Seismic Tomography of Southern California Using Adjoint Methods

Thesis by

Carl Tape

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2009
(Defended April 1, 2009)

Acknowledgments

I thank Jeroen Tromp for his supervision, collaboration, and scientific inspiration over the past five years. His expertise, ambition, and optimism provided a steady platform for undertaking the work in this thesis. Our scientific endeavors were exciting because they were uncertain. They followed a steady progression from theoretical foundations, to extensive modeling, to real applications. As a scientist, mentor, and teacher, Jeroen set a standard of excellence.

I thank my closest colleague and good friend, Qinya Liu. Qinya provided critical technical and scientific support in my initial years at Caltech. Her PhD research (e.g., *Liu and Tromp*, 2006), along with previous work by Dimitri Komatitsch (*Komatitsch et al.*, 2004), was vital for our tomographic inversion in southern California. I have learned a lot from Qinya over the years, and I hope I am beginning to return the favor.

I have appreciated interactions with past and present members of our research group. In particular, collaborations with Alessia Maggi and Min Chen have strengthened my thesis. I also thank Viola Carter and Donna Mireles, honorary group members, for handling many logistics over the past five years, and for providing a fun atmosphere in the Seismo Lab. I am grateful to John Woodhouse for supervising my Master's thesis at Oxford, and I benefitted from collaborations with John, Daniel Peter, and Lapo Boschi.

I benefited greatly from academic endeavors at Caltech that are not within this thesis. I thank Mark Simons and Pablo Musé for collaborating on a project involving the estimation of GPS velocity fields using spherical wavelets (*Tape et al.*, 2009). Mike Gurnis, Hiroo Kanamori, and Mark Simons have enthusiastically supported my pursuit of subduction zone parameters. In this endeavor, I hope to better understand what controls the maximum-moment earthquake and the dip of slabs in subduction zones (*Ruff and Kanamori*, 1980; *Jarrard*, 1986). I look forward to continuing both of these collaborations in the future.

I am grateful to the scientists at Caltech who helped turn me into a geophysicist: Jean-Philippe Avouac, Rob Clayton, Mike Gurnis, Egill Hauksson, Don Helmberger, Hiroo Kanamori, Nadia Lapusta, Tapio Schneider, Mark Simons, and Joann Stock. I have learned much from them through classes, seminars, and personal interactions. Two classes by Hiroo Kanamori set a standard of teaching excellence I will aim to emulate.

I am also grateful for scientists who visited Caltech, in particular, Albert Tarantola and Malcolm Sambridge. Tarantola changed the way I view geophysics, science, and life around me. His scientific approach is idealistic, ambitious, and feasible. Sambridge provided key insights for our tomographic problem, and these are incorporated in Chapter 4. His class on inverse methods, his contributions in group meeting, and his outstanding performance on the soccer field made his visit memorable in many ways.

Two components were essential for conducting our tomographic study of southern California: data and computers. Without the data, there is no inverse problem. I thank the Southern California Data Center (Egill Hauksson, Ellen Yu), the Northern California Data Center (Doug Neuhauser), the IRIS data center (Tim Ahern), and the University of Nevada Reno (Glenn Biasi) for providing seismograms used in our study. Without the computing power, our technique is not feasible. I thank Jeroen Tromp and Mike Gurnis for having the vision of computational infrastructure within the Division of Geological and Planetary Sciences. I used 0.8 million CPU-hours in the tomographic inversion, and I recognize the luxury of having access to such a computational facility. I thank Naveed Near-Ansari and John Lilley for keeping the cluster up and running, and for hundreds of replies to my emails. I thank the code management facility of the Computational Infrastructure for Geodynamics (CIG), in particular Wei Mi, Sue Kientz, and Leif Strand. I also thank the phenomenal IT support in the GPS Division at Caltech: Mike Black, Scott Dungan, and Ken Ou.

Regarding southern California earthquakes and structure, I received scientific feedback or unpublished data from the following people: Jeanne Hardebeck, Guoqing Lin, Egill Hauksson, Don Helmberger, Ying Tan, Shengji Wei, Rob Clayton, Hiroo Kanamori, Tom Brocher, Christiane Stidham, John Shaw, Andreas Plesch, Ying Tan, Jeff McGuire, Gary Fuis, Janice Murphy, Nik Christensen, and Paul Friberg. I benefitted from discussions of geology or geodynamics with Mark Legg, Jason Saleeby, Steve Kidder, Alan Chapman, Elisabeth Nadin, Mihai Ducea, Peter Luffi, Laetitia Le Pourhiet, and Leon Silver. Regarding structure and simulations in the Bay Area region, I benefitted from discussions with Artie

Rodgers and Brad Aagaard.

I participated in several field trips that left me with indelible memories and helped shape my geological background. I joined Joe Kirshvink on five outstanding Bob-Sharp-style Ge136 field trips in the southwestern United States. Jean-Philippe Avouac led a phenomenal active tectonics trip to Tian Shan in far western China. Jason Saleeby and Paul Asimow led a memorable trip to Hawaii, where we even observed graded sedimentary bedding in the form of peridotite xenoliths within a relict lava river channel. And Joann Stock supported an epic (at least in the life of a graduate student) geophysics research cruise from Antarctica to New Zealand.

I thank my thesis committee: Rob Clayton, Mark Simons, Don Helmberger, Egill Hauksson, and Jeroen Tromp.

I am grateful to all of my Caltech acquaintances, officemates, and friends, especially Willy Amidon (Susan Parsons) and Yoshi Kaneko (Misaki Miyazawa). My positive experience at Caltech is partly attributable to Yoshi.

I thank my family: my brother Ken, my mom Jeanne, and my dad Walt.

And I thank my incredible, wonderful wife, Elisabeth Nadin.

Abstract

A fundamental objective of seismology is to produce detailed tomographic images of Earth's interior by fitting simulated seismograms to recorded seismograms. The quality of the image depends on the quality of the observations and on the accuracy of the modeling tool. We present a seismic tomography approach that employs accurate numerical methods of seismic wave propagation. Our approach follows successive steps of a minimization problem. First, specify an initial tomographic model in terms of shear wave speed, compressional wave speed, and density. Next, collect a dataset of well-recorded earthquakes. Specify a misfit function that quantifies the difference between sets of recorded and simulated seismograms. For each earthquake, evaluate both the misfit function and the gradient of the misfit function. Adjoint methods are used to compute the gradient via the interaction of a "forward wavefield," propagating from source to stations, with an "adjoint wavefield," propagating from stations to source. Using the gradient for each earthquake, we then compute an update to the initial model. This procedure is then iterated to obtain a better model of Earth's interior structure.

We iteratively improve a three-dimensional (3D) seismological model of the southern California crust. The resulting model is constructed from 16 tomographic iterations, which required 6800 wavefield simulations and a total of 0.8 million CPU hours. The new crustal model reveals strong heterogeneity, including local changes of $\pm 30\%$ with respect to the initial 3D model provided by the Southern California Earthquake Center. The improved crustal model illuminates features at the surface that agree with geology, such as the southern San Joaquin basin. It also reveals crustal features at depth that aid in the tectonic reconstruction of southern California, such as possible Farallon oceanic crustal fragments beneath the western Transverse Ranges. The new model enables more accurate assessments of seismic hazard for scenario earthquakes.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis involves the development and application of adjoint methods to the seismic tomographic inverse problem. The success of an inverse problem depends primarily on the quality and coverage of the data, and on the accuracy of the forward modeling tool within the inverse problem. Our forward modeling tool is the spectral-element method (SEM), which has been developed for regional and global scales of seismic wave propagation (*Komatitsch and Vilotte*, 1998; *Komatitsch and Tromp*, 2002a,b; *Komatitsch et al.*, 2004).

The remaining chapters emphasize the inverse problem, so we will briefly state the essential equations of the forward problem of seismic wave propagation. The equation of motion for an anelastic Earth is given by

$$\rho \frac{\partial^2 \mathbf{s}}{\partial t^2} = \boldsymbol{\nabla} \cdot \mathbf{T} + \mathbf{f}, \tag{1.1}$$

where $\rho(\mathbf{x})$ denotes the density distribution, $\mathbf{s}(\mathbf{x}, t)$ the seismic wavefield, $\mathbf{T}(\mathbf{x}, t)$ the stress tensor, and $\mathbf{f}(\mathbf{x}, t)$ the earthquake source. (We neglect rotation and self-gravitation.)

If the medium is elastic, then we apply Hooke's law

$$\mathbf{T} = \mathbf{c} : \boldsymbol{\nabla}\mathbf{s} \, , \tag{1.2}$$

which states that the stress is linearly related to the displacement gradient $\boldsymbol{\nabla}\mathbf{s}$ (strain) through the fourth-order elastic tensor $\mathbf{c}(\mathbf{x})$.

If the elastic medium is isotropic, then the elements of $\mathbf{c}$ are described by two parameters:

$$c_{jklm} = (\kappa - \tfrac{2}{3}\mu)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl}), \tag{1.3}$$

where $\mu(\mathbf{x})$ is the shear modulus and $\kappa(\mathbf{x})$ is the bulk modulus.

