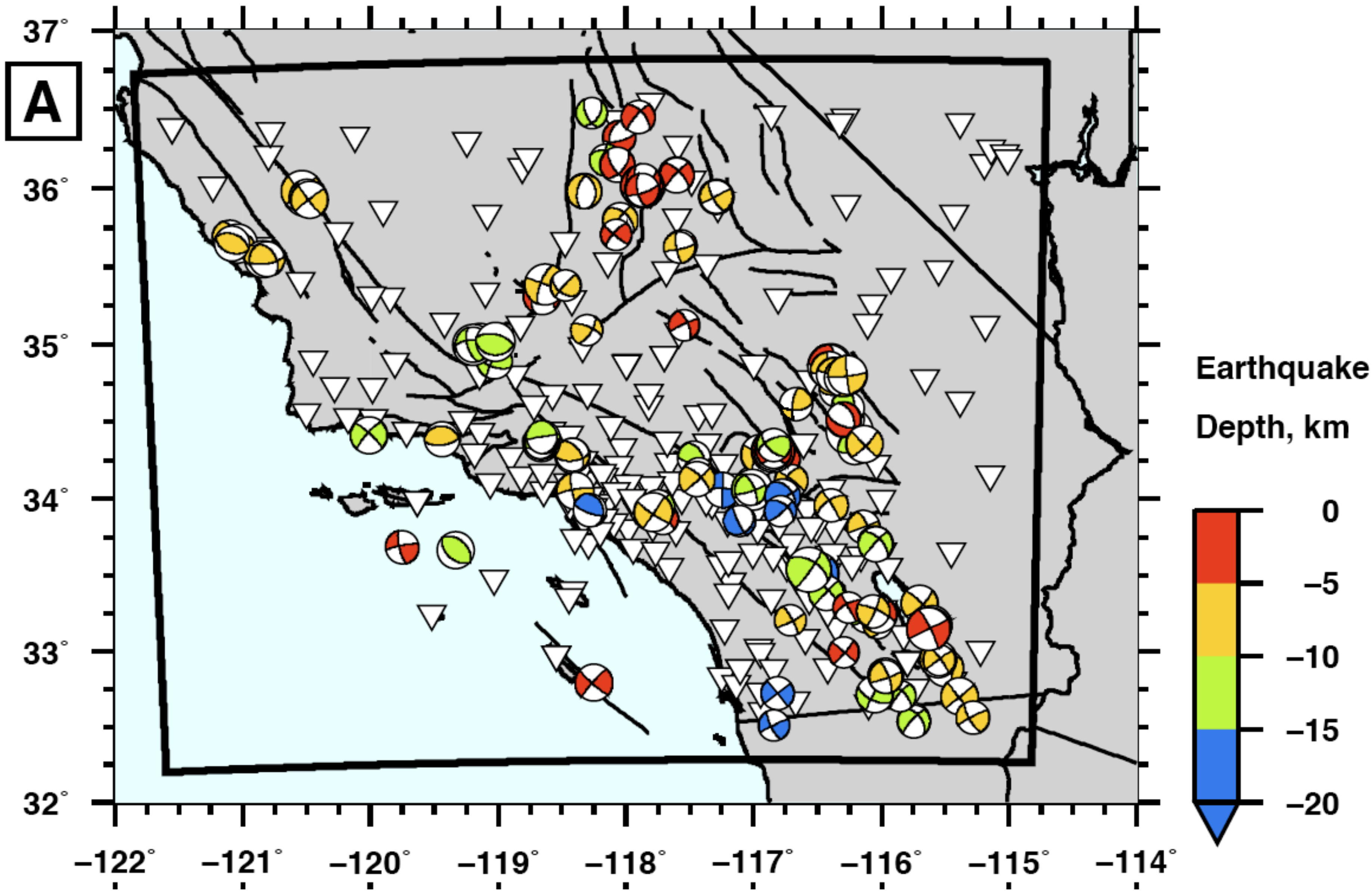
3. Collect sources that best sample the medium by obtaining spatial coverage (key: depth) and different mechanisms.

4. Events in the same location are okay.
    1. Hypocenters apart by a couple km can have totally different waveforms at shorter periods.
    2. At longer periods, discrepancies among event kernels for similar sources may help isolate bad sources.

5. Set aside validation events for misfit only but not the inversion.

Figure D.26: Source mechanisms considered in the southern California tomography study (201 through 208 out of 294).

143 earthquakes for tomographic inversion



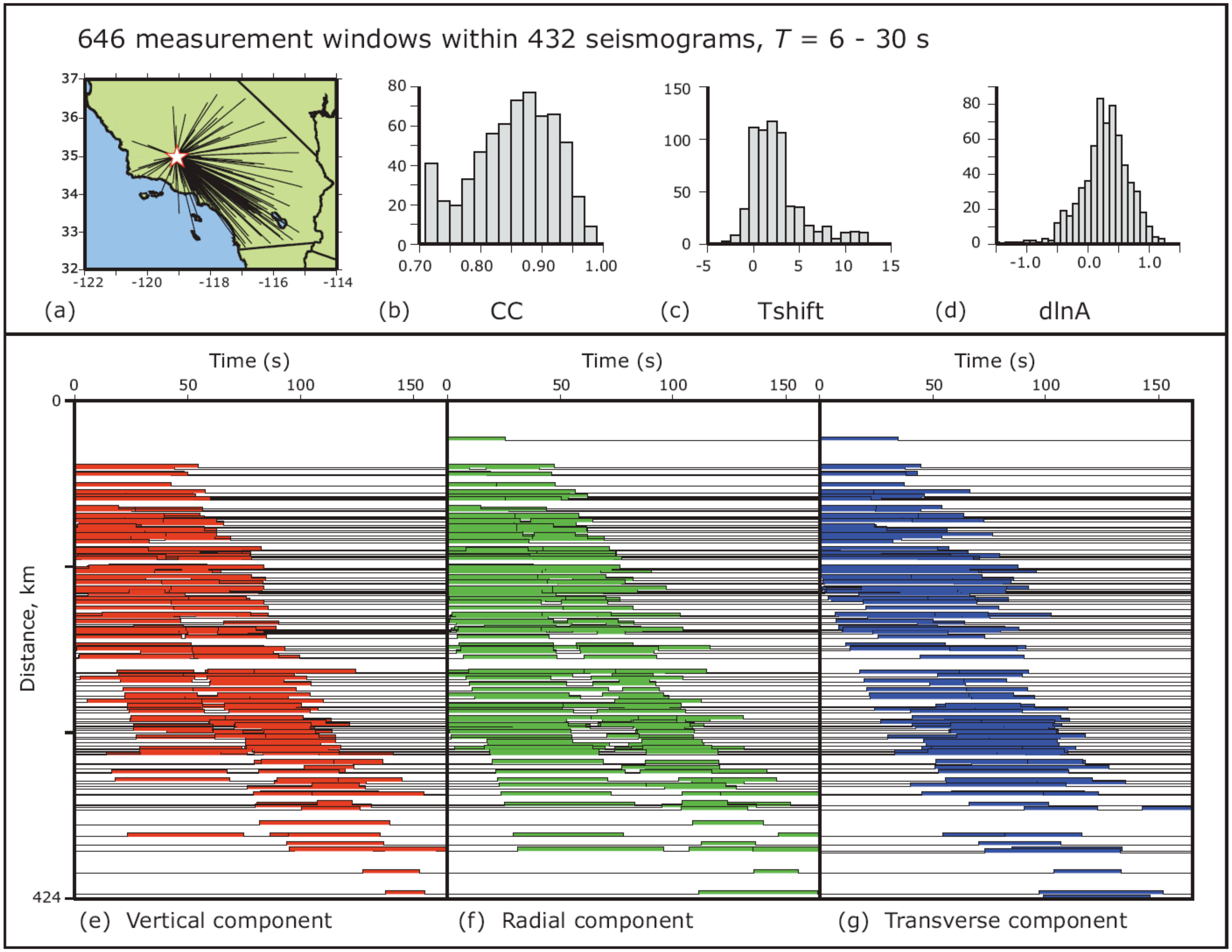
Question: What misfit function?

1. **What are the data?**
2. **What are the uncertainties in the data? Or: How are the data weighted?**
3. **What is the forward model for the synthetics?**
4. **What are the model variables (e.g., Vp, Vs)? What is the model parameterization?**
5. **What is the regularization? Or: What prior smoothness is assumed?**

Least-squares misfit function (Tarantola, 2005)

$$
\begin{aligned}
2\,S(\mathbf{m}) &= \|\,\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\,\|^2_{\text{D}} + \|\,\mathbf{m} - \mathbf{m}_{\text{prior}}\,\|^2_{\text{M}} \\
&= (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}})^t\, \mathbf{C}_{\text{D}}^{-1}\, (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}) + (\mathbf{m} - \mathbf{m}_{\text{prior}})^t\, \mathbf{C}_{\text{M}}^{-1}\, (\mathbf{m} - \mathbf{m}_{\text{prior}}) \ .
\end{aligned}
\tag{3.32}
$$

4 3 1 2 5

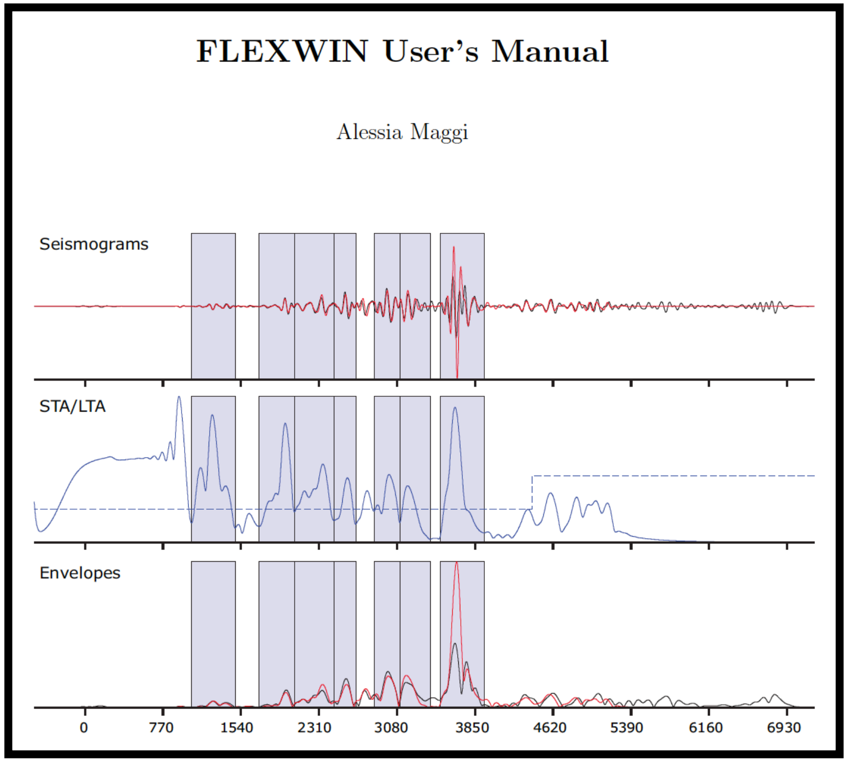


646 measurement windows within 432 seismograms, T = 6 - 30 s
(a)
(b) CC
(c) Tshift
(d) dlnA
Time (s)
Distance, km
(e) Vertical component
(f) Radial component
(g) Transverse component

Geophys. J. Int. (2009) doi: 10.1111/j.1365-246X.2009.04099.x

An automated time-window selection algorithm for seismic tomography

Alessia Maggi,[1] Carl Tape,[2] Min Chen,[2,*] Daniel Chao[2] and Jeroen Tromp[2,†]
[1]*Institut de Physique du Globe de Strasbourg (*UMR 7516 CNRS, *Université de Strasbourg/EOST), 5 rue René Descartes,* 67084 *Strasbourg Cedex, France.*
E-mail: alessia@sismo.u-strasbg.fr
[2]*Seismological Laboratory, California Institute of Technology, Pasadena,* CA 91125, *USA*



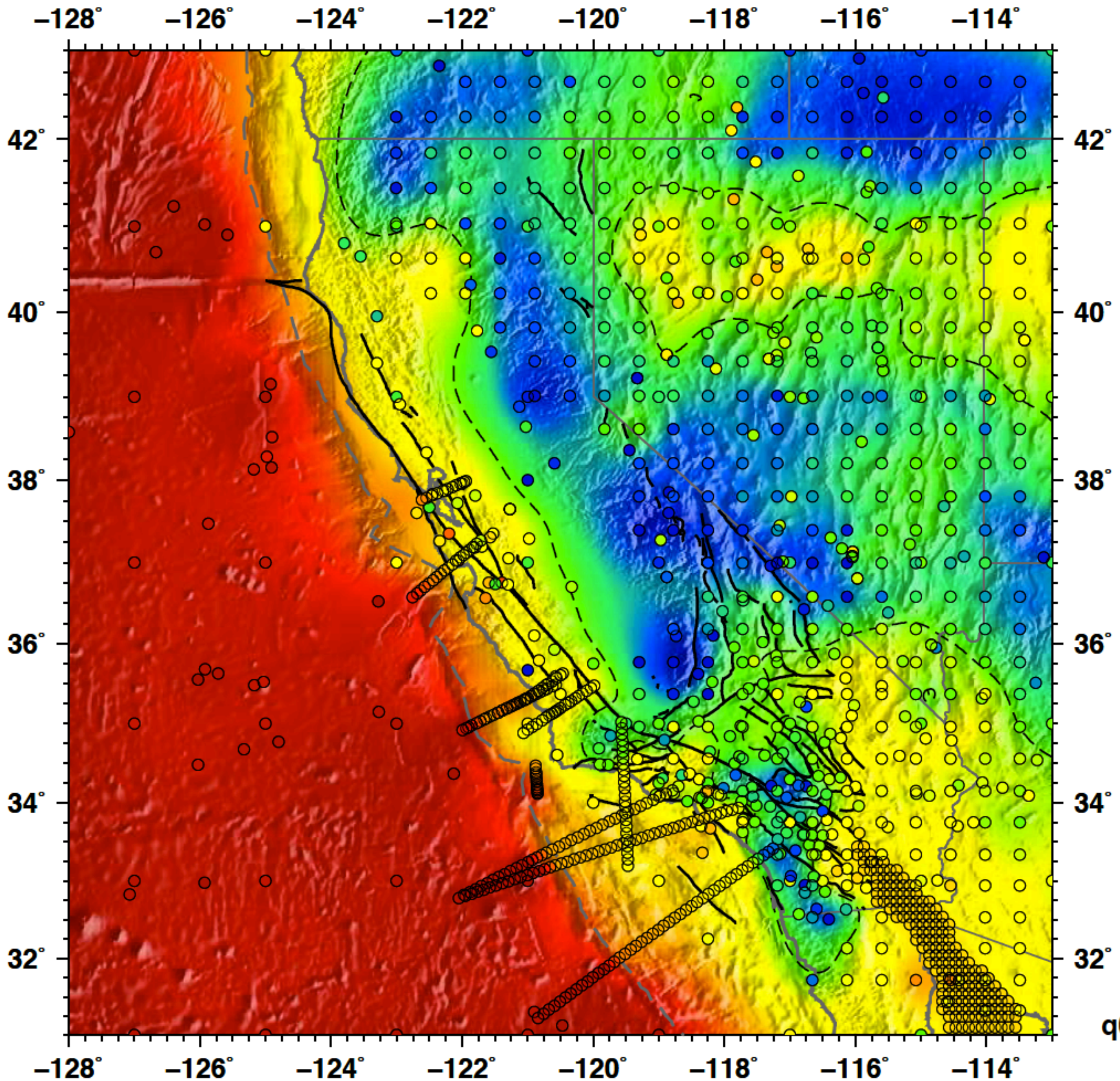
http://www.geodynamics.org/

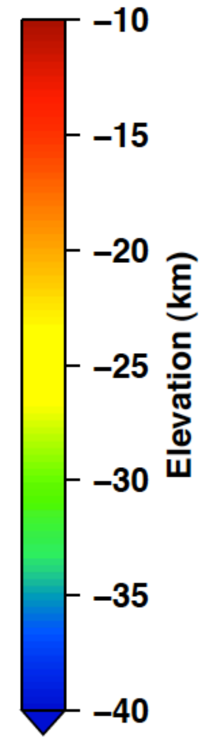
Question: What model do I start with?

1. Implement the definite constraints: topography, basin geometry from industry data, shallow basin Vp values from industry measurements.

2. Test the available 3D seismic velocity models. What data were used in constructing each model?

Moho Elevation (CVM-H 6.3)

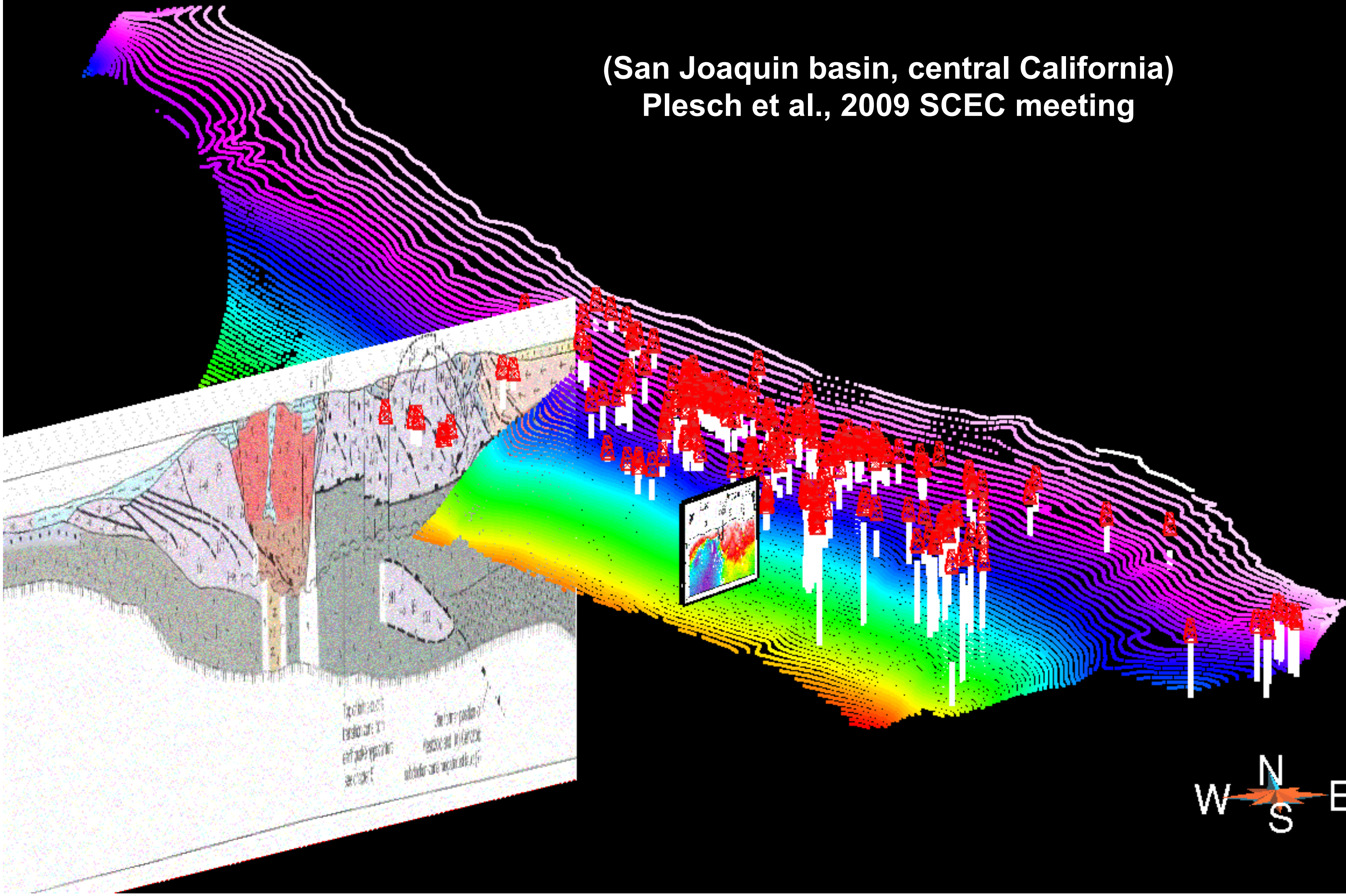




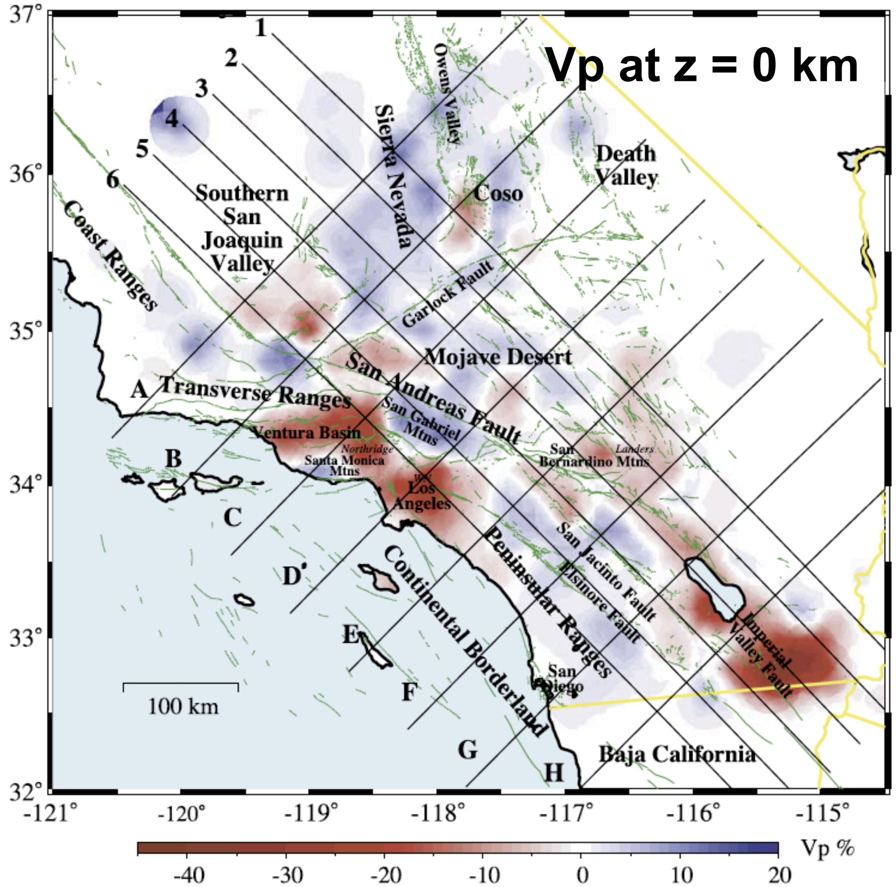
q02_q08_ic01_im01

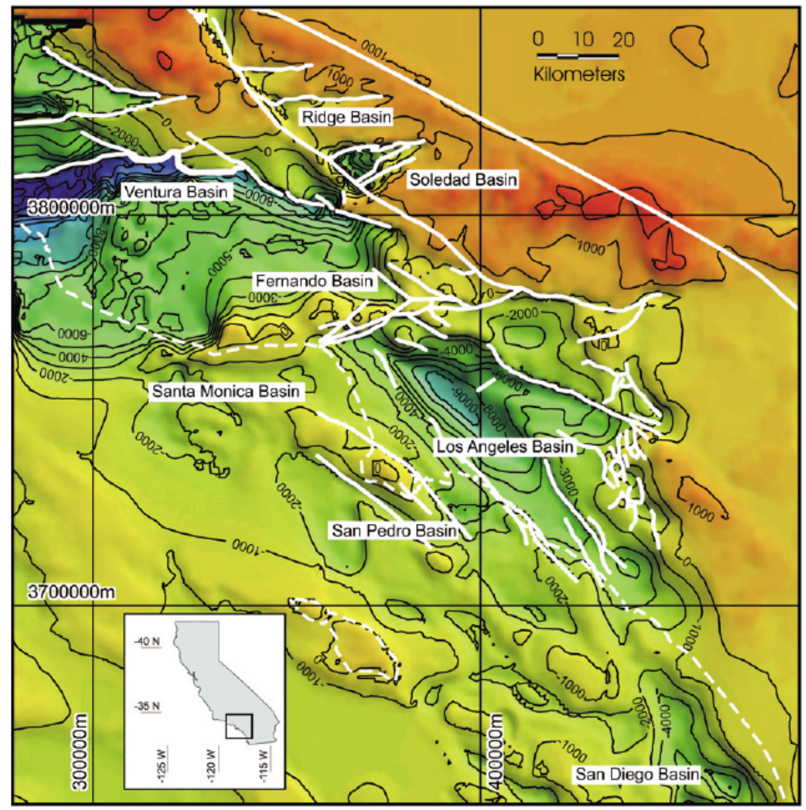
Many data sets are useful!

(San Joaquin basin, central California)
Plesch et al., 2009 SCEC meeting



Initial model of 3D wavespeed structure: SCEC Community Velocity Model





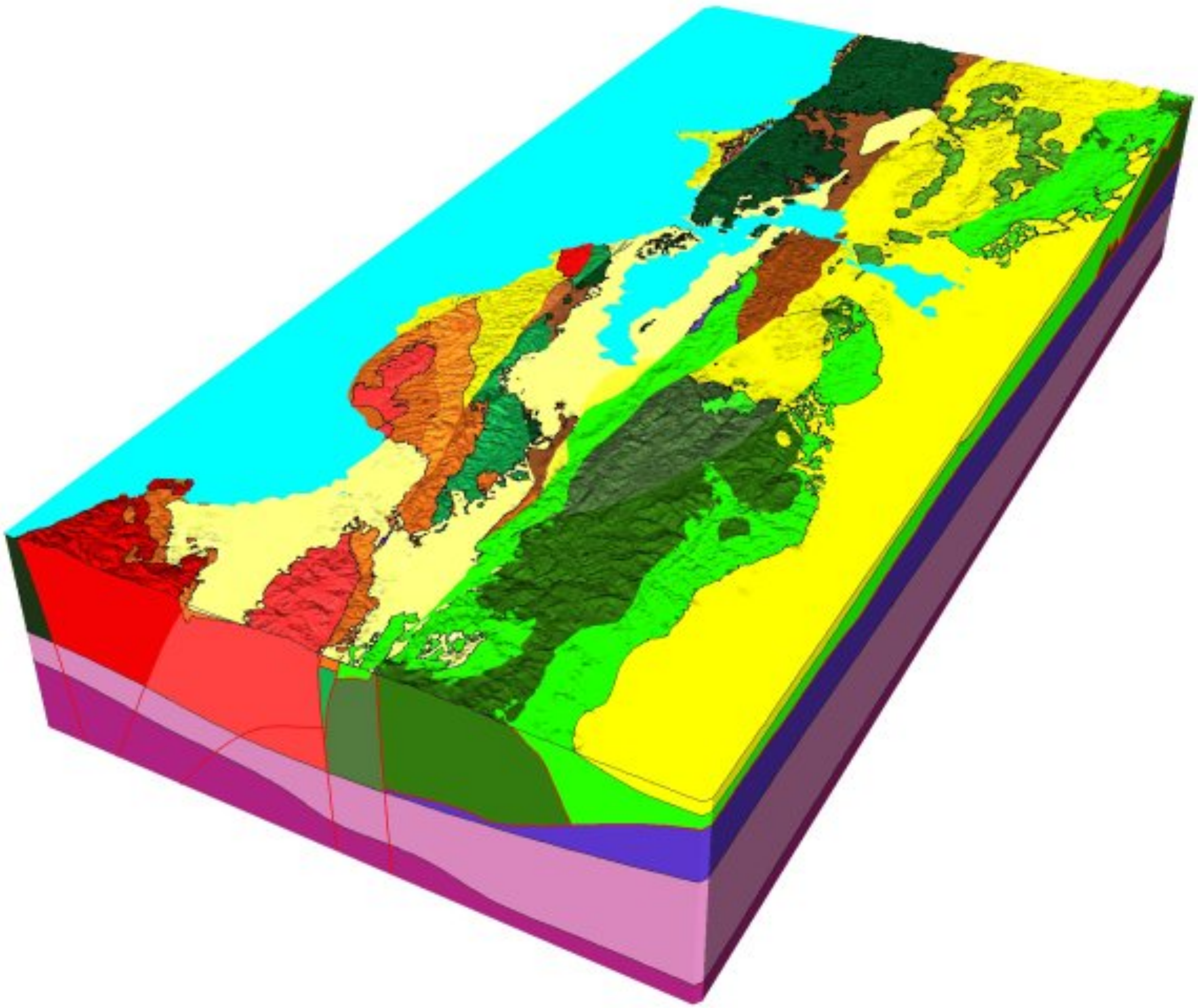
Background model:

Lin et al. (2007b)

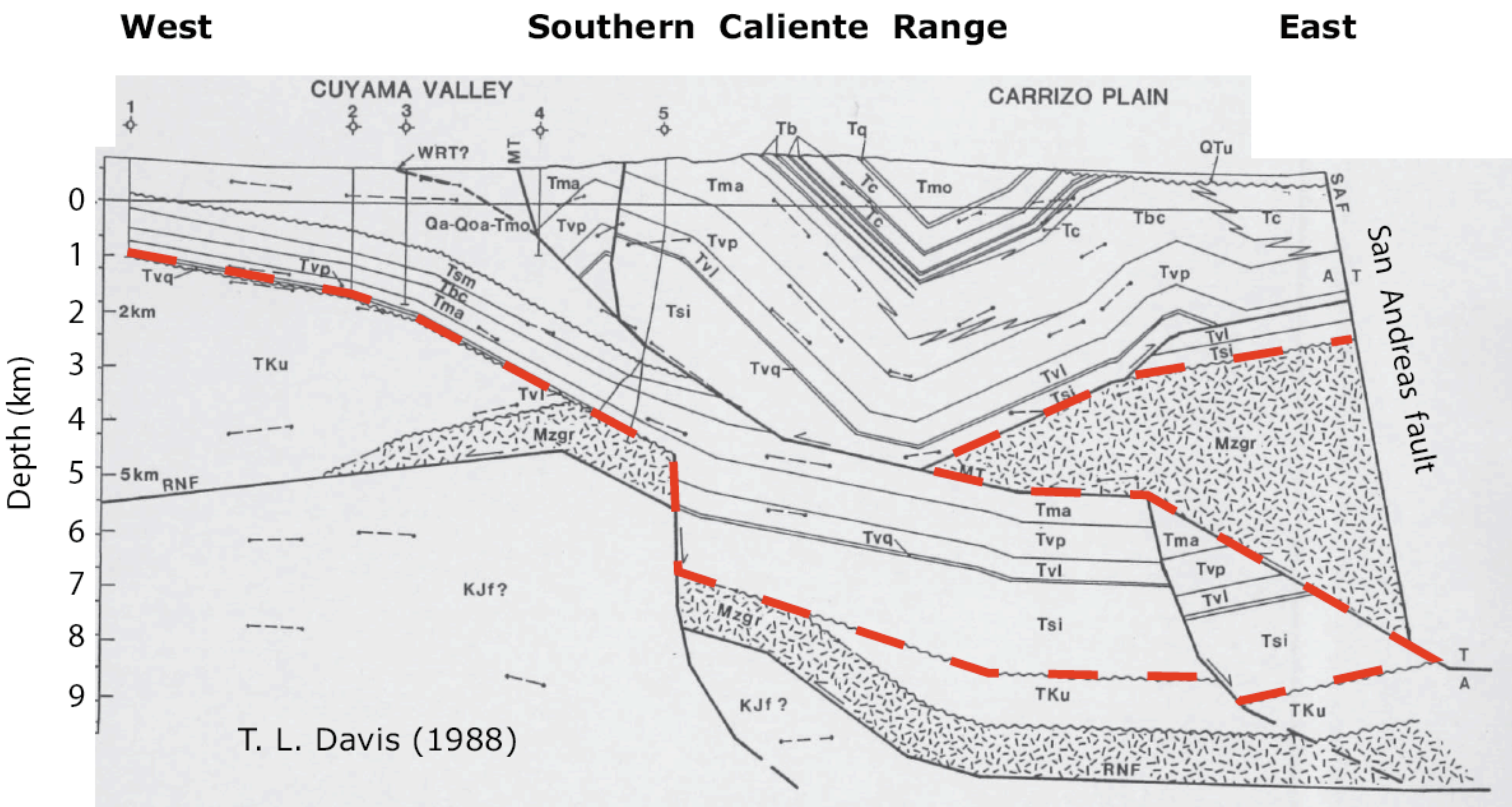
Basin models:

Suss and Shaw (2003)

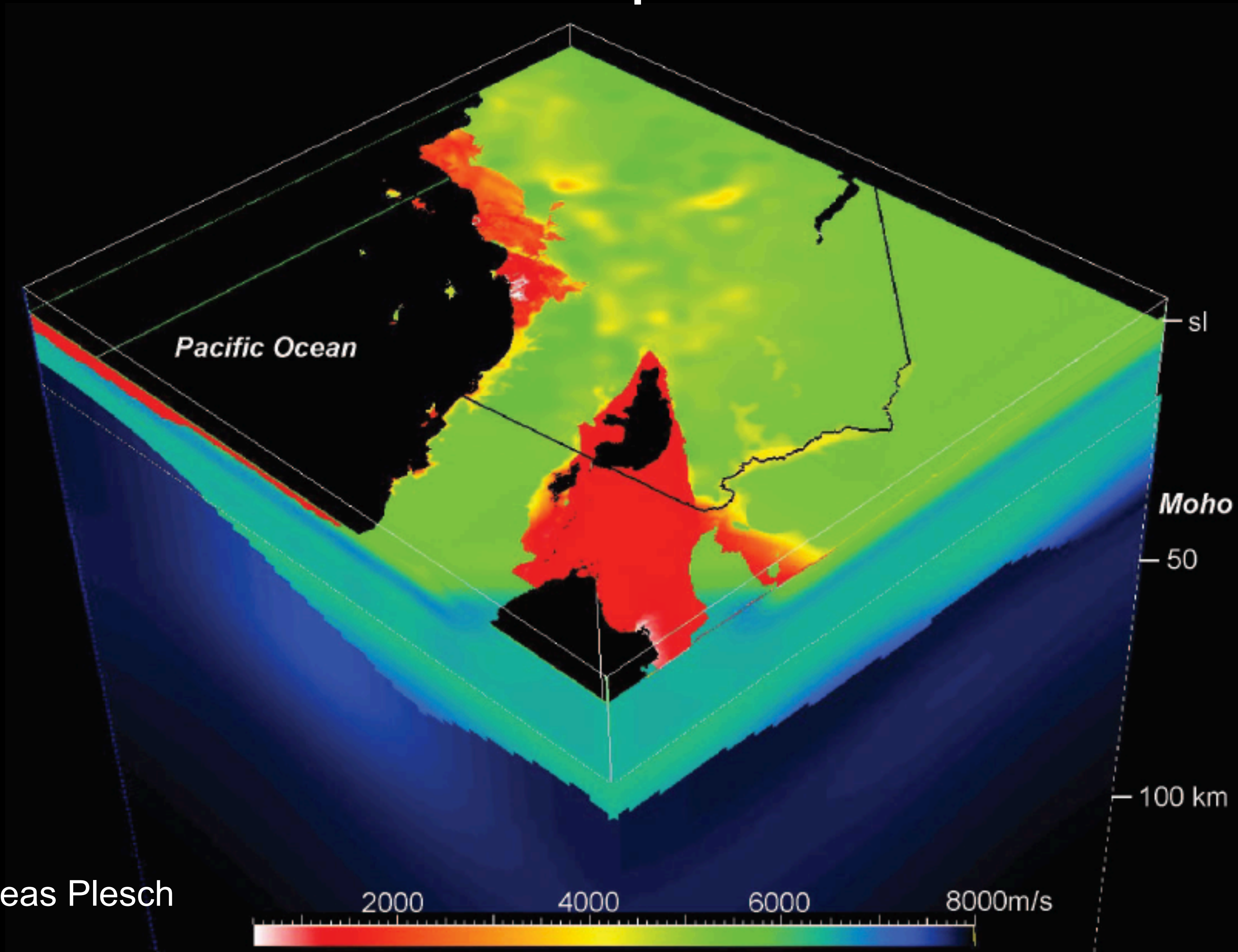
USGS Bay Area model



How do we incorporate 2D constraints?

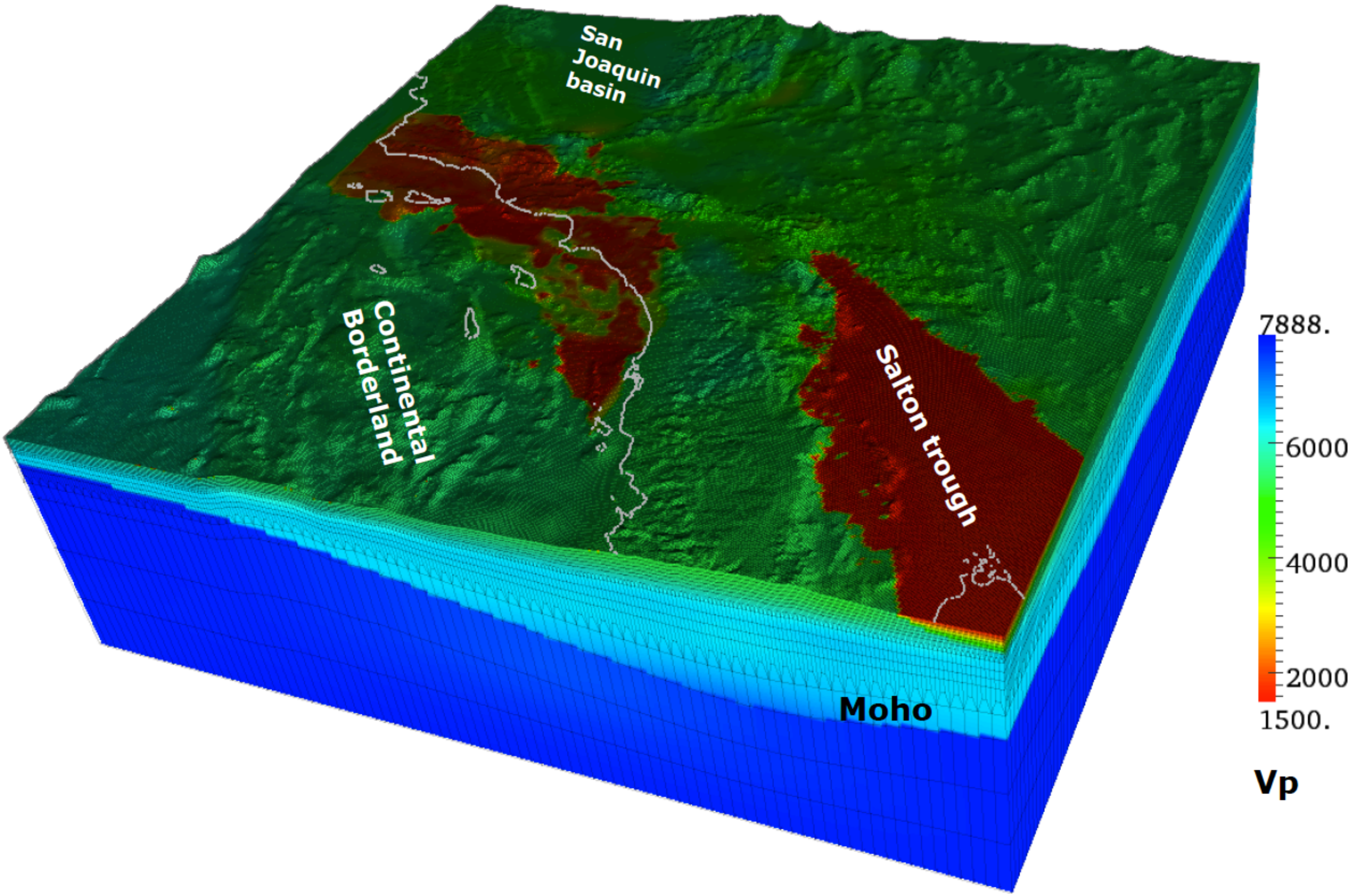


Southern California Earthquake Center CVM-H 6.3



Andreas Plesch

Hexahedral mesh using GEOCUBIT (E. Casarotti)



Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes

GJI 2011

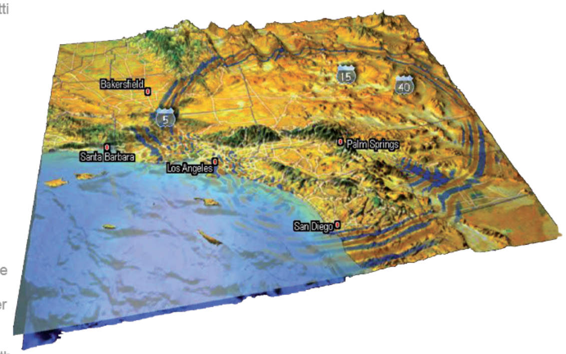
Daniel Peter,[1] Dimitri Komatitsch,[2,3] Yang Luo,[1] Roland Martin,[2] Nicolas Le Goff,[2] Emanuele Casarotti,[4] Pieyre Le Loher,[2] Federica Magnoni,[4] Qinya Liu,[5] Céline Blitz,[2] Tarje Nissen-Meyer,[6] Piero Basini[6] and Jeroen Tromp[1,7]

COMPUTATIONAL INFRASTRUCTURE FOR GEODYNAMICS (CIG)
PRINCETON UNIVERSITY (USA)
UNIVERSITY OF PAU, CNRS and INRIA (FRANCE)

SPECFEM 3D

User Manual
Version 2.0

Piero Basini
Céline Blitz
Ebru Bozdağ
Emanuele Casarotti
Min Chen
Hom Nath Gharti
Vala Hjörleifsdóttir
Sue Kientz
Dimitri Komatitsch
Jesús Labarta
Nicolas Le Goff
Pieyre Le Loher
Qinya Liu
Yang Luo
Alessia Maggi
Federica Magnoni
Roland Martin
Dennis McRitchie
Matthias Meschede
David Michéa
Tarje Nissen-Meyer
Daniel Peter
Brian Savage
Bernhard Schuberth
Anne Sieminski
Leif Strand
Carl Tape
Jeroen Tromp
Hejun Zhu

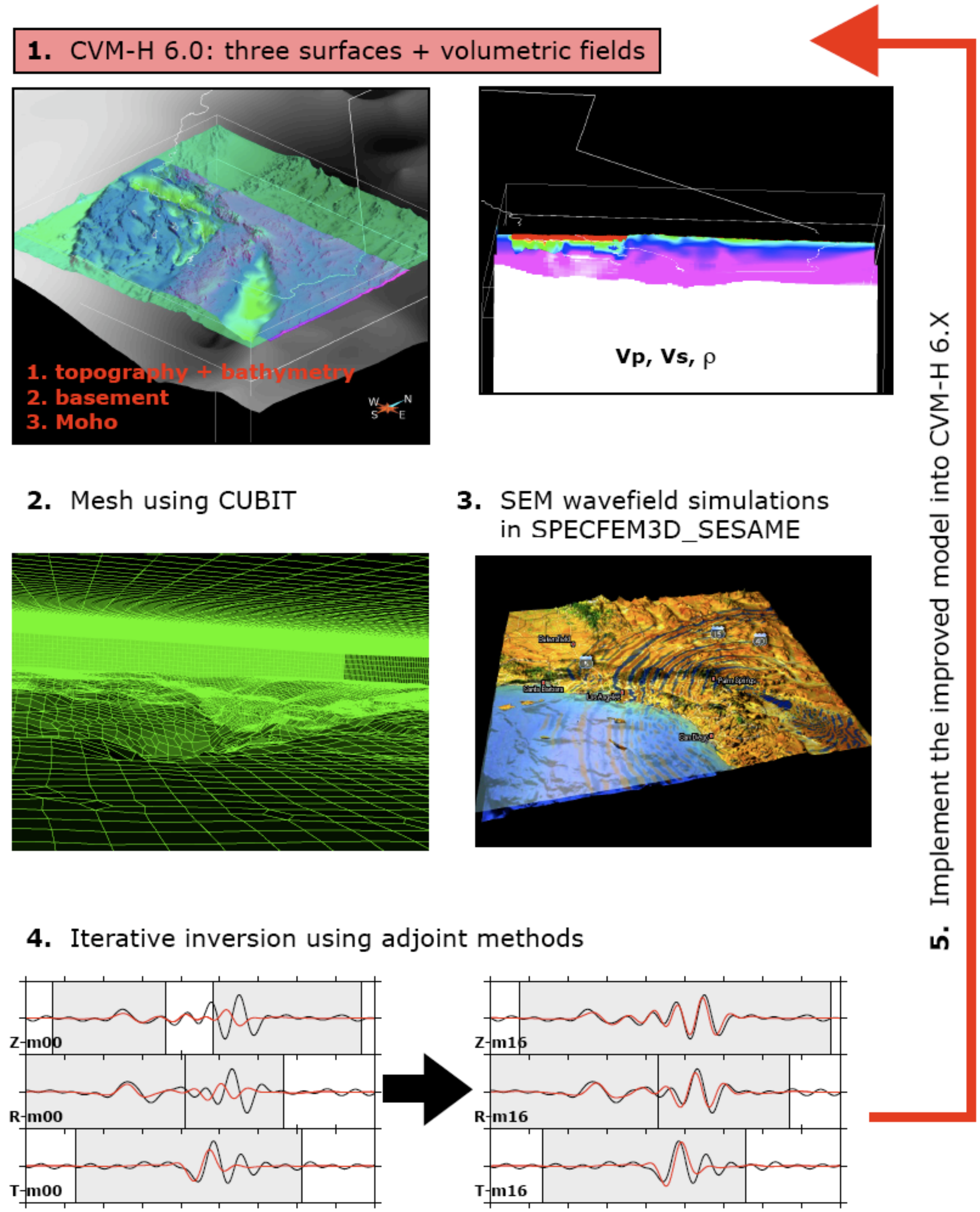


INRIA

TOMOGRAPHY STEPS

1. Specify initial model in terms of wavespeed structure and a set of earthquake sources.
2. Use “forward model” to generate synthetic seismograms.
3. Make measurements (evaluate the misfit function).
4. Compute the gradient of the misfit function (adjoint methods).
5. Iterate to new tomography models (minimize the misfit function).