In general, the `SPECFEM3D` software (*Komatitsch and Tromp*, 2002a; *Komatitsch et al.*, 2004) takes into account the full complexity of seismic wave propagation, including attenuation, full anisotropy, topography, and ocean loads, as well as asymmetries relevant for very long-period waves, such as Earth's ellipticity and self-gravitation. In this thesis, the SEM (2D and 3D version) has been used primarily for elastic, isotropic Earth models. Attenuation is only implemented within the sedimentary basins in the Los Angeles region for 3D wavefield simulations.

1.1 The inverse problem (and thesis overview)

The forward modeling tool within our tomographic inverse problem is the SEM, either in a 2D wave propagation code (*Tromp et al.*, 2005; *Tape et al.*, 2007) or a 3D version (*Komatitsch et al.*, 2004). The challenge for tomographers is how to harness the accuracy of these forward-modeling methods for the inverse problem. One approach is to utilize so-called adjoint methods (*Tarantola*, 1984; *Talagrand and Courtier*, 1987; *Courtier and Talagrand*, 1987), which are related to concepts developed in seismic imaging (*Claerbout*, 1971; *McMechan*, 1982). *Tromp et al.* (2005) demonstrated the theoretical connections between adjoint methods, seismic tomography, time reversal imaging (e.g., *Fink*, 1997), and finite-frequency "banana-doughnut" kernels (e.g., *Dahlen et al.*, 2000).

Chapter 2 (*Tape et al.*, 2007) extends the study of *Tromp et al.* (2005) in the direction of an iterative inverse problem based on adjoint methods. The 2D synthetic inversion experiments are performed using three different approaches: (1) a gradient-based fully numerical approach using adjoint methods — we call this "adjoint tomography"; (2) a classical approach using finite-frequency kernels; and (3) a classical approach using rays. By "classical," we mean that a model is expanded into basis functions, and the sensitivity of each measurement is described using rays or kernels derived from a simple (homogeneous or 1D) reference model (e.g., Table A.1).

Chapter 3 contains excerpts from *Tromp et al.* (2005) that demonstrate the ability of

the adjoint approach to isolate the volumetric region that a seismic waveform "sees" as it propagates through a (1D or 3D) reference model. We implemented adjoint methods within a 2D SEM code, and then designed and conducted series of numerical experiments. (For a quick visual understanding of how finite-frequency kernels are formed via the interaction between the forward wavefield and the adjoint wavefield, one should begin with Chapter 3.)

The frequency dependence of the seismic wavefield plays a critical role in the forward and inverse problems, and we illustrate some basic features in Figures 1.1 and 1.2. A seismic waveform containing longer-period energy will sample (or "see") a broader region along the path between the source and station (Figure 1.1). Thus, by making measurements over multiple frequency bands, we should be able to better sample the model, and to make changes that are required by the data. With the implementation of adjoint methods within a 3D SEM code (*Liu and Tromp*, 2006), it became possible to visualize finite-frequency kernels that could be complicated, even for simple 1D layers models (Figure 1.2). These adjoint capabilities within the 3D SEM code provided the possibility for a full inverse problem using actual data (Chapter 6).

Chapter 4 presents an approach to compute an optimal model update for a given iteration within the inverse problem. The approach relies on subspace projection techniques (*Kennett et al.*, 1988; *Sambridge et al.*, 1991), and it does not require any additional forward or adjoint wavefield simulations.

In preparation for a tomographic inversion that begins with an initial 3D reference model, *Maggi et al.* (2009) developed an automated algorithm for picking measurement time windows containing recorded and synthetic waveforms. Chapter 5 contains excerpts from *Maggi et al.* (2009), emphasizing examples for southern California.

Chapter 6 presents an application of adjoint tomography to southern California. A preview of this study is exemplified in Figures 1.3–1.5. Figure 1.3 demonstrates the highly variable frequency content within a three-component seismogram. It also shows the appearance of the wavefield over different period ranges.

This thesis marks the beginning of an endeavor into seismic tomography using adjoint methods. Future work will undoubtedly refine the current procedures and include additional complexity. First, we will implement — or possibly invert for — attenuation in southern California in regions of tomographically documented sedimentary basins.[1] Sec-

[1] Attenuation is already implemented in the basins of *Komatitsch et al.* (2004) and *Lovely et al.* (2006)

ond, we will consider crustal anisotropy (e.g., *Christensen and Mooney*, 1995; *Paulssen*, 2004) and boundary surfaces (e.g., *Fuis et al.*, 2003; *Yan and Clayton*, 2007; *Bleibinhaus et al.*, 2007) as inversion parameters. The prospects of inverting for these parameters are discussed in *Sieminski et al.* (2007) for anisotropy and in *Dahlen* (2005) and *Tromp et al.* (2005) for boundary surfaces.

Finally, we will apply seismic reflection imaging techniques (e.g., *Kiyashchenko et al.*, 2007) to identify and quantify lateral and vertical reflectors in southern California. Figure 1.6 shows an example of a Rayleigh-wave reflection off of the Tehachapi Mountains. The reflected waveform is not apparent in the synthetic seismogram from the initial 3D model ($\mathbf{m}_{00}$) but is in the final model ($\mathbf{m}_{16}$). The ability to capture such waveforms on individual seismograms demonstrates the possibility of enhancing tomographic coverage by delving deeper into seismograms while using the same set of earthquakes and stations.

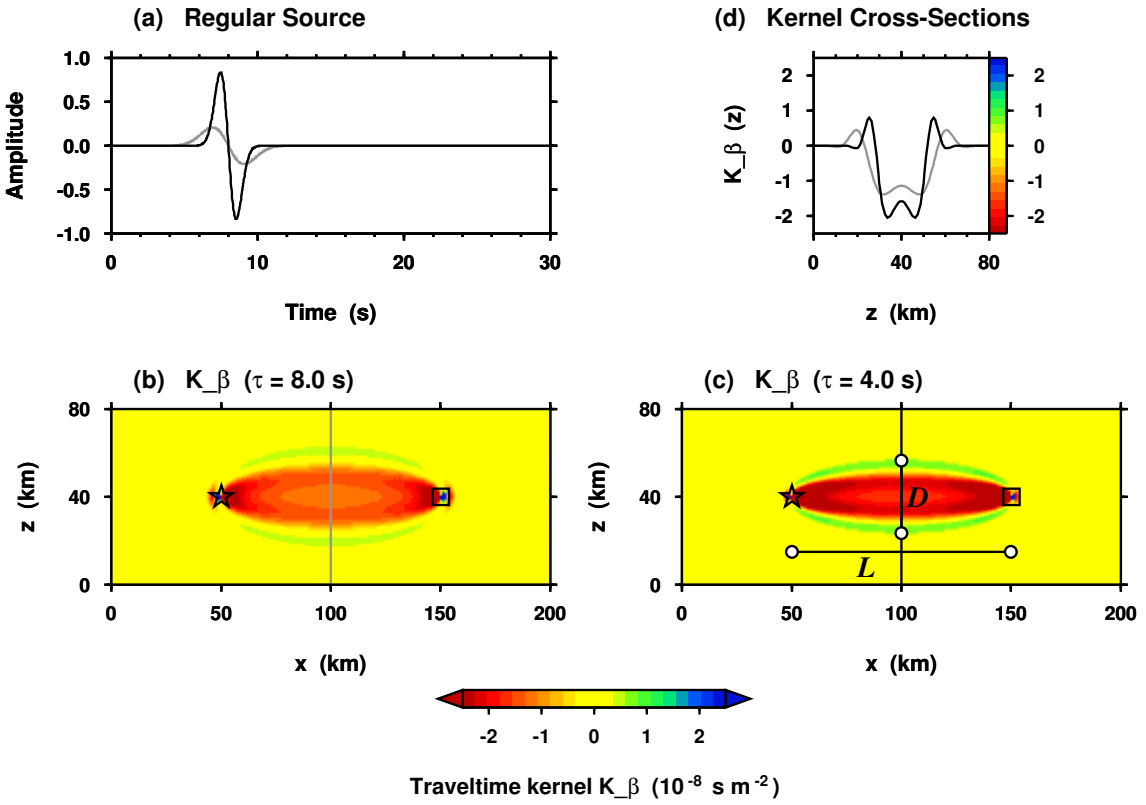


Figure 1.1: Frequency dependence of sensitivity kernels (after *Tromp et al.*, 2005, Figure 6). (a) Two source-time functions for the regular wavefield with durations of $\tau = 8.0$ s (gray) and $\tau = 4.0$ s (black) (see Eq. 3.2). (b) $\bar{K}_{\beta(\alpha\rho)}$ for $\tau = 8.0$ s. (c) $\bar{K}_{\beta(\alpha\rho)}$ for $\tau = 4.0$ s. $D = 33$ km is the width of the first Fresnel zone, estimated as $D \approx \sqrt{\lambda L} = \sqrt{\beta T L}$, where $T = 3.4$ s is the dominant period of the seismic wave, and $\beta = 3.2$ km/s is the shear wavespeed and $L = 100$ km is the path length. (d) Depth cross sections of (b) and (c) at a horizontal distance of $x = 100$ km. As expected, the higher-frequency kernel is narrower in width and greater in amplitude. See Chapter 3 for details on how these kernels were constructed.