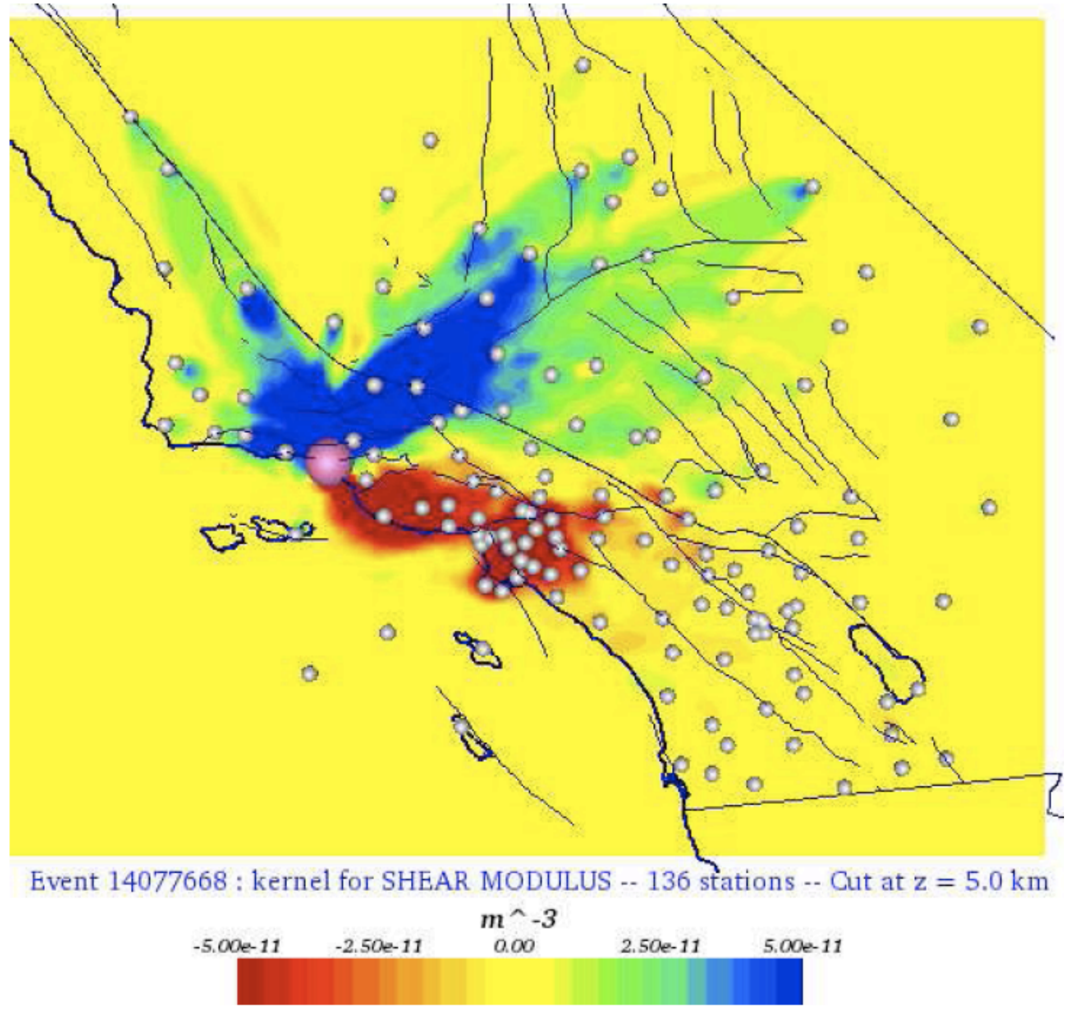
Yes, Step 1 may take longer than you expect, but it may substantially reduce the time involved with later steps.



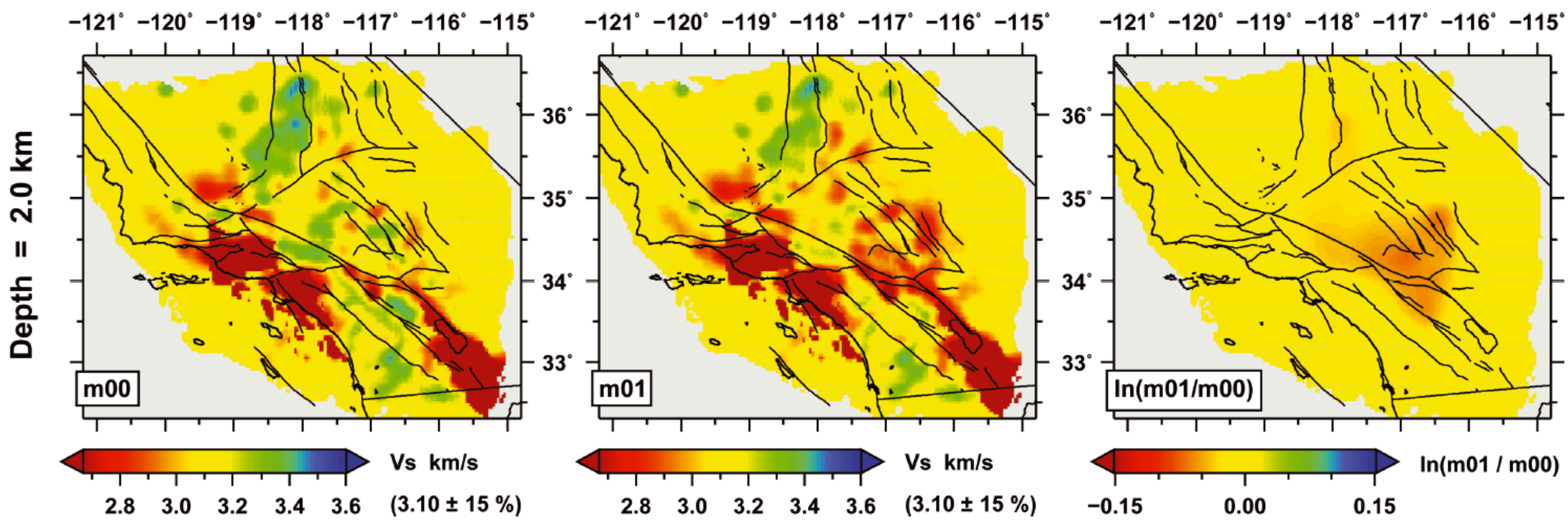
1. CVM-H 6.0: three surfaces + volumetric fields
1. topography + bathymetry
2. basement
3. Moho
W N S E
Vp, Vs, ρ
2. Mesh using CUBIT
3. SEM wavefield simulations in SPECFEM3D_SESAME
4. Iterative inversion using adjoint methods
Z-m00
R-m00
T-m00
Z-m16
R-m16
T-m16
5. Implement the improved model into CVM-H 6.X

Question: I've got the gradient of the misfit function – now what?

1. Standard gradient-based optimization algorithms apply, e.g., conjugate gradient.
2. Preconditioning should help, especially at later iterations.
3. Check the step length (for the model update) by evaluating the misfit function for a few representative events.



Question: What if my model update looks trivial?



A simple perturbation is still a valuable scientific contribution, because it improves the **absolute** wavespeed values in the model.

Question: When do I stop iterating?

1. When the misfit values start oscillating between iterations.

2. When normalized data misfit is <=1 (per event? per station? overall?).

3. When there is the possibility that unaccounted-for parameters may be responsible for the remaining misfit (inaccuracies of internal surfaces, source parameters, attenuation, anisotropy, etc).

After you stop iterating…

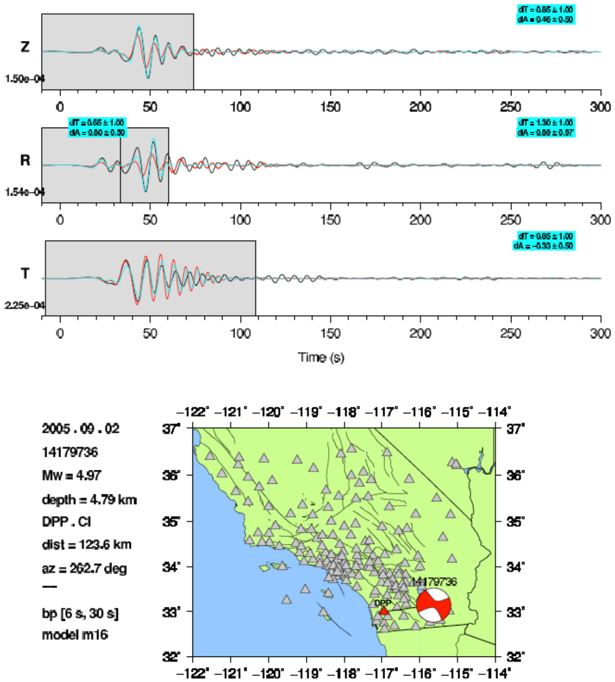
1. Misfit assessment (as many different measures as you want).
2. Compute the volumetric sensitivity of the data set (e.g., for Vs and Vb).
3. Compute individual sensitivity kernels for any seismic waveforms.
4. Time reversal imaging to identify new reflectors (Stich et al.).
5. Physical interpretation (geology, tectonics, dynamics).
6. Email any station problems to the network operators!
7. Consider reinverting with different model variables (anisotropy, attenuation, topography of internal surfaces).

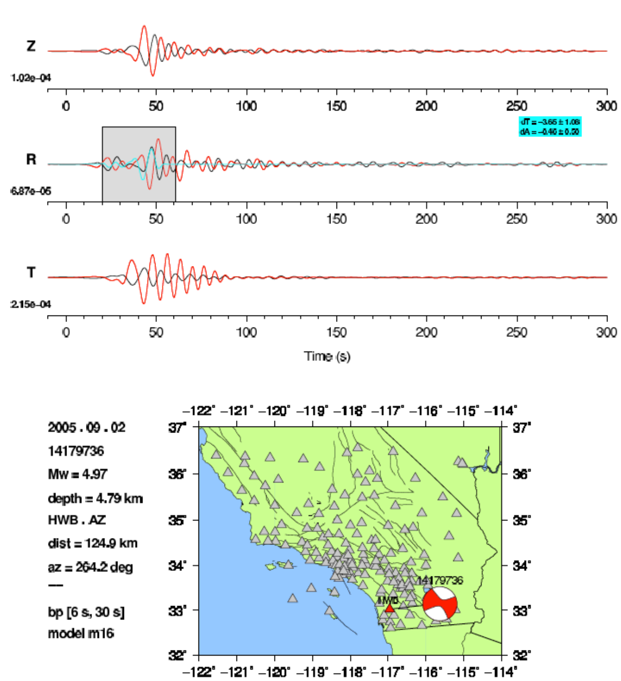
Synthetic seismograms can help improve seismic networks

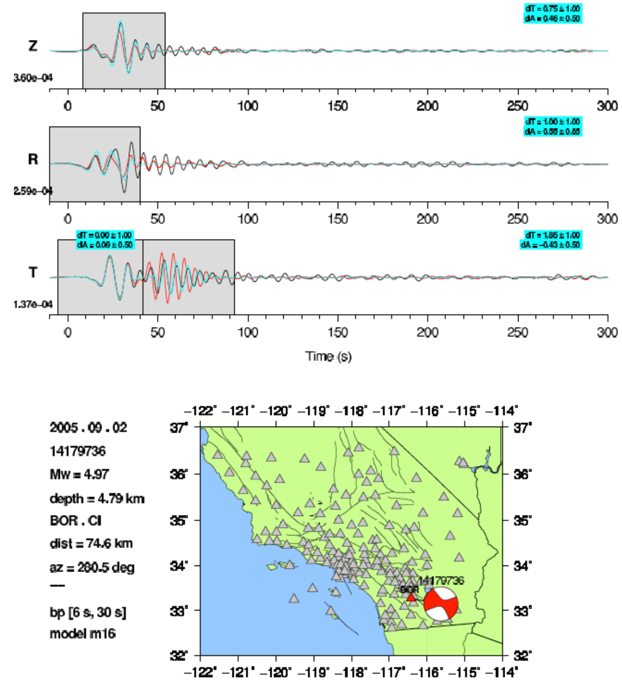
Table E.1: Southern California station–epochs with problematic polarity. "Earthquake dates" indicates the earliest and latest earthquakes within my dataset that exhibit the identified polarity problem on records bandpassed 6–30 s. These dates were used to identify the problematic epochs for each station.

Station	Earthquake dates		Corresponding Epochs		Channels	Figures
	Earliest	Latest	Start	End	(BH_)	
CRP.CI	2003–12–25	2006–06–30	2003.297	2003.301	Z, E, N	E.1
			2003.301	2006.114		
			2006.114	2006.212		
HWB.AZ	2003–05–24	2008–07–29	2003.099	2004.056	Z, E, N	E.3–E.4
			2004.056	99999		
BVDA2.AZ	2003–05–24	2007–02–09	2003.133	2004.056	Z, E, N	E.5–E.6
			2004.056	99999		
PER.CI	2003–12–04	2009–01–31	2003.141	2003.147	E, N	E.7–E.9
			2003.147	2006.157		
			2006.157	2008.305		
			2008.305	99999		
BTP.CI	2002–10–29	2003–03–11	2002.297	2003.071	E, N	E.10–E.12
NSS2.CI	2004–09–29	2005–09–02	2004.077	2006.125	E, N	E.13–E.15
109C.TA	2004–07–14	2005–10–18	2004.125	2005.101	E, N	E.16–E.18
			2005.101	2007.242		
OSI.CI	1998-01-05	1998-10-27	1995.179(?)	2002.196(?)	E(?)	E.19–E.22

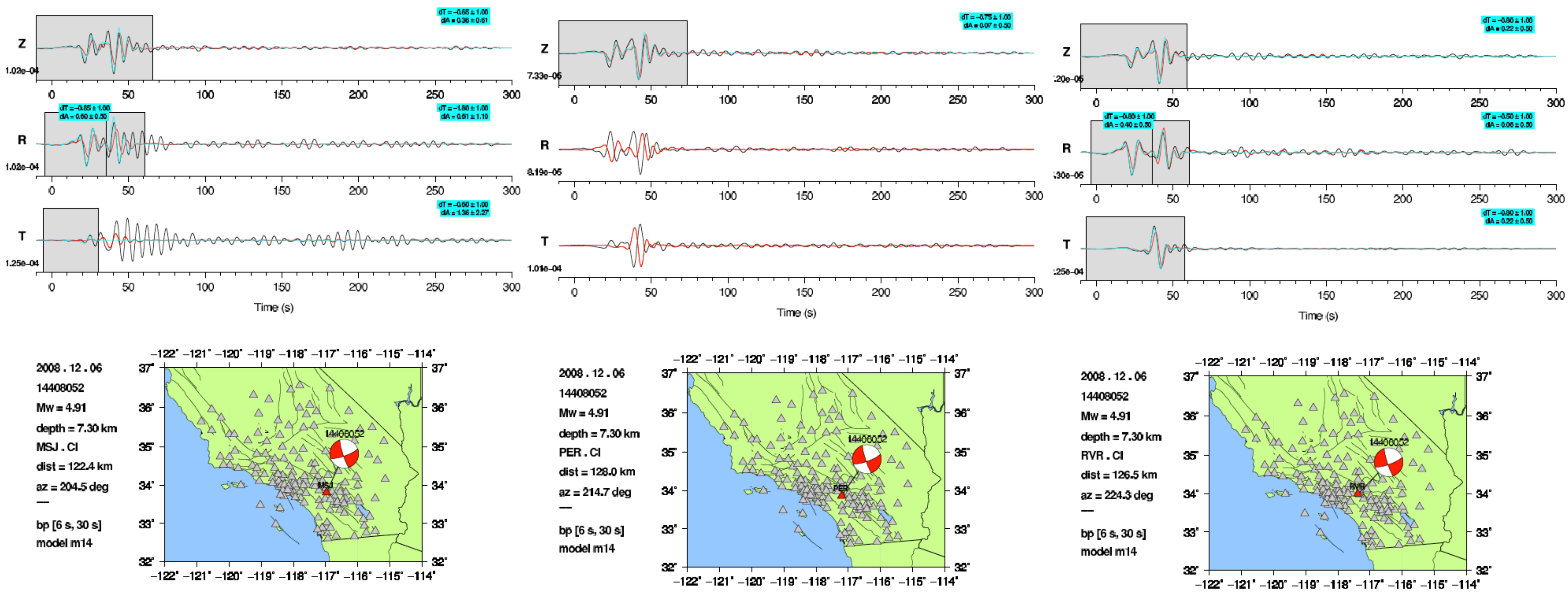
Synthetic seismograms can help improve seismic networks