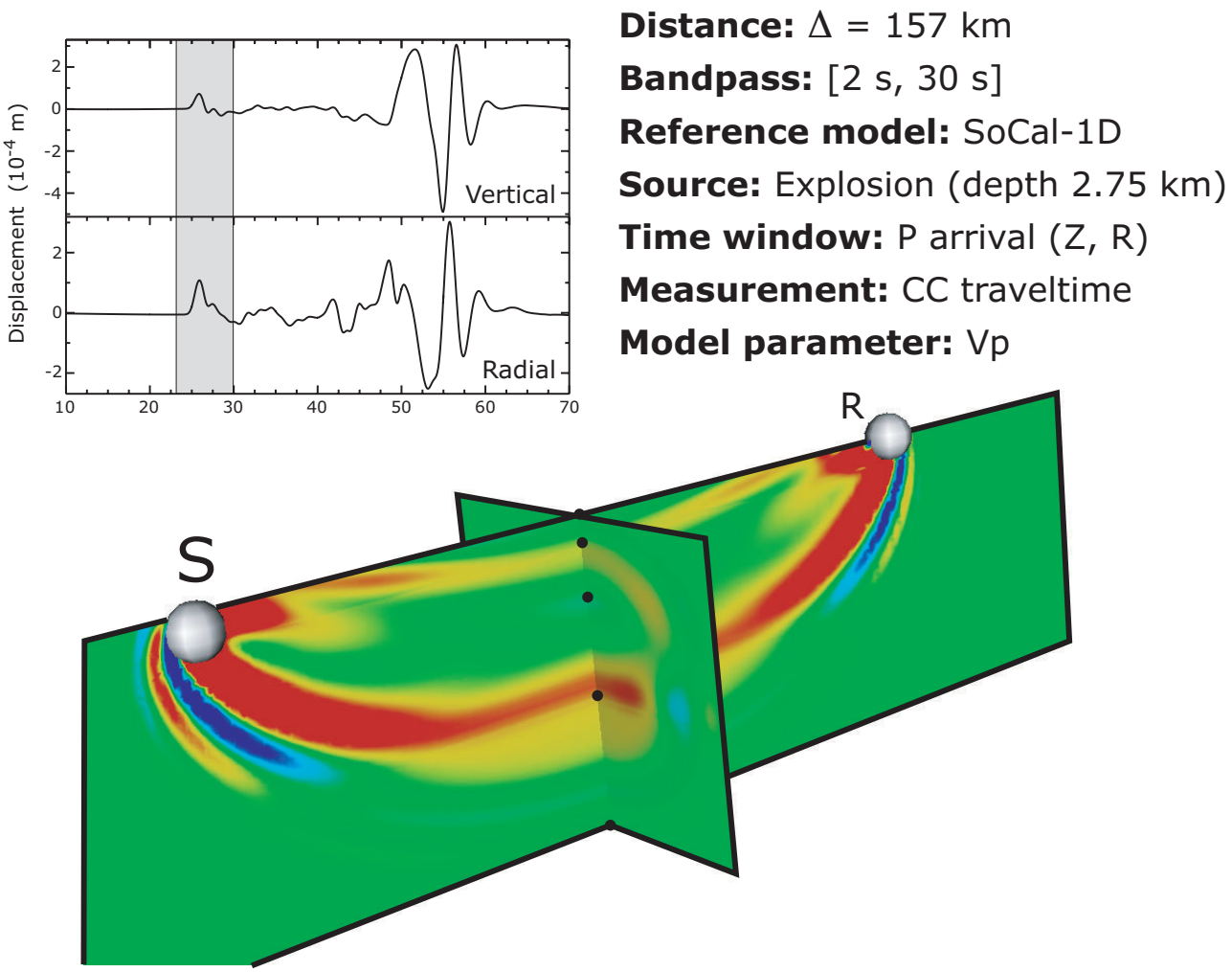


Figure 1.2: Sensitivity kernel for a crustal P wave in a 1D southern California model (*Kanamori and Hadley*, 1975; *Wald et al.*, 1995). The time window highlighted in the vertical and radial components of the seismograms corresponds to the volumetric sensitivity kernel, shown in two perpendicular cross sections. S is the source location, and R is the receiver. The dots mark the layer boundaries of the 1D model at 5.5, 16.0, and 32.0 km, with the bottom at 60.0 km. This sensitivity kernel indicates that the P arrival is a combination of Pn, diffracted along the Moho, and PmP, reflected at the Moho. Several factors, listed in the upper right, influence the nature of the sensitivity kernel, which can be interpreted as follows: "Given a cross-correlation traveltime measurement, ΔT, between observed and simulated seismic waveforms (bandpass period range 2–30 s) within the selected time window, the sensitivity kernel illuminates the volumetric region of the V_P wavespeed model that should be perturbed in order to reduce ΔT." (However, this does not ensure that the wavespeed perturbation is a step in the direction of the actual wavespeed structure, because the ΔT could have arisen from an incorrect description of the earthquake source.)

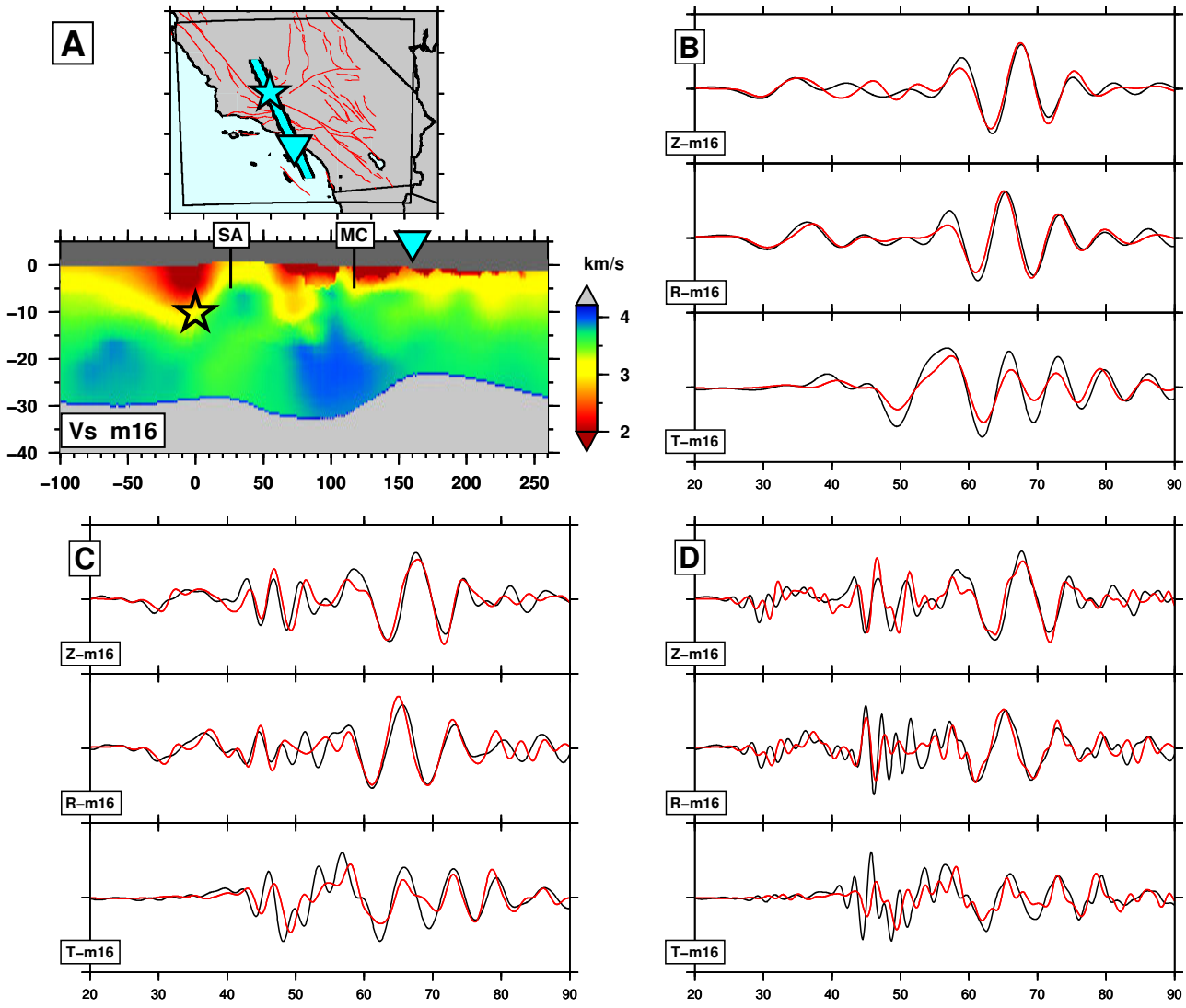


Figure 1.3: The frequency dependence of the seismic wavefield. (a) Cross section of adjoint tomography crustal model $\mathbf{m}_{16}$ (Chapter 6) from event 14186612 to station FMP.CI. SA, San Andreas fault; MC, Malibu Coast fault. (b) Data (black) and 3D synthetics (red), filtered in the period range 6–30 s. Z, vertical component, R, radial component, T, transverse component. (c) Same as (b), for the period range 3–30 s. (d) Same as (b), for the period range 2–30 s.