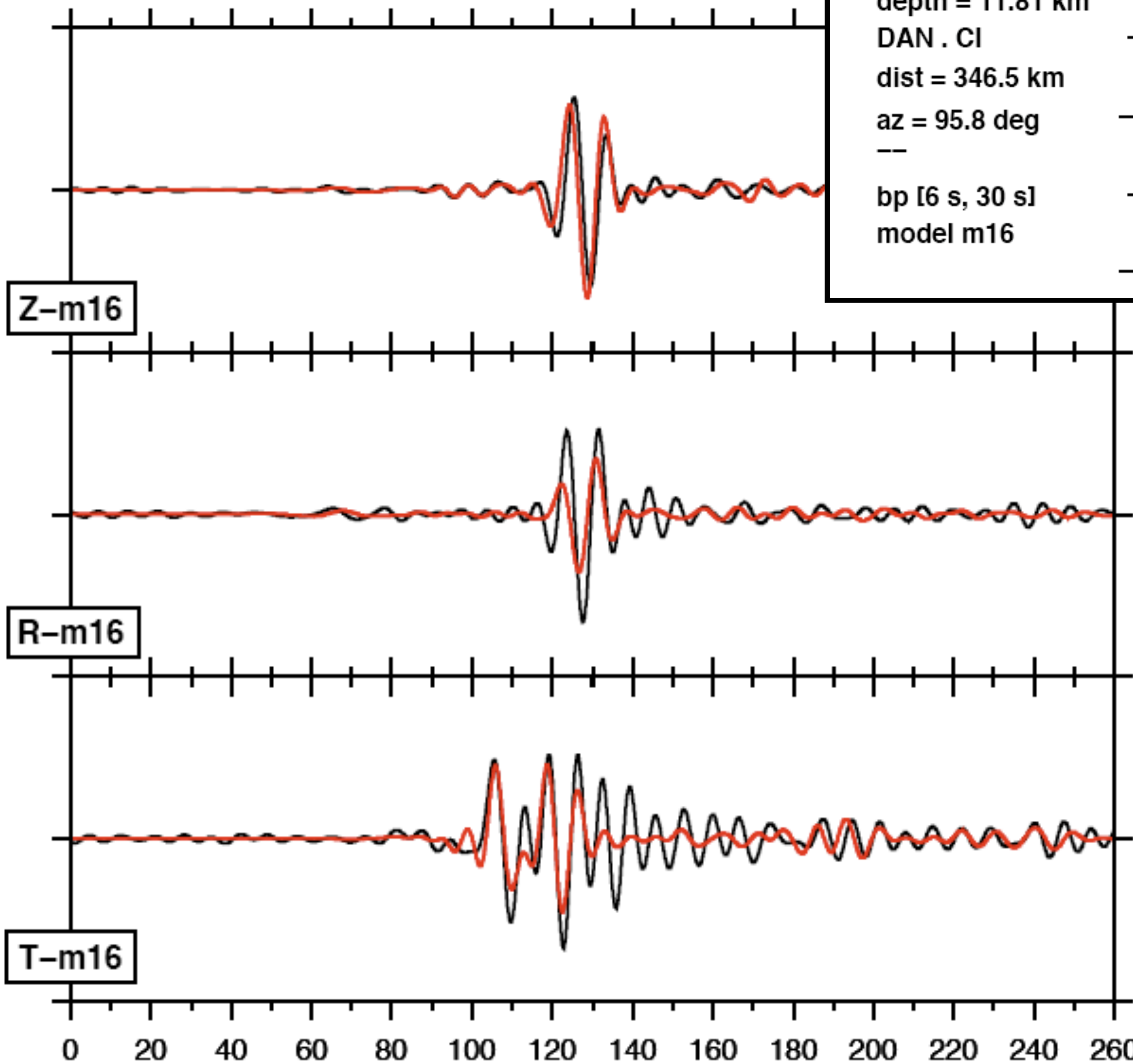


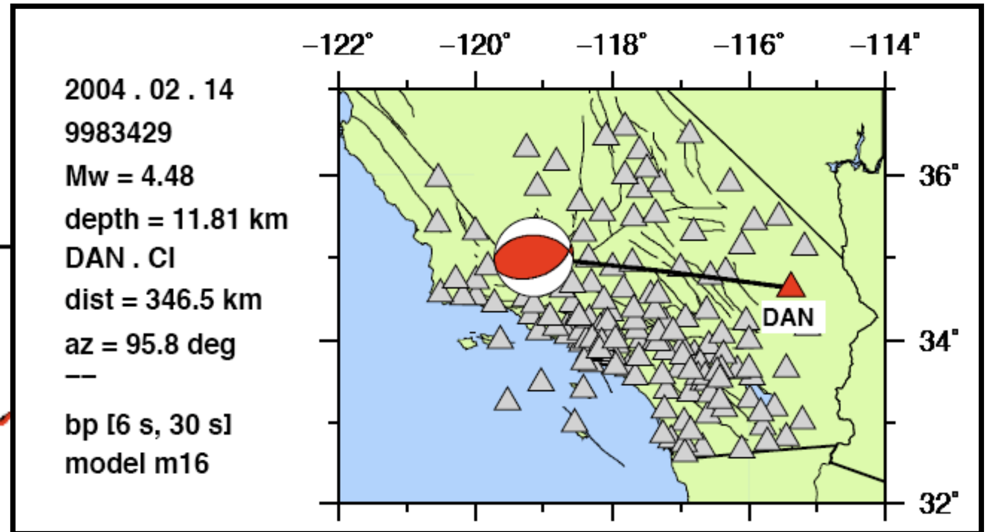


Synthetic seismograms can help improve seismic networks

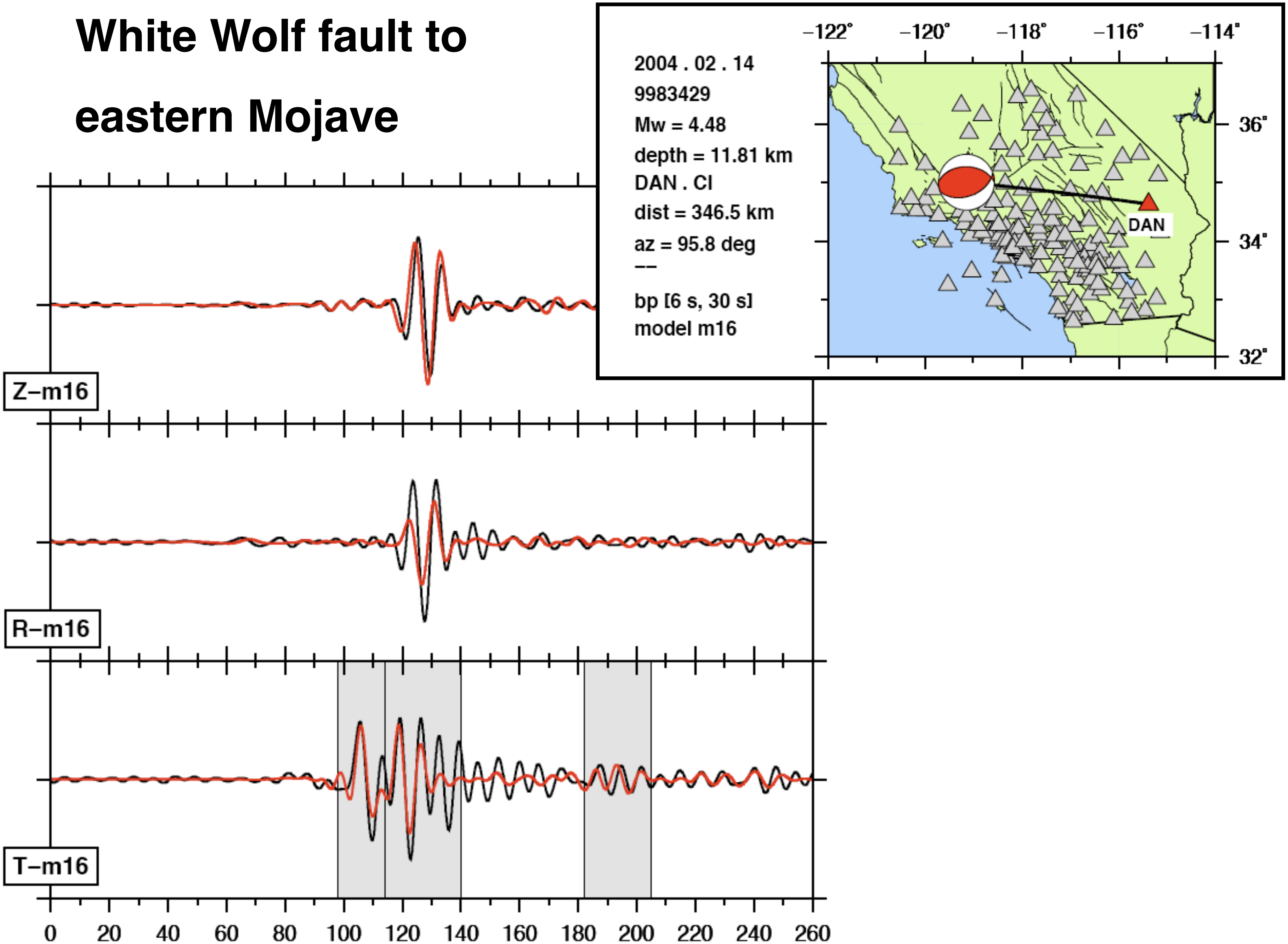


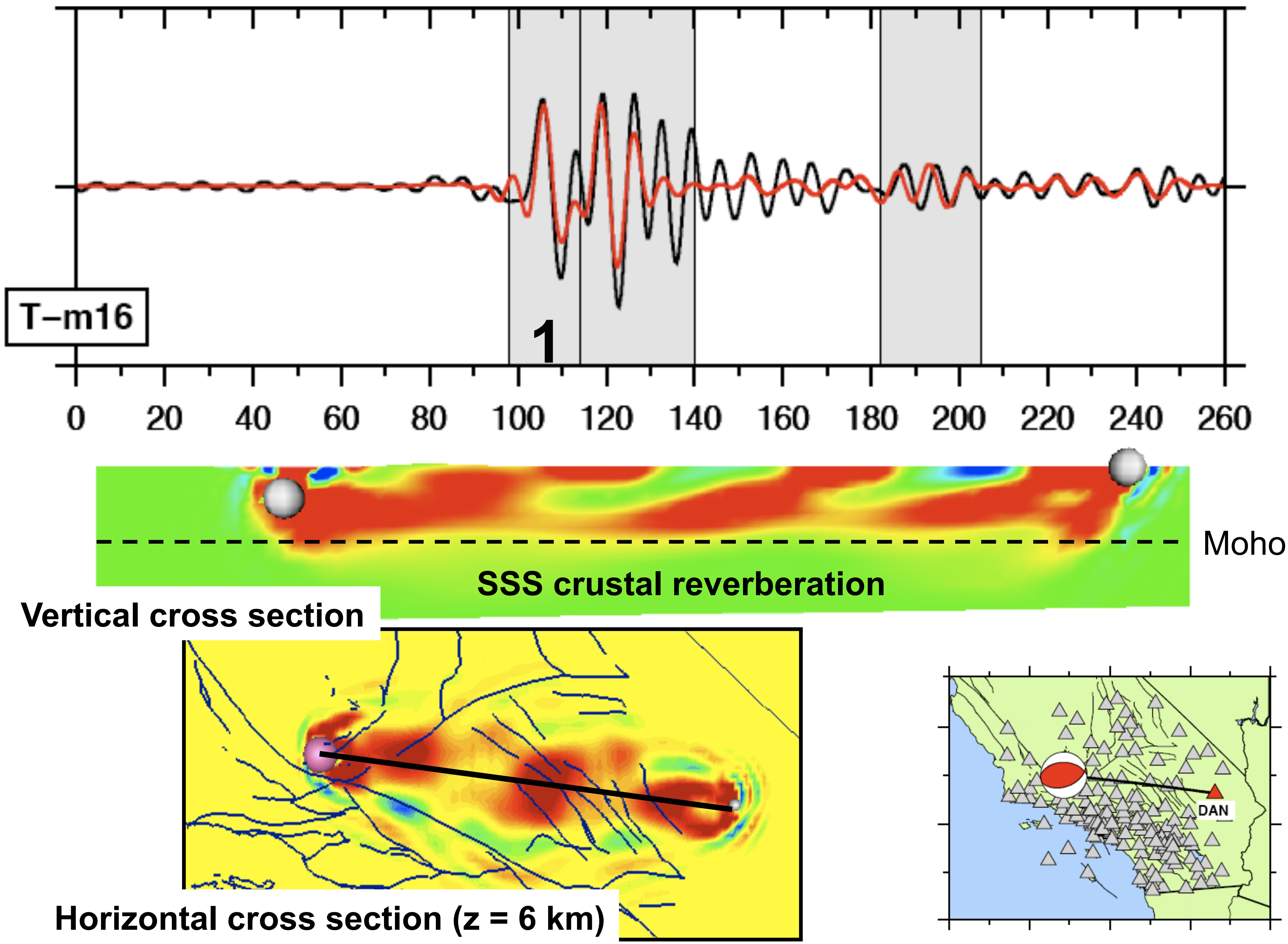
White Wolf fault to eastern Mojave

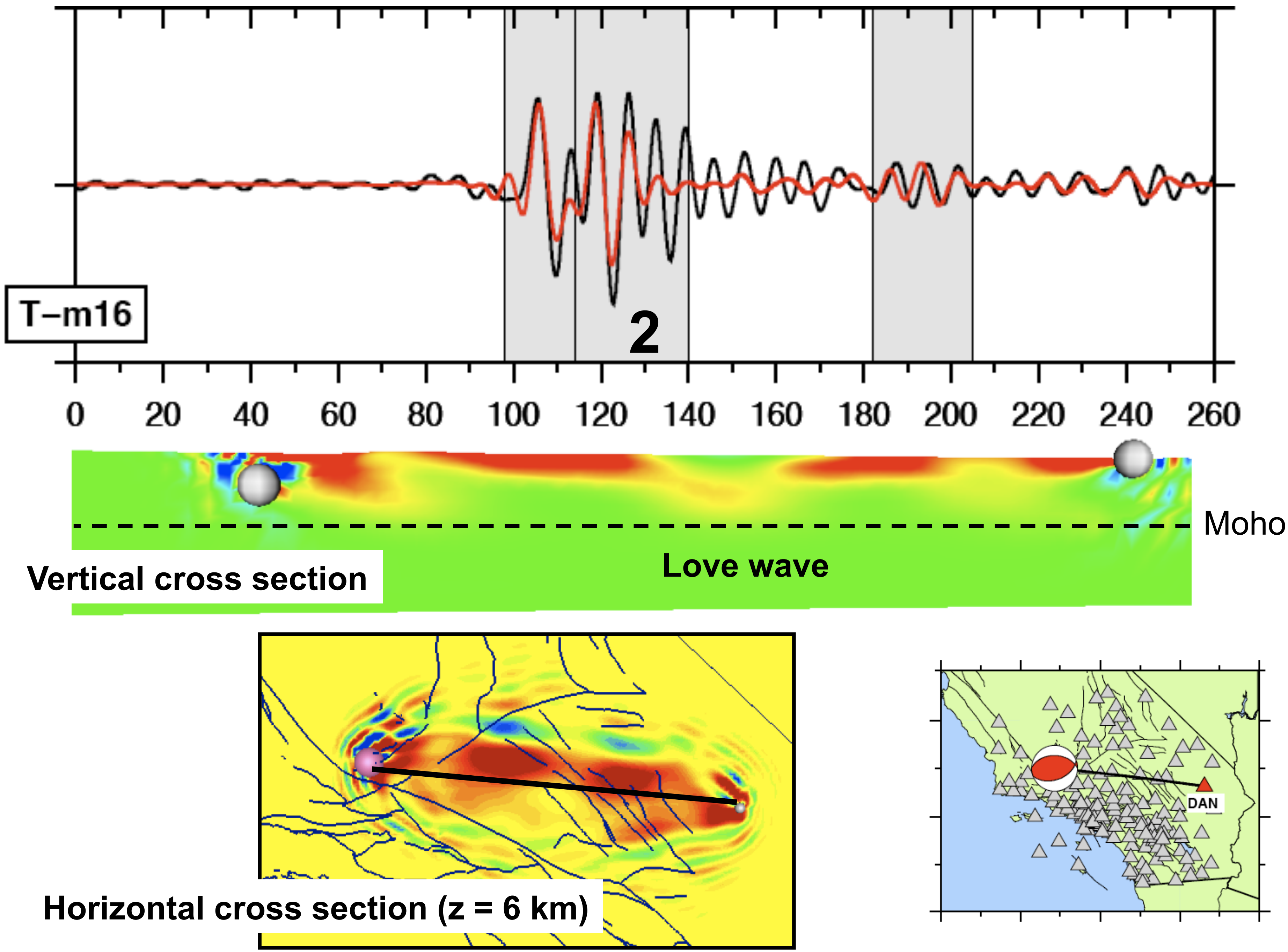


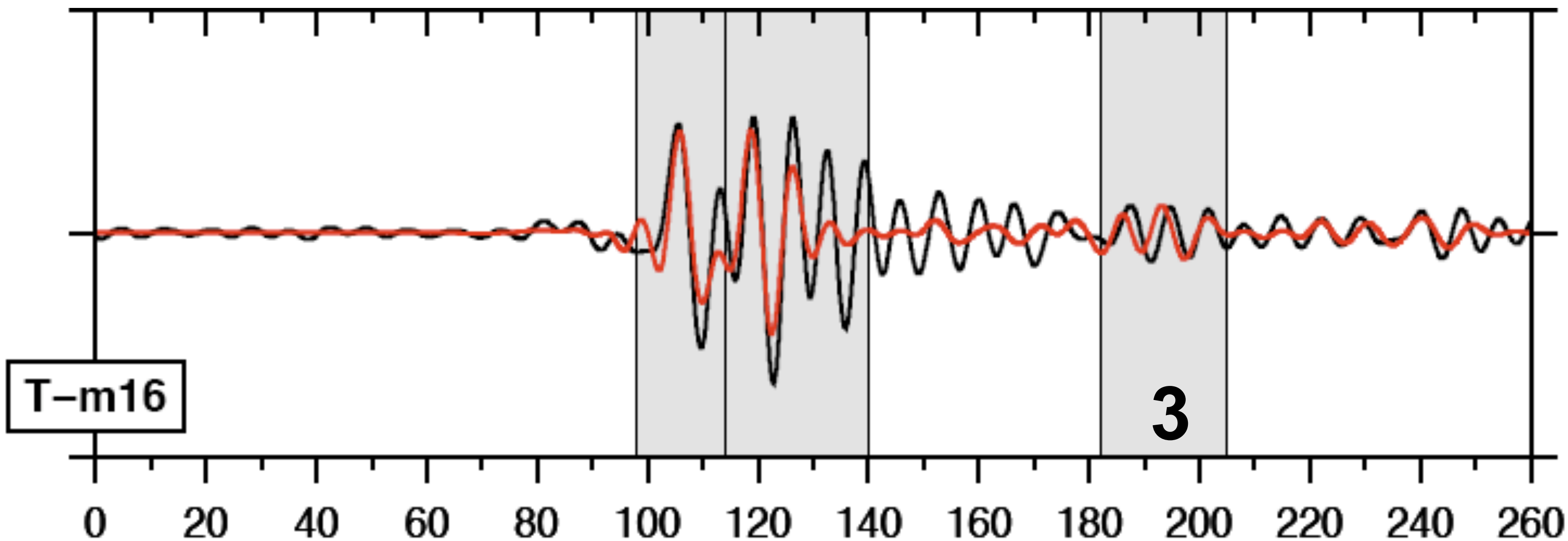


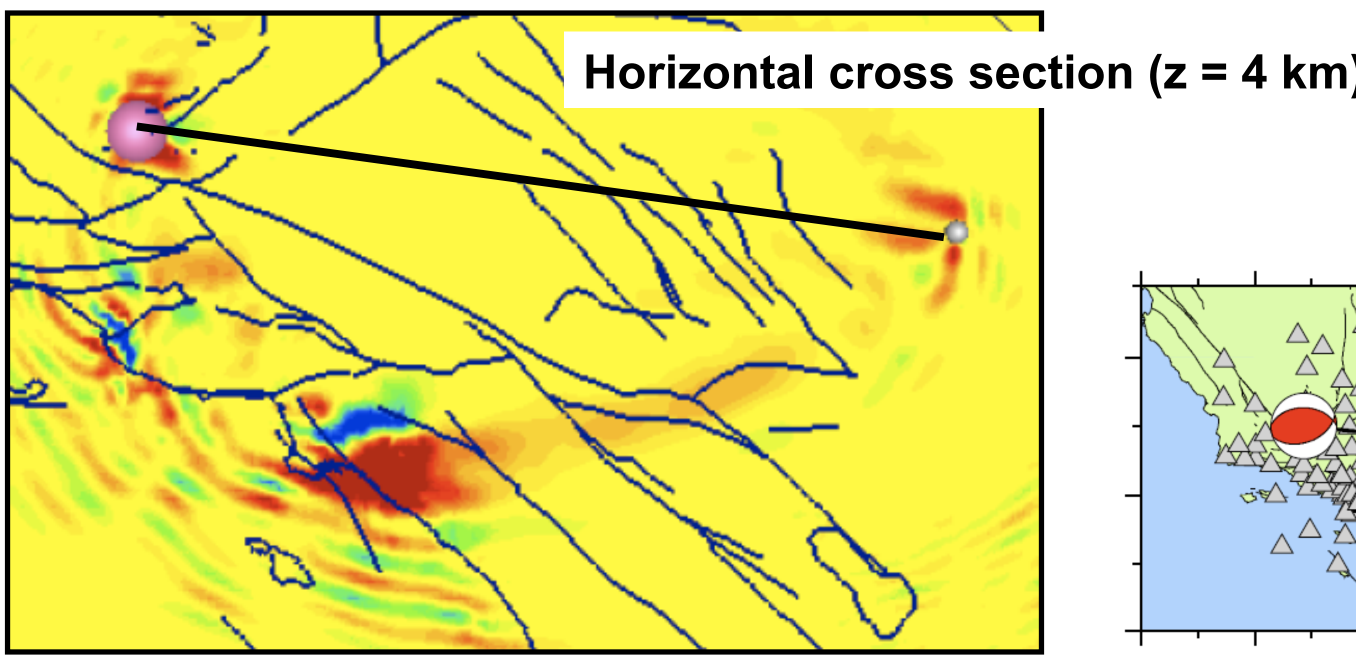
White Wolf fault to eastern Mojave

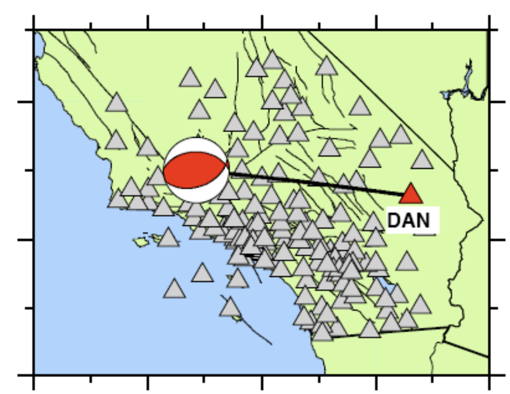



T−m16
1
0
20
40
60
80
100
120
140
160
180
200
220
240
260
Moho
SSS crustal reverberation
Vertical cross section
Horizontal cross section (z = 6 km)
DAN


T–m16
2
0
20
40
60
80
100
120
140
160
180
200
220
240
260
Moho
Love wave
Vertical cross section
Horizontal cross section (z = 6 km)
DAN