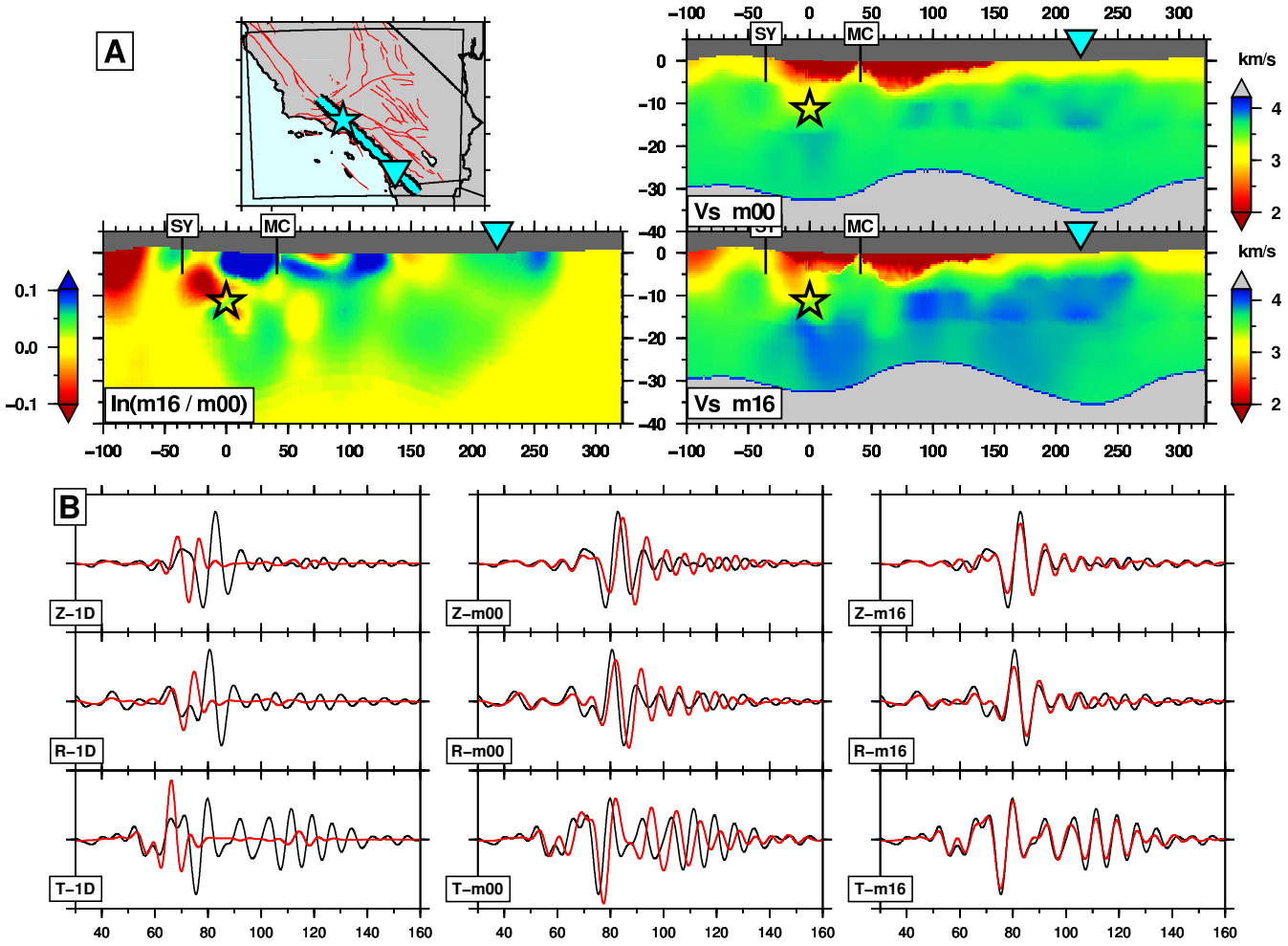


Figure 1.4: Iterative improvement in seismic waveforms. (a) Initial 3D model $\mathbf{m}_{00}$, final 3D model $\mathbf{m}_{16}$, and the difference between the two models, $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$. (b) Data (black) and 3D synthetics (red), filtered in the period range 6–30 s. Z, vertical component, R, radial component, T, transverse component. Left column: synthetics generated using the standard 1D southern California model (*Kanamori and Hadley*, 1975; *Wald et al.*, 1995). Center column: synthetics generated using $\mathbf{m}_{00}$, the 3D model of *Komatitsch et al.* (2004). Right column: synthetics generated using $\mathbf{m}_{16}$, the 16th iteration of the crustal model.

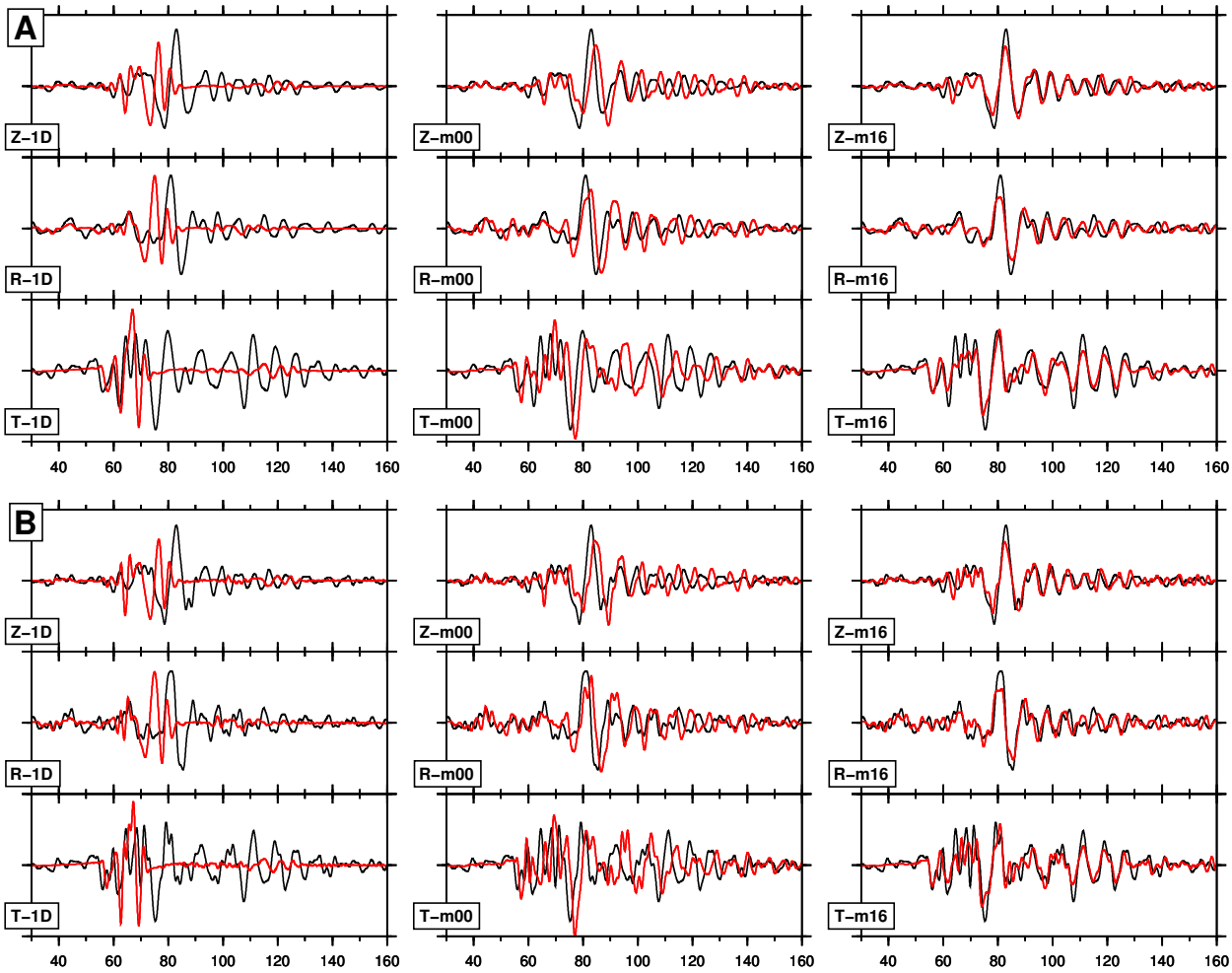


Figure 1.5: Same seismograms as in Figure 1.4b, but for the period ranges 3–30 s (a) and 2–30 s. Z, vertical component, R, radial component, T, transverse component. Left column: synthetics generated using the standard 1D southern California model (*Kanamori and Hadley*, 1975; *Wald et al.*, 1995). Center column: synthetics generated using $\mathbf{m}_{00}$, the 3D model of *Komatitsch et al.* (2004). Right column: synthetics generated using $\mathbf{m}_{16}$, the 16th iteration of the crustal model.

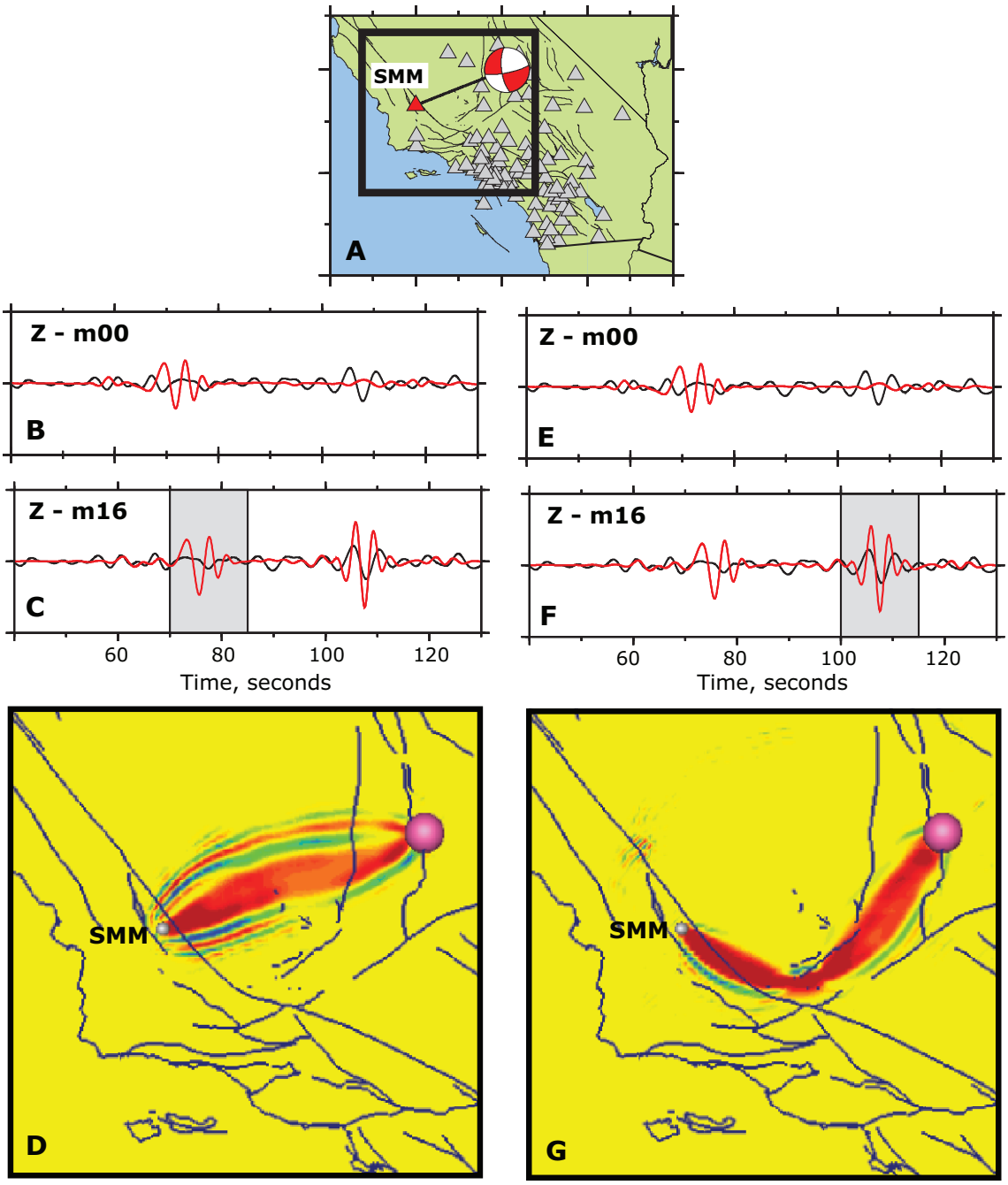


Figure 1.6: Reflected Rayleigh wave at the Tehachapi Mountains. The manually-picked measurement windows shown in (c) and (f) highlight two different surface waves. All seismograms are filtered in the period range 3–30 s. (a) Map showing earthquake source (10992159) and station coverage. Black box shows the region in (d) and (g). (b) Data (black) and $\mathbf{m}_{00}$ 3D synthetics (red). (c) Data (black) and $\mathbf{m}_{16}$ 3D synthetics (red). Synthetic waveform within the time window follows the propagation path shown in (d). The phase of the synthetic waveform is about right ($\Delta T = 1.7$ s), but the amplitude is much to large $\Delta \ln A = -1.5$. (d) Horizontal cross section at 4 km depth of a volumetric sensitivity kernel corresponding to the windowed synthetic waveform in (c). The path illuminates the "direct" Rayleigh wave, though the path does not lie on the great circle between source and station. (e)–(g) Same as (b)–(d), but highlighting the second time window, which is only apparent in the $\mathbf{m}_{16}$ synthetics. The Rayleigh wave reflects at the Tehachapi Mountains, near the southernmost San Joaquin basin, before reaching SMM.

Chapter 2

Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves

Note

This chapter was published as a paper by C. Tape, Q. Liu, and J. Tromp in *Geophysical Journal International* in 2007. Supplemental derivations and tables are included in Appendix A.

Summary

We employ adjoint methods in a series of synthetic seismic tomography experiments to recover surface-wave phase-speed models of southern California. Our approach involves computing the Fréchet derivative for tomographic inversions via the interaction between a forward wavefield, propagating from the source to the receivers, and an "adjoint" wavefield, propagating from the receivers back to the source. The forward wavefield is computed using a 2D spectral-element method (SEM) and a phase-speed model for southern California. A "target" phase-speed model is used to generate the "data" at the receivers. We specify an objective or misfit function that defines a measure of misfit between data and synthetics.

For a given receiver, the remaining differences between data and synthetics are time reversed and used as the source of the adjoint wavefield. For each earthquake, the interaction between the regular and adjoint wavefields is used to construct finite-frequency sensitivity kernels, which we call *event kernels*. An event kernel may be thought of as a weighted sum of phase-specific (e.g., P) *banana-doughnut kernels*, with weights determined by the measurements. The overall sensitivity is simply the sum of event kernels, which defines the *misfit kernel*. The misfit kernel is multiplied by convenient orthonormal basis functions that are embedded in the SEM code, resulting in the gradient of the misfit function, i.e., the Fréchet derivative. A nonlinear conjugate gradient algorithm is used to iteratively improve the model while reducing the misfit function. We illustrate the construction of the gradient and the minimization algorithm, and consider various tomographic experiments, including source inversions, structural inversions, and joint source-structure inversions. Finally, we draw connections between classical Hessian-based tomography and gradient-based adjoint tomography.

2.1 Introduction

Seismic tomography is in a state of transition from ray-based inversions using 1D reference models toward finite-frequency-kernel-based inversions using 3D reference models (*Akçelik et al.*, 2003; *Zhao et al.*, 2005). The transition from ray- to kernel-based inversions has been motivated in part by the pioneering studies of *Marquering et al.* (1999), *Zhao et al.* (2000), and *Dahlen et al.* (2000), which were based on 1D reference models but showed that seismological measurements are sensitive to structure away from the ray path and are affected by wavefront healing. The transition from 1D to 3D reference models has been motivated by computational advances coupled with success in modeling the forward problem of seismic wave propagation in complex media (e.g., *Komatitsch and Vilotte*, 1998; *Komatitsch et al.*, 2002; *Capdeville et al.*, 2003).

The purpose of this paper is to illustrate an approach for "3D–3D" seismic tomography, by which we mean seismic tomography based on a 3D reference model, 3D numerical simulations of the complete seismic wavefield, and finite-frequency sensitivity kernels. The success of 3D–3D tomography depends largely on two factors: (1) the accuracy and efficiency of the technique used to generate 3D synthetic seismograms, and (2) the efficiency of the inver-

sion algorithm. We have implemented numerical methods — the spectral-element method (SEM) — on parallel computers to simulate 3D seismic wave propagation at regional and global scales (e.g., *Komatitsch and Tromp*, 1999; *Komatitsch et al.*, 2004; *Komatitsch and Tromp*, 2002a,b). The inverse problem can be cast as a minimization problem, where the objective or misfit function measures some difference between data and synthetic seismograms computed from a 3D model. Our approach to the inverse problem utilizes adjoint methods (*Tarantola*, 1984; *Talagrand and Courtier*, 1987), which provide the gradient of the misfit function but not its second derivatives, i.e., the Hessian. The efficiency of the inverse algorithm is controlled by the computation of the gradient, which requires only two 3D simulations per earthquake (i.e., the gradient is independent of the number of receivers or the number of measurements), as well as an effectively chosen gradient method.