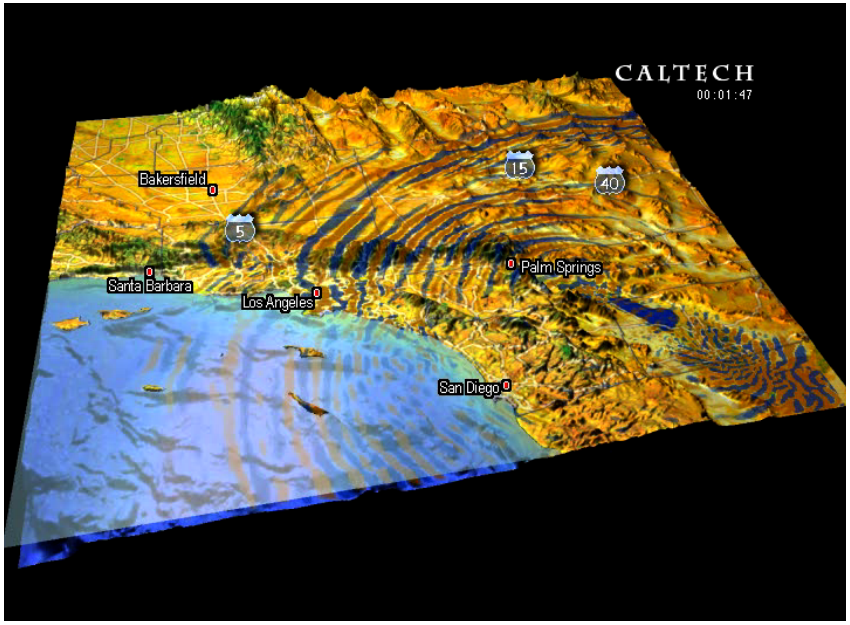

T−m16
3
0
20
40
60
80
100
120
140
160
180
200
220
240
260

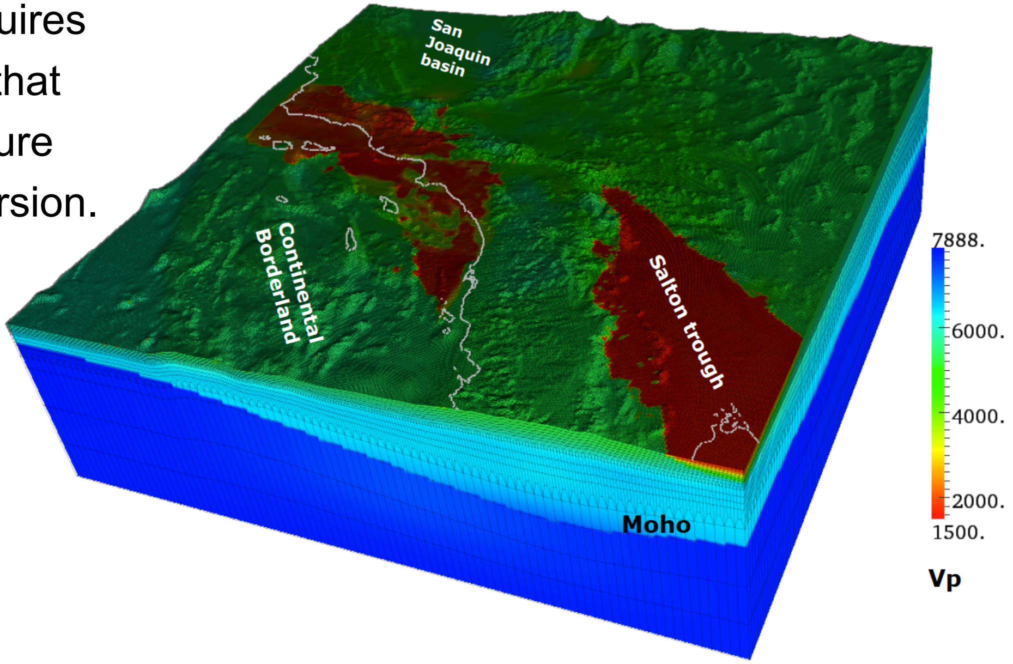

Horizontal cross section (z = 4 km)


DAN

Final remarks

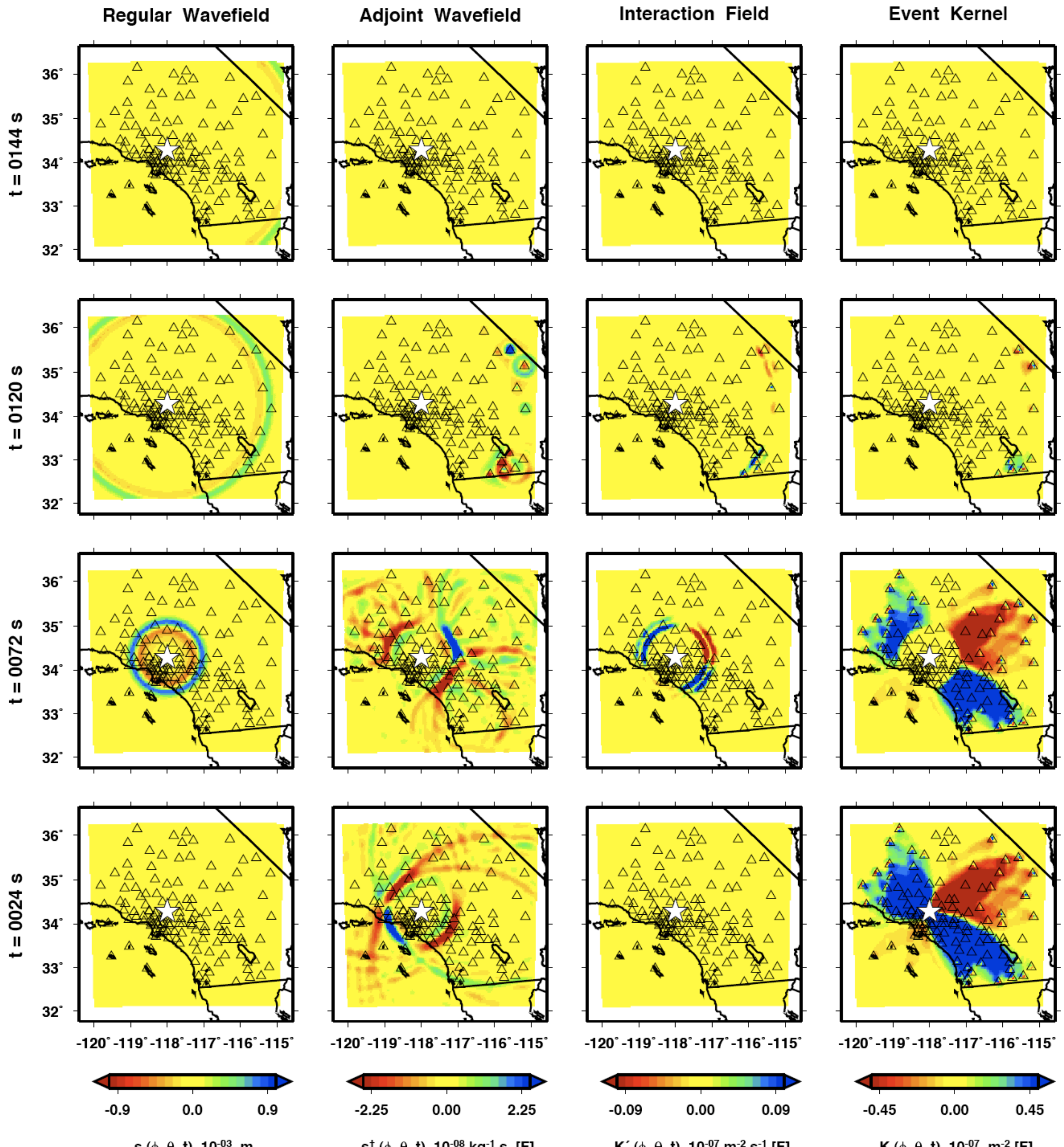
1. We have the ability to **interrogate and improve** existing 3D structural models.
2. The resolvable scale is primarily controlled by data coverage (assuming the entire waveform is used).
3. It should be possible to fit the entire recorded seismic wavefield (at a given period range) for a set of events.
4. Tomography at the crustal scale requires integration of data sets and models that are perhaps more sensitive to structure than the earthquake data in the inversion.
5. There are many opportunities for improving these techniques and applications!