The framework for 3D–3D tomographic inversions using adjoint methods was developed in exploration geophysics (e.g., *Tarantola*, 1984; *Gauthier et al.*, 1986; *Mora*, 1987; *Pratt et al.*, 1998; *Pratt*, 1999). These studies illustrated the computation of the gradient and the related inversion technique using 2D heterogeneous models and 2D numerical algorithms. Applications of 3D–3D tomographic techniques are presented in *Bijwaard and Spakman* (2000), *Zhao et al.* (2005), *Capdeville et al.* (2005), and *Akçelik et al.* (2003), among others. *Bijwaard and Spakman* (2000) performed 3D ray-tracing though 3D models to iteratively improve a global P-wave model. *Zhao et al.* (2005) used fully numerical methods (finite differencing) to compute traveltime misfit function gradients for 3D models of the greater Los Angeles area. *Capdeville et al.* (2005), using synthetic data, demonstrated a technique of stacking synthetic records that limits the number of forward simulations to one per event (per model iteration); however, the technique requires modification when the data set is incomplete, as is generally the case. *Akçelik et al.* (2003), using synthetic data, illustrated a tomographic inversion using a finite-element method together with an adjoint approach within a conjugate gradient framework. They also addressed multiscale approaches to the inverse problem in an attempt to avoid reaching local minima during the inversion.

This paper is an extension of *Tromp et al.* (2005), which synthesized the work on adjoint methods with studies in finite-frequency tomography (*Marquering et al.*, 1999; *Dahlen et al.*, 2000; *Zhao et al.*, 2000) and time-reversal imaging (*Fink et al.*, 1989; *Fink*, 1992, 1997). In *Tromp et al.* (2005) we illustrated how the computation of a sensitivity kernel for a particular model and a particular type of measurement could be achieved via the interaction of two

wavefields, one constituting the "regular" wavefield traveling from source to receiver, and the other constituting the "adjoint" wavefield traveling from receiver to source, constructed by a suitable time-reversed synthetic seismogram recorded at the receiver. We performed a simple source inversion to illustrate the conjugate gradient algorithm, whereby only the gradient of the misfit function is used to iteratively invert for the source parameters. In this paper, we use the conjugate gradient approach to illustrate wave-speed inversions, source inversions, and joint (source and structure) inversions. In each example, the "observed" seismograms are computed for a "target" model, and the synthetic seismograms are computed from a current model that iteratively improves toward the target model over the course of the inversion. All of the simulations illustrated in this paper were performed on a single Linux PC.

We begin by highlighting the differences between classical and adjoint tomography in the context of a minimization problem. We define *classical tomography* as a Newton inversion scheme that computes model sensitivities for each measurement by constructing the gradient and Hessian of the misfit function (Section 2.3) (e.g., *Woodhouse and Dziewonski*, 1984; *Ritsema et al.*, 1999). In *adjoint tomography* only the gradient is computed, and it is computed via adjoint methods (e.g., *Gauthier et al.*, 1986; *Akçelik et al.*, 2003). In Section 2.5 we illustrate the construction of a *misfit kernel*, which can be thought of as the gradient of the misfit function. In Section 2.6 we show how this gradient is used in the conjugate gradient algorithm to iteratively improve the model. We finish by showing several tomographic experiments, including simultaneous source-structure inversions, as well as a comparison between ray- and kernel-based classical inversions and adjoint tomography.

2.2 General formulation of the inverse problem

Our objective will be to minimize a measure of the misfit between a set of data, for example waveforms or traveltimes, and a complementary set of synthetics. The generated synthetics are based on a model $\mathbf{m}$, for example a set of structural and source parameters, and our aim is to reduce the misfit between the data and the synthetics by making (successive) model corrections $\delta\mathbf{m}$. We define the *misfit function* $F(\mathbf{m})$ to be a measure of misfit between the data and synthetics computed for model $\mathbf{m}$. The function F is alternatively called an "objective" or "cost" function. For example, F could represent least-squares measures of

waveform or traveltime differences.

Let us suppose we have a particular model $\mathbf{m}$, and we wish to obtain an updated model $\mathbf{m} + \delta\mathbf{m}$ that brings us closer to a minimum of the misfit function F (*Nolet*, 1987; *Tarantola*, 2005, Appendix 6.22). We make a quadratic Taylor expansion of $F(\mathbf{m} + \delta\mathbf{m})$:

$$F(\mathbf{m} + \delta\mathbf{m}) \approx F(\mathbf{m}) + \mathbf{g}(\mathbf{m})^T \delta\mathbf{m} + \frac{1}{2}\delta\mathbf{m}^T \mathbf{H}(\mathbf{m})\, \delta\mathbf{m}, \tag{2.1}$$

where the gradient vector $\mathbf{g}(\mathbf{m})$ is defined in terms of the first derivative of the misfit function (also known as the Fréchet derivative) by

$$\mathbf{g}(\mathbf{m}) = \left.\frac{\partial F}{\partial \mathbf{m}}\right|_{\mathbf{m}}, \tag{2.2}$$

and the Hessian matrix $\mathbf{H}(\mathbf{m})$ is defined in terms of the second derivatives of the misfit function by

$$\mathbf{H}(\mathbf{m}) = \left.\frac{\partial^2 F}{\partial \mathbf{m} \partial \mathbf{m}}\right|_{\mathbf{m}}. \tag{2.3}$$

The "$|_{\mathbf{m}}$" dependence is used to emphasize that the preceding variable is evaluated at model $\mathbf{m}$.

The gradient of (2.1) with respect to $\delta\mathbf{m}$ is given by

$$\mathbf{g}(\mathbf{m} + \delta\mathbf{m}) \approx \mathbf{g}(\mathbf{m}) + \mathbf{H}(\mathbf{m})\, \delta\mathbf{m}, \tag{2.4}$$

which can be set equal to zero to obtain the (local) minimum of (2.1):

$$\mathbf{H}(\mathbf{m})\, \delta\mathbf{m} = -\mathbf{g}(\mathbf{m}). \tag{2.5}$$

An updated model $\mathbf{m} + \delta\mathbf{m}$ may be obtained with or without the Hessian $\mathbf{H}$. If the gradient and Hessian (or approximate Hessian) are both available, then the inverse approach is known as a *Newton method*; if only the gradient is available, then it is a *gradient method* (e.g., steepest descent, conjugate gradient). In classical traveltime tomography, one generally has access to both the gradient $\mathbf{g}$ and the Hessian $\mathbf{H}$ of the misfit function, in which case the model update $\delta\mathbf{m}$ may be obtained based on (2.5). For complex, heterogeneous models, computation of the gradient is generally still feasible, but computation of the Hes-

sian is not. In the absence of the Hessian, one can minimize the misfit function using only the gradient (2.2) based on iterative methods.

2.3 Classical tomography

We begin by investigating 2D surface-wave traveltime tomography based on either ray or finite-frequency sensitivity kernels. These classical inversions, which involve access to both the gradient and the Hessian of the misfit function, serve as a reference and standard for subsequent iterative inversions based on only the gradient (Section 2.6). In particular, we will investigate how many iterations of the conjugate-gradient adjoint approach are required to obtain a similar misfit to the data as an inversion based on knowledge of the gradient and Hessian. Of course our ultimate goal is to use the adjoint approach to address inverse problems for fully 3D reference models, when the calculation of the Hessian is generally not feasible, and the experiments in this paper serve as a guide to the implementation and convergence of such iterative inversions.

2.3.1 Theory

The traveltime misfit function may be expressed as

$$F(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{N}\left[T_i^{\text{obs}} - T_i(\mathbf{m})\right]^2, \tag{2.6}$$

where T_i^{obs} denotes the observed traveltime for the ith source-receiver combination, $T_i(\mathbf{m})$ the predicted traveltime based on the current model $\mathbf{m}$, and N the number of traveltime measurements. The variation of the misfit function (2.6) is given by

$$\delta F = -\sum_{i=1}^{N}\Delta T_i\ \delta T_i, \tag{2.7}$$

where δT_i is the theoretical traveltime perturbation and

$$\Delta T_i = T_i^{\text{obs}} - T_i(\mathbf{m}) \tag{2.8}$$

denotes the traveltime anomaly. The sign convention for the traveltime anomaly follows that of *Dahlen et al.* (2000) and *Dahlen and Baig* (2002), such that a negative traveltime

indicates a delay in the synthetic arrival relative to the recorded arrival. Throughout this paper, an uppercase delta, Δ, will denote a differential measurement, and a lowercase delta, δ, will denote a mathematical perturbation.

In ray-based tomography, the predicted traveltime anomaly δT_i along the ith ray path may be related to fractional wave-speed perturbations $\delta \ln c = \delta c/c$ based on the relationship

$$\delta T_i = -\int_{\text{ray}_i} c^{-1}\, \delta \ln c \, \mathrm{d}s, \tag{2.9}$$

where $\mathrm{d}s$ denotes a segment of the ith ray.