Thank you!

EXTRA SLIDES

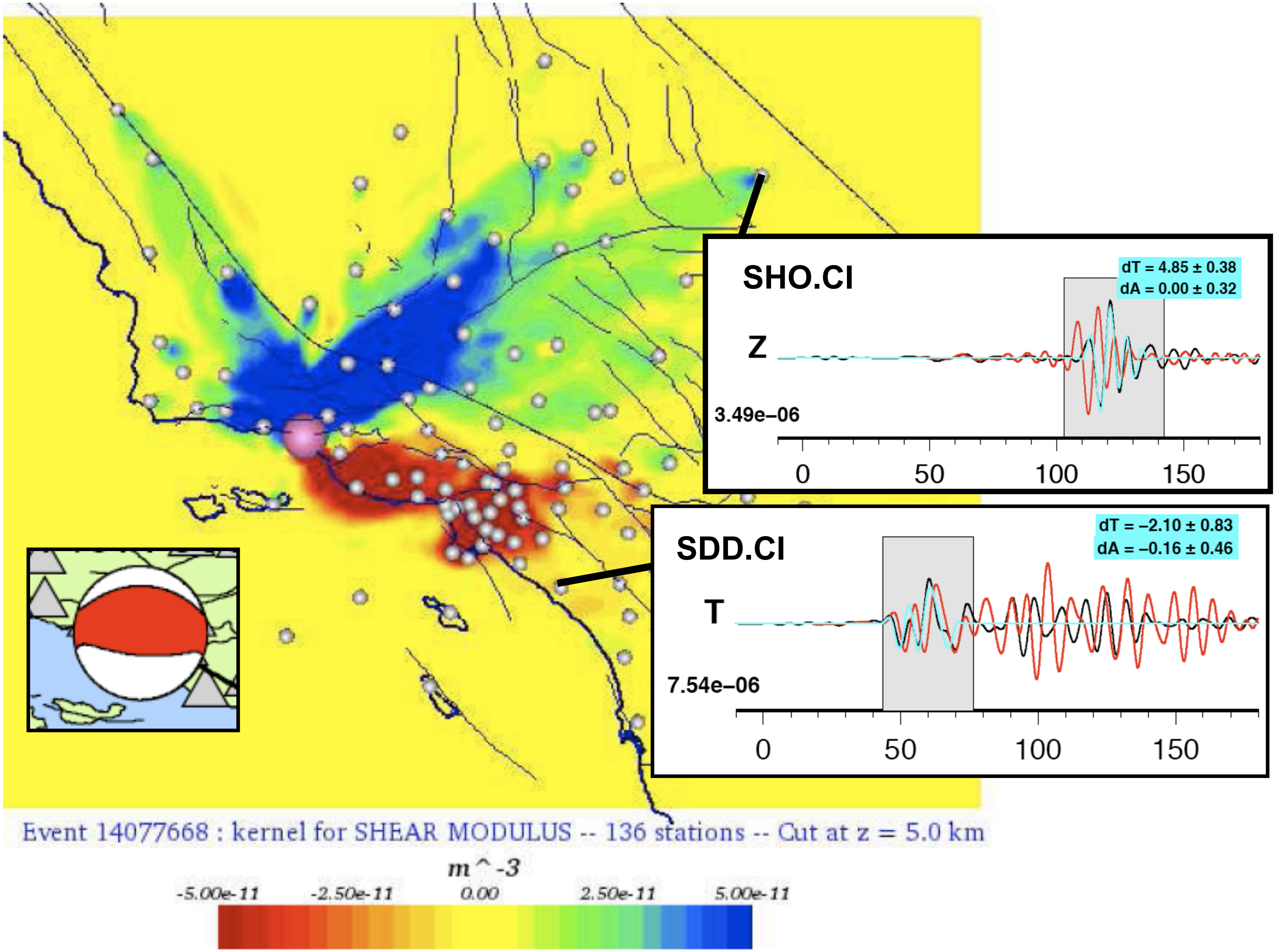


Tape, Liu, Tromp (2007)

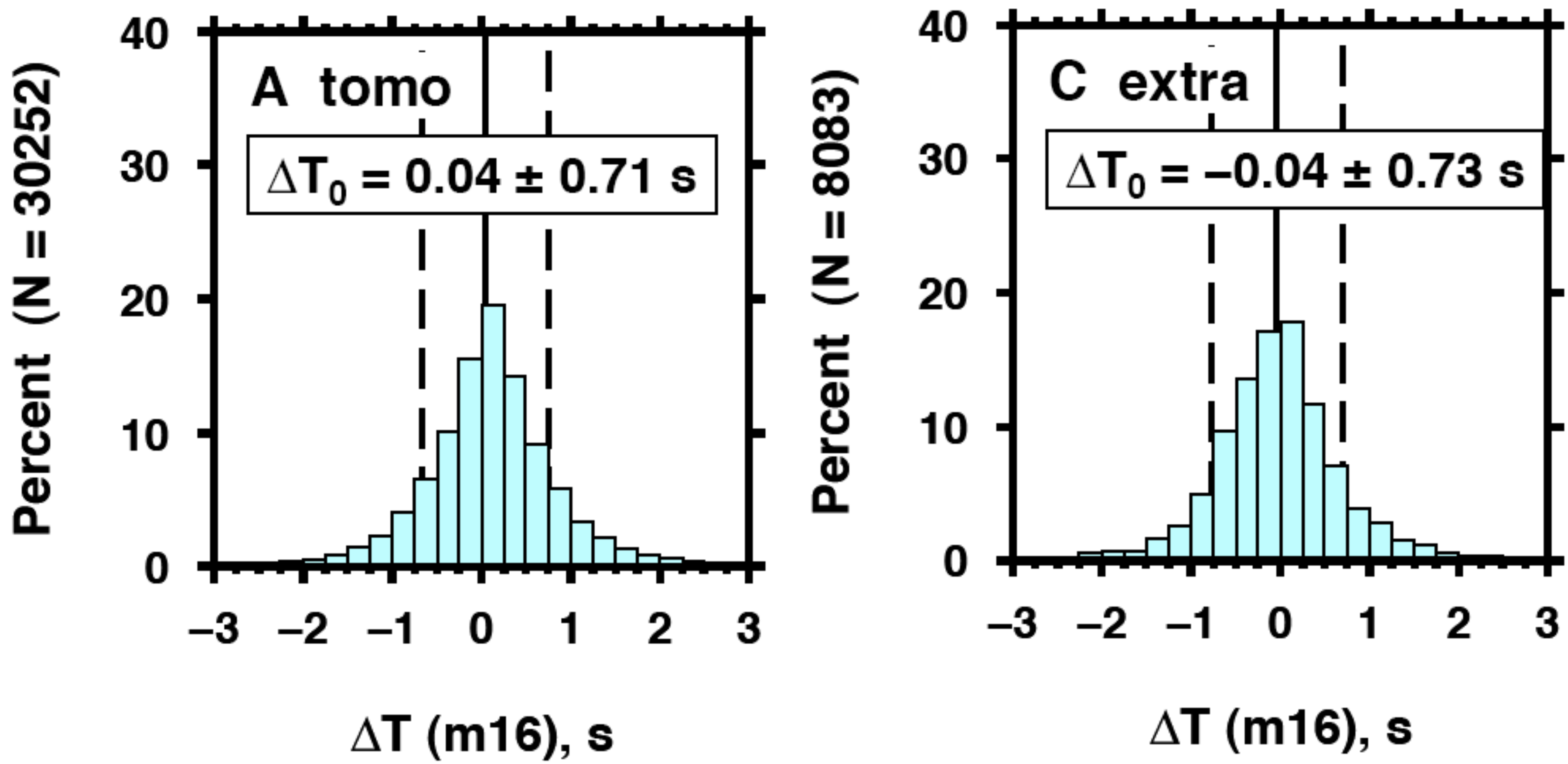
Trying to measure as much as possible

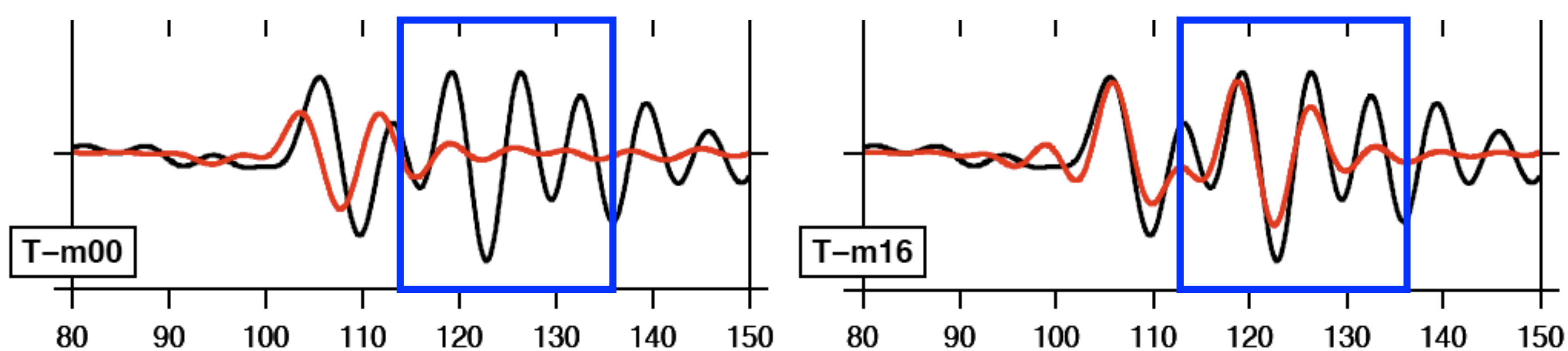
components (Z,R,T)	3	
earthquakes	143	+ 91 = 234
stations	203	
paths	12583	
seismograms	52138	
windows	61673	

	6–30 s	3–30 s	2–30 s	total
vertical (Z)	10319	5623	4864	20806
radial (R)	9276	5443	4579	19298
transverse (T)	10657	5684	5228	21569
total	30252	16750	14671	61673