Taking into account finite-frequency effects, *Marquering et al.* (1999), *Zhao et al.* (2000), and *Dahlen et al.* (2000) demonstrate that the traveltime anomaly may alternatively be related to relative wave-speed perturbations based on a finite-frequency sensitivity kernel $K_i(\mathbf{x})$ for the ith source-receiver combination by

$$\delta T_i = \int_V K_i \, \delta \ln c \, \mathrm{d}^3\mathbf{x}. \tag{2.10}$$

Marquering et al. (1999) dubbed these finite-frequency kernels "banana-doughnut kernels" on account of their shape in smooth, spherically symmetric Earth models for cross-correlation traveltime measurements. These kernels are also referred to as "sensitivity," "finite-frequency," or "Born" kernels. For our purposes, the key point is that a banana-doughnut kernel does not incorporate the traveltime measurement, whereas the event and misfit kernels discussed in Section 2.5 do incorporate measurements.

Unlike the ray-theoretical expression (2.9), equation (2.10) relates the traveltime anomaly to 3D heterogeneity $\delta \ln c$ throughout the entire Earth model, as seen through the kernel K_i. The relations (2.9) and (2.10) are valid for any model. Frequently the model is chosen to be one-dimensional because this makes the ray and finite-frequency kernel calculations much simpler, but this is not required (*Zhao et al.*, 2005).

Substituting (2.10) into (2.7), we express the variation of the traveltime misfit function for finite-frequency tomography as

$$\delta F = \int_V K \, \delta \ln c \, \mathrm{d}^3\mathbf{x}, \tag{2.11}$$

where the traveltime *misfit kernel* $K(\mathbf{x})$ is a weighted sum of the kernels $K_i(\mathbf{x})$:

$$K(\mathbf{x}) = -\sum_{i=1}^{N} \Delta T_i \, K_i(\mathbf{x}), \tag{2.12}$$

such that the weight associated with the kernel for the ith source-receiver combination K_i is the corresponding traveltime anomaly ΔT_i. It is important to note the distinction that misfit kernels $K(\mathbf{x})$ depend on the data, whereas the banana-doughnut kernels $K_i(\mathbf{x})$ are data independent.

To make the tomographic inversions practical, we need to choose a finite set of basis functions in which to expand our model. Let $B_k(\mathbf{x})$, $k = 1, \ldots, M$, denote a set of M basis functions. We expand our fractional phase-speed perturbations, $\delta \ln c(\mathbf{x})$, into these basis functions:

$$\delta \ln c(\mathbf{x}) = \sum_{k=1}^{M} \delta m_k \, B_k(\mathbf{x}), \tag{2.13}$$

where δm_k, $k = 1, \ldots, M$, represent the perturbed model coefficients, which are determined in terms of the gradient $\mathbf{g}$ and Hessian $\mathbf{H}$ of the misfit function by (2.5).

Next, we determine $\mathbf{g}$ and $\mathbf{H}$ for this classical traveltime tomography problem. Substituting (2.13) into (2.9) and (2.10), respectively, we obtain

$$\delta T_i = \sum_{k=1}^{M} \delta m_k \, G_{ik}, \tag{2.14}$$

where for ray theory,

$$G_{ik} \equiv \left. \frac{\partial T_i}{\partial m_k} \right|_{\mathbf{m}} = -\int_{\mathrm{ray}_i} c^{-1} B_k \, \mathrm{d}s, \tag{2.15}$$

whereas for finite-frequency tomography,

$$G_{ik} \equiv \left. \frac{\partial T_i}{\partial m_k} \right|_{\mathbf{m}} = \int_V K_i B_k \, \mathrm{d}^3\mathbf{x}. \tag{2.16}$$

We note that in either case G_{ik} will depend on the source-receiver geometry (index i), the choice of basis functions (index k), and the choice of reference model ($\mathbf{m}$).

Using (2.13) we express the variation in the misfit function (2.11) as

$$\delta F = \sum_{k=1}^{M} \int_V K B_k \, \mathrm{d}^3\mathbf{x} \, \delta m_k. \tag{2.17}$$

Upon comparing this result to

$$\delta F = \mathbf{g} \cdot \delta\mathbf{m} = \sum_{k=1}^{M} g_k \, \delta m_k, \tag{2.18}$$

we deduce that the elements of the gradient vector, g_k, are determined by

$$g_k = \frac{\partial F}{\partial m_k} = \int_V K B_k \, \mathrm{d}^3\mathbf{x}, \qquad k = 1, \ldots, M. \tag{2.19}$$

This highlights the simple relationship between the misfit kernel and the gradient of the misfit function. Substituting (2.12) into (2.19), we obtain

$$g_k = -\sum_{i=1}^{N} \int_V K_i B_k \, \mathrm{d}^3\mathbf{x} \, \Delta T_i = -\sum_{i=1}^{N} G_{ik} \, \Delta T_i, \qquad k = 1, \ldots, M, \tag{2.20}$$

which in matrix notation becomes

$$\mathbf{g} = -\mathbf{G}^T \mathbf{d}. \tag{2.21}$$

Here $\mathbf{G}$ is the $N \times M$ design matrix constructed using (2.15) for rays or (2.16) for finite-frequency kernels, a superscript T denotes the transpose, and $\mathbf{d}$ is defined as an N-dimensional data vector of cross-correlation traveltime measurements:

$$\mathbf{d} = (\Delta T_1, \ldots, \Delta T_i, \ldots, \Delta T_N)^T. \tag{2.22}$$

Note that the data vector depends on model $\mathbf{m}$ through the synthetics.

The second derivatives of the misfit function are given by (2.3), and thus the elements of the Hessian $\mathbf{H}$ are given by

$$H_{kk'} = \left.\frac{\partial^2 F}{\partial m_k \partial m_{k'}}\right|_{\mathbf{m}} = \left.\frac{\partial g_k}{\partial m_{k'}}\right|_{\mathbf{m}} = \sum_{i=1}^{N} \left[G_{ik'} G_{ik} + \Delta T_i \left.\frac{\partial^2 T_i}{\partial m_k \partial m_{k'}}\right|_{\mathbf{m}} \right], \tag{2.23}$$

where G_{ik} is defined in (2.16). We introduce an *approximate Hessian* $\tilde{\mathbf{H}}$ by ignoring the second-order terms:

$$\tilde{H}_{kk'} \equiv \sum_{i=1}^{N} G_{ik} G_{ik'}, \qquad k, k' = 1, \ldots, M, \tag{2.24}$$

which in matrix notation is

$$\tilde{\mathbf{H}} \equiv \mathbf{G}^T \mathbf{G}. \tag{2.25}$$

Henceforth, we will refer to $\tilde{\mathbf{H}}$ as the Hessian. This approximation, $\tilde{\mathbf{H}} \approx \mathbf{H}$, characterizes the *Gauss–Newton method* and is exact if the model perturbations are linearly related to the traveltime measurements.

Having established the gradient (2.21) and Hessian (2.25), the model correction $\delta\mathbf{m}$ is determined by (2.5):

$$\mathbf{G}^T \mathbf{G}\, \delta\mathbf{m} = \mathbf{G}^T \mathbf{d}, \tag{2.26}$$

where $\delta\mathbf{m}$ is defined in (2.13), $\mathbf{d}$ is defined in (2.22), and $\mathbf{G}$ is defined according to (2.15) or (2.16).

In general, the Hessian matrix (2.25) is not full rank, which means that its inverse does not exist. To stabilize the inverse problem, one introduces a damping matrix $\mathbf{D}$ typically involving the norm, gradient, or second derivative of the wave-speed perturbations, and a damping parameter γ:

$$\tilde{\mathbf{H}}_\gamma = \mathbf{G}^T \mathbf{G} + \gamma^2 \mathbf{D}. \tag{2.27}$$

The damping parameter γ is chosen in a subjective manner, generally by inspecting a graph that trades off misfit of the solution against complexity of the model. Having stabilized the inverse of the Hessian, the solution to (2.26) may now be expressed as

$$\delta\mathbf{m} = (\mathbf{G}^T \mathbf{G} + \gamma^2 \mathbf{D})^{-1} \mathbf{G}^T \mathbf{d}, \tag{2.28}$$

from which the updated model, $\mathbf{m} + \delta\mathbf{m}$, may be obtained. In Section 2.10 we show how (2.28) is obtained by adding a regularization term to the misfit function. More generally,

for nonlinear inverse problems one uses an iterative Gauss–Newton method to minimize the misfit function. In that case (2.28) is replaced by an iterative expression that relates model $k+1$ to model k and the initial model (e.g., *Tarantola*, 2005).

2.3.2 Experimental setup

We simulate 2D elastic wave propagation using a spectral-element method (SEM), which combines the flexible spatial parameterization of finite-element methods with the accuracy of pseudospectral methods (e.g., *Komatitsch and Vilotte*, 1998; *Komatitsch and Tromp*, 1999). For simplicity, we consider "membrane waves" (*Tanimoto*, 1990; *Peter et al.*, 2007) traveling in the x-y plane with a vertical (z) component of motion. The elastic wave equation for the vertical component of displacement $s(x, y, t)$ is given by

$$\rho\,\partial_t^2 s = \partial_x(\mu\,\partial_x s) + \partial_y(\mu\,\partial_y s) + f, \tag{2.29}$$

where $\rho(x, y)$ denotes the density distribution and $\mu(x, y)$ the shear modulus. The source $f(x, y, t)$ is given by

$$f(x, y, t) = h(t)\,\delta(x - x_s)\,\delta(y - y_s), \tag{2.30}$$

where $h(t)$ denotes the source time function and (x_s, y_s) the source location. All four membrane edges are absorbing, and attenuation and anisotropy are not incorporated. The relationship between membrane-wave phase speed, c, and rigidity is $\mu = \rho c^2$.

We take southern California as our region of interest (Figure 2.1) in anticipation of eventually improving the present 3D reference wave-speed models (*Hauksson*, 2000; *Magistrale et al.*, 2000; *Süss and Shaw*, 2003). The modeled region is 480 km by 480 km. The numerical simulations are carried out on a planar grid with $N_{\text{glob}} = 25921$ gridpoints. The source time function of the point source (2.30) used in the simulations is a Gaussian of the form

$$h(t) = (-2\alpha^3/\sqrt{\pi})\,(t - t_s)\,\exp[-\alpha^2(t - t_s)^2], \tag{2.31}$$

where $\alpha = 2\tau_0/\tau$, $\tau_0 = 2.628$ s, $\tau = 20.0$ s is the duration of $h(t)$, and $t_s = 48.0$ s is the origin time (e.g., Figure 2.6a). The duration of each simulation is $T = 240$ s unless

otherwise noted.

The synthetic records are computed using source locations of actual events ($M \geq 4$) recorded in southern California between 1990 and 2005 (Figure 2.1). The initial set of synthetics is computed using a model with homogeneous phase speed c. In general, the synthetics in our experiments are generated from a laterally varying model, while the data are generated from what is designated as the "target" model. Computationally, the model correction is expressed as a fractional perturbation, $\delta c/c = \delta \ln c$, with current phase speed c. In the figures, however, each phase-speed model is plotted as a percent perturbation from the phase-speed value for the initial model. In Section 2.8 we allow for additional perturbations in the source parameters, so that in general the synthetics are computed from a model with perturbed sources and perturbed structure.

2.3.3 2D tomographic example

To illustrate a classical tomographic inversion, we begin by choosing a set of basis functions, $B_k(\mathbf{x})$, in which to expand the fractional wave-speed perturbations $\delta \ln c(\mathbf{x})$ (2.13). We use spherical spline basis functions (*Wang and Dahlen*, 1995; *Wang et al.*, 1998), which are well-suited for regional models where multiscale parameterization is desired because of nonuniform path coverage (e.g., *Boschi et al.*, 2004). (We do not exploit the multiscale aspects here.) An example of a spherical spline basis function is plotted in Figure 2.2b. We choose $M = 286$ spherical spline basis functions to cover the southern California region.

The data are computed using the phase-speed model in Figure 2.1b, and the synthetics are computed for a homogeneous phase-speed model with $c = 3.78$ km/s. We make cross-correlation traveltime measurements between data and synthetics to obtain the data vector $\mathbf{d}$ (2.22). The total number of measurements is $N = N_{\text{events}} \times N_{\text{receivers}} = 25 \times 132 = 3300$.

We illustrate the classical tomographic approach using both rays and banana-doughnut kernels to represent the sensitivities of the measurements to the model parameters. Thus we compute two $N \times M$ design matrices, $\mathbf{G}^{\text{ray}}$ and $\mathbf{G}^{\text{ker}}$, respectively. Figure 2.2a–c shows the computation of a single G_{ik}^{ray} element, and Figure 2.2d–f shows an example for G_{ik}^{ker}. Figure 2.2 illustrates why the choice between kernels or rays may be moot, depending on the resolution of the basis functions. The infinitesimally thin ray path is smeared out by the relatively smooth basis functions. Thus, in our example, $\mathbf{G}^{\text{ray}} \approx \mathbf{G}^{\text{ker}}$, and we will simply use a generic $\mathbf{G}$ to denote either the ray or kernel design matrix.

The (approximate) Hessian matrix, $\tilde{\mathbf{H}} = \mathbf{G}^T\mathbf{G}$, and the gradient vector, $\mathbf{g} = -\mathbf{G}^T\mathbf{d}$, are visualized in Figure 2.3. The Hessian and gradient are determined by the source-receiver geometry and the banana-doughnut kernels (or ray paths), but only the gradient is controlled by the data.

Figure 2.4 shows the model recovery using classical tomography based on a single iteration of the Gauss–Newton method. The recovered model is strongly dependent on the damping parameter γ. When $\gamma \approx 0$, the inverse is unstable and structure is artificially introduced into regions where there is no coverage, i.e., the edges of the domain and the oceans (Figure 2.4a). When $\gamma \to \infty$, the recovered model is simply the initial model (Figure 2.4f), although the spatial pattern is that of the gradient (e.g., compare Figure 2.3c with Figure 2.4g). The reason for this is that for large values of the damping parameter γ the damped Hessian (2.27) is dominated by the damping matrix $\mathbf{D}$, which in our case is the identity matrix $\mathbf{I}$. In this case the solution to the inverse problem given by (2.28) is effectively a scaled version of the gradient $\mathbf{g}$. For the example in Figure 2.4, the L-curve suggests that $\gamma = 10.0$ is a reasonable model selection; this model is shown in Figure 2.4c and Figure 2.20c.

2.4 Computation of the gradient and Hessian

Obtaining the Hessian involves computing banana-doughnut kernels K_i for each source-receiver combination. Thus, the cost of computing the Hessian is the cost of computing all the kernels. For a problem involving N_{events} earthquakes, $N_{\text{receivers}}$ stations, $N_{\text{comp}} = 3$ component seismograms, and N_{picks} measurements per seismogram one would need to calculate $N_{\text{events}} \times N_{\text{receivers}} \times N_{\text{comp}} \times N_{\text{picks}}$ kernels.

In adjoint tomography one computes a misfit kernel K from which only the gradient is obtained. One of the primary benefits of adjoint tomography is that the misfit kernel need not be computed by summing over individual banana-doughnut kernels for each source-receiver pair, as in (2.12). Instead, the measurements, ΔT_i, are incorporated into the adjoint source, which is used to compute the misfit kernel (Section 2.5). This kernel is constructed via the interaction between a forward wavefield and an adjoint wavefield, requiring only two simulations per earthquake (*Tromp et al.*, 2005). So if our inverse problem involves N_{events} earthquakes, obtaining the gradient of the misfit function involves $2N_{\text{events}}$

numerical simulations, i.e., this calculation is *independent of the number of receivers, components, and picks*. The main drawback of adjoint tomography is that the Hessian is not available, which means that iterative techniques must be used to determine the minimum of the objective function.

Thus, a fundamental distinction between classical and adjoint tomography is whether or not individual banana-doughnut kernels are computed. In the context of classical tomography, there are several ways to compute the kernels. For 1D Earth models, they may be calculated cheaply and rapidly, in particular if approximate expressions are used (*Dahlen et al.*, 2000). Using normal modes, *Zhao and Jordan* (2006) computed global finite-frequency kernels for spherically symmetric models. The kernels may be used to construct the design matrix $\mathbf{G}$, which has $N_{\text{events}} \times N_{\text{receivers}} \times N_{\text{comp}} \times N_{\text{picks}} \times M$ elements. The parameterization of the model (2.13) must be carefully considered, since M scales $\mathbf{G}$. Once $\mathbf{G}$ is obtained, the Hessian follows from (2.25).

The computation of the kernels K_i for 3D models may be accomplished in two ways:

1. We may perform an adjoint simulation for every single measurement, which requires a total of $2N_{\text{events}} \times N_{\text{receivers}} \times N_{\text{comp}} \times N_{\text{picks}}$ simulations (two for each measurement). For 3D models the numerical cost is prohibitive.

2. Alternatively, we may invoke source-receiver reciprocity and for every source and receiver calculate and store Green's functions as a function of both space and time. This requires one to perform and store $N_{\text{events}} + 3N_{\text{receivers}}$ simulations: one simulation for each event and one simulation for each receiver component. For realistic 3D simulations the storage requirements are formidable, although for small problems the approach is feasible, as demonstrated by *Zhao et al.* (2005).

Our goal is to improve fully 3D reference models. Therefore, to make the inverse problem tractable, we are forced to consider an approach based on knowledge of the value of the misfit function $F(\mathbf{m})$, its gradient $\mathbf{g}$, but not its Hessian $\tilde{\mathbf{H}}$. Minimization of the misfit function based on this information may be accomplished using a nonlinear conjugate gradient method, as discussed in Section 2.6. But first we demonstrate how we compute the gradient using adjoint methods.

2.5 The gradient: construction of a misfit kernel

In this section we demonstrate how we compute the gradient of the misfit function, $\mathbf{g} = \partial F/\partial \mathbf{m}$, using adjoint methods. The gradient of the misfit function is obtained from (2.19):

$$g_k = \int_\Omega K B_k \, \mathrm{d}^2\mathbf{x}, \tag{2.32}$$

where for the 2D examples in this paper the integration is over the model surface Ω. Given the misfit