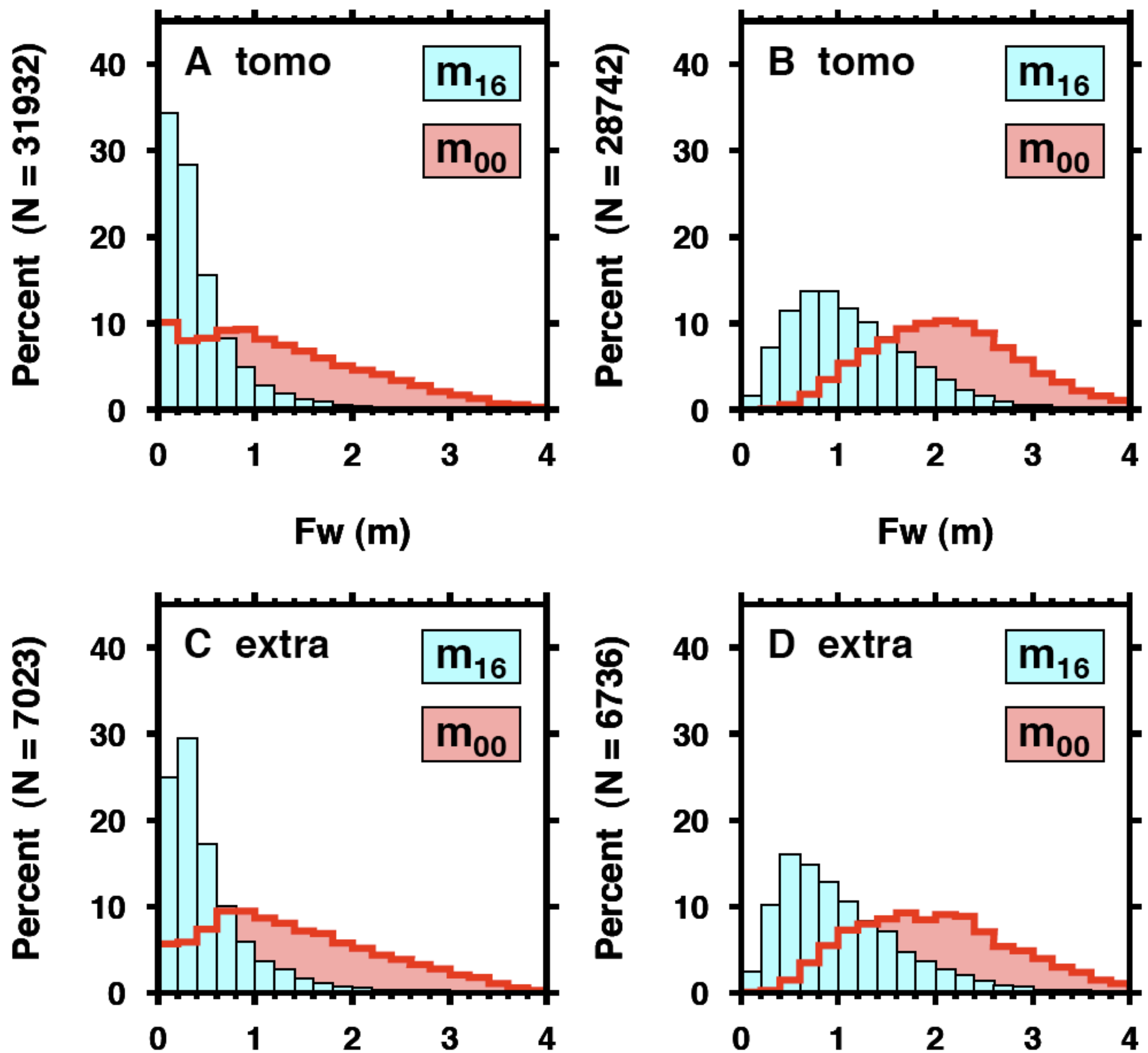


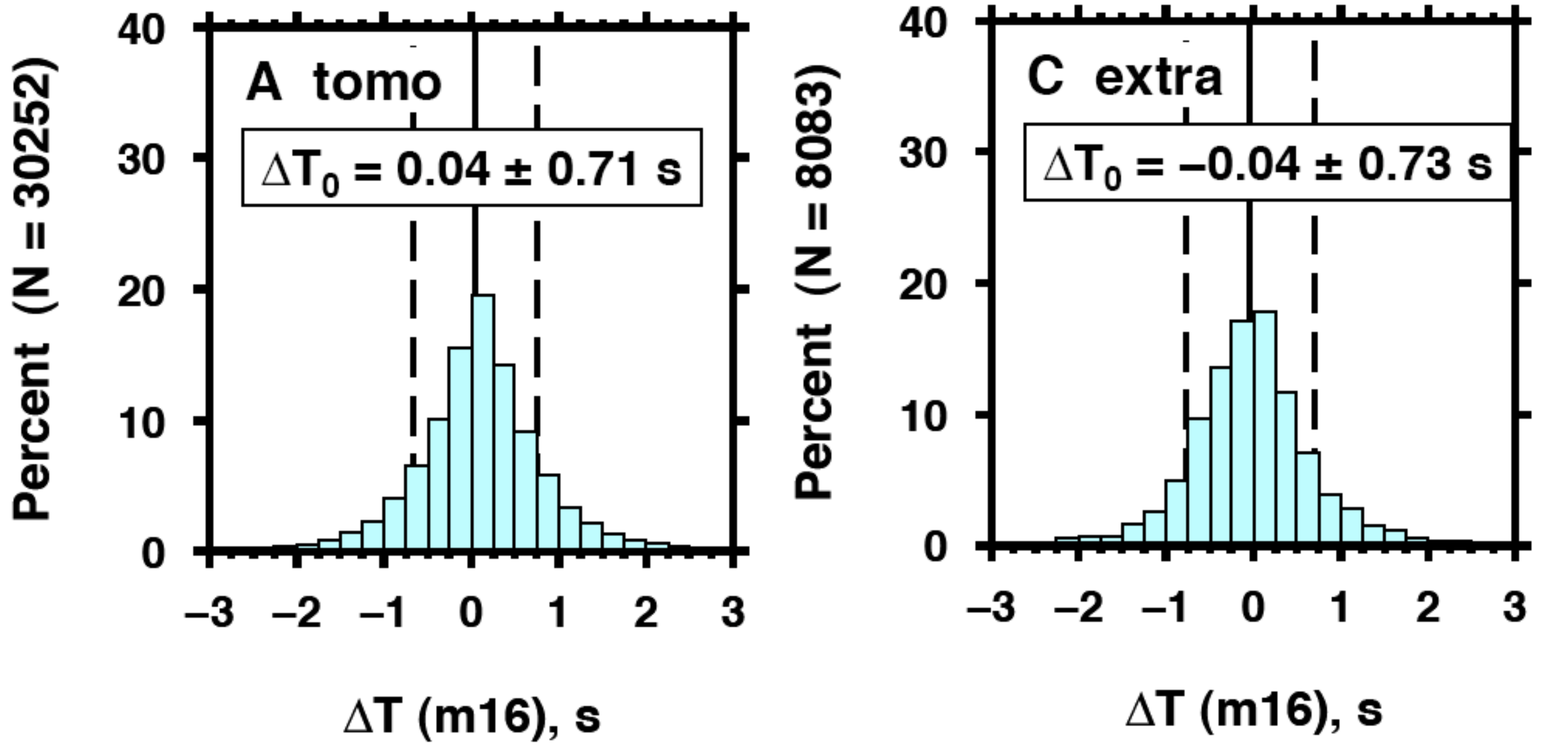
SHO.CI
dT = 4.85 ± 0.38
dA = 0.00 ± 0.32
Z
3.49e–06
0
50
100
150
SDD.CI
dT = –2.10 ± 0.83
dA = –0.16 ± 0.46
T
7.54e–06
0
50
100
150
Event 14077668 : kernel for SHEAR MODULUS -- 136 stations -- Cut at z = 5.0 km
m^-3
-5.00e-11
-2.50e-11
0.00
2.50e-11
5.00e-11

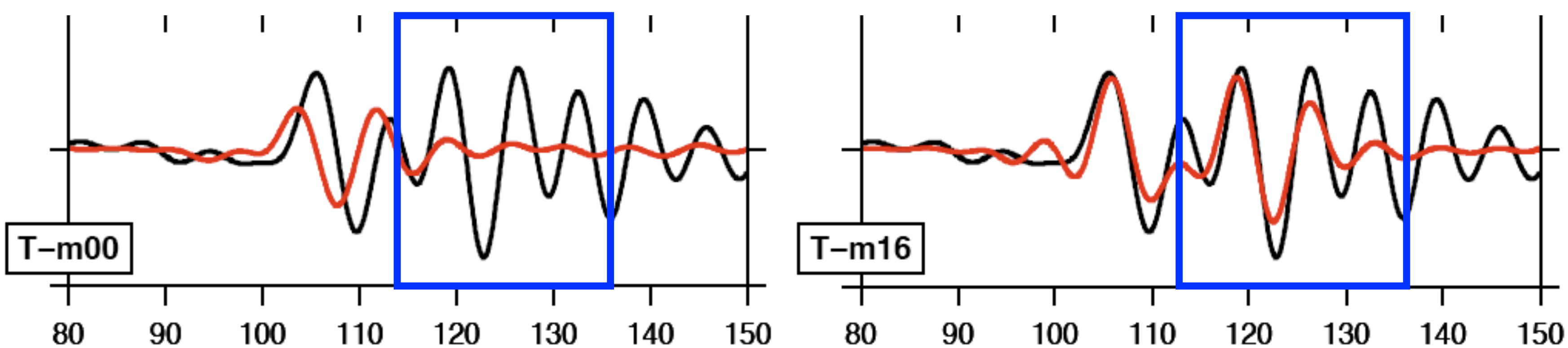

A tomo
$\Delta T_0 = 0.04 \pm 0.71$ s
Percent (N = 30252)
ΔT (m16), s
C extra
$\Delta T_0 = -0.04 \pm 0.73$ s
Percent (N = 8083)
ΔT (m16), s
0
10
20
30
40
–3
–2
–1
0
1
2
3

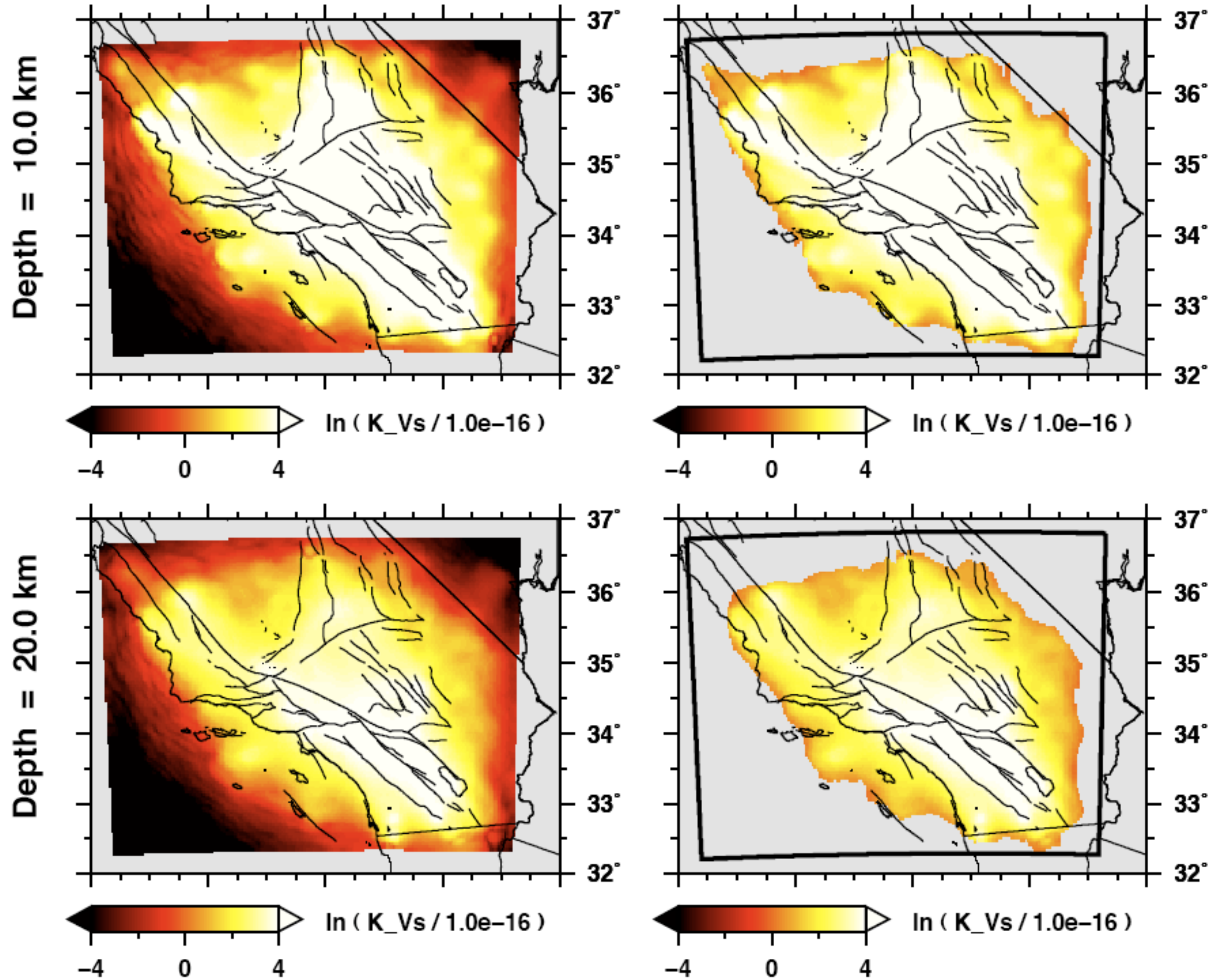

T–m00
T–m16
80
90
100
110
120
130
140
150

Waveform difference misfit reduction




A tomo
ΔT$_0$ = 0.04 ± 0.71 s
Percent (N = 30252)
ΔT (m16), s
C extra
ΔT$_0$ = –0.04 ± 0.73 s
Percent (N = 8083)
ΔT (m16), s


T–m00
T–m16
80
90
100
110
120
130
140
150

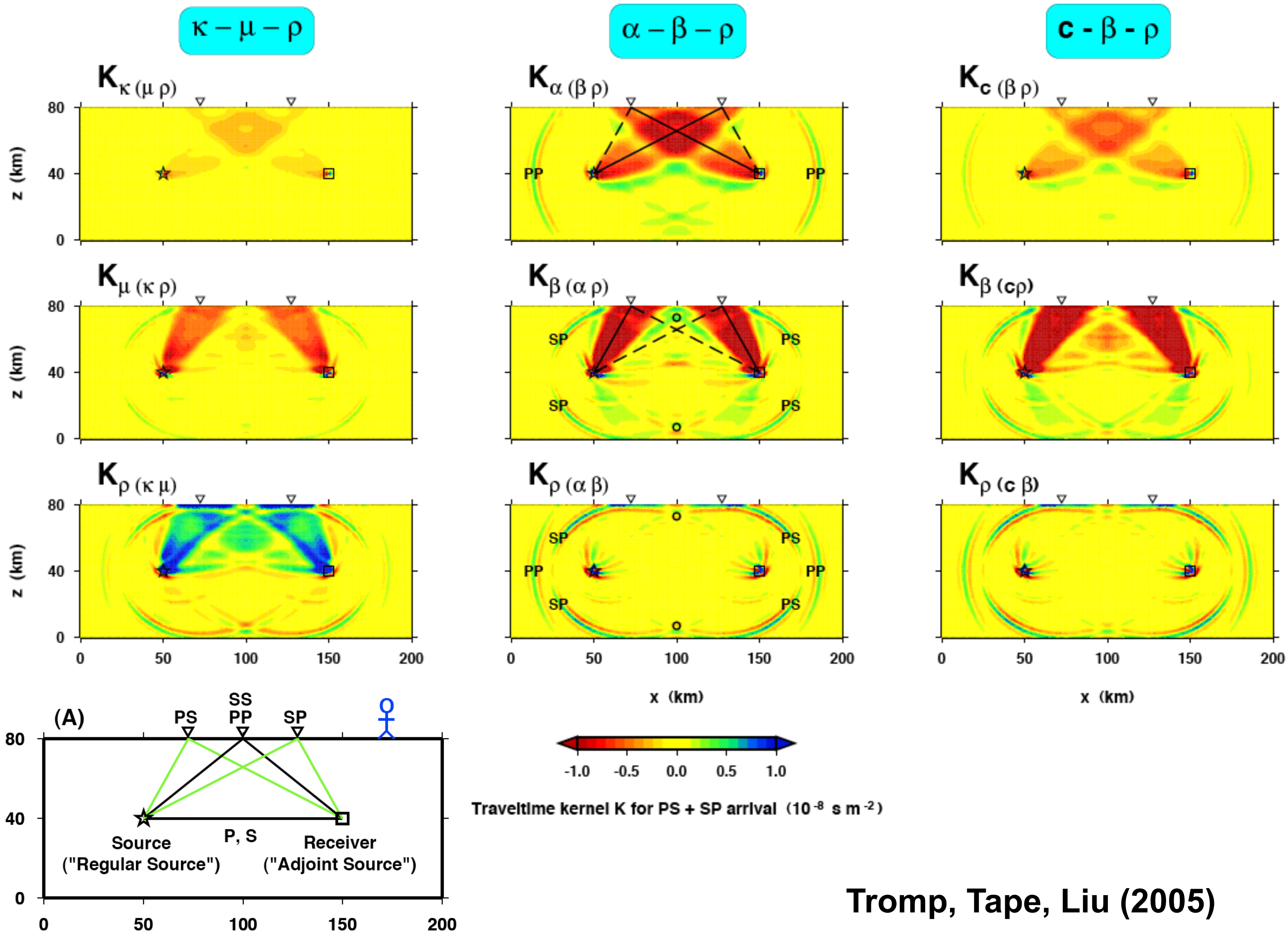


Depth = 10.0 km
Depth = 20.0 km
37˚
36˚
35˚
34˚
33˚
32˚
ln (K_Vs / 1.0e−16)
−4
0
4

Choice of model variables

Model Parameter		Notation	Kernel Expression
Bulk modulus	κ	$K_{\kappa(\mu\rho)}(\mathbf{x})$	$-\kappa \int_0^T [\nabla \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)][\nabla \cdot \mathbf{s}(\mathbf{x}, t)]\, dt$
Shear modulus	μ	$K_{\mu(\kappa\rho)}(\mathbf{x})$	$-2\mu \int_0^T \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t)\, dt$
Density	ρ	$K_{\rho(\kappa\mu)}(\mathbf{x})$	$-\rho \int_0^T \mathbf{s}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t)\, dt$
Bulk sound structure	c	$K_{c(\beta\rho)}(\mathbf{x})$	$2\, K_{\kappa(\mu\rho)}$
S-wave structure	β	$K_{\beta(c\rho)}(\mathbf{x})$	$2\, K_{\mu(\kappa\rho)}$
Density	ρ	$K_{\rho(c\beta)}(\mathbf{x})$	$K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}$
P-wave structure	α	$K_{\alpha(\beta\rho)}(\mathbf{x})$	$\left(2 + \frac{8\mu}{3\kappa}\right) K_{\kappa(\mu\rho)}$
S-wave structure	β	$K_{\beta(\alpha\rho)}(\mathbf{x})$	$2\, K_{\mu(\kappa\rho)} - \frac{8\mu}{3\kappa} K_{\kappa(\mu\rho)}$
Density	ρ	$K_{\rho(\alpha\beta)}(\mathbf{x})$	$K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}$

- $\mathbf{s}(\mathbf{x}, t)$ is the wavefield; $\mathbf{D}$ is the strain deviator; κ, μ, and ρ vary with $\mathbf{x}$



Tromp, Tape, Liu (2005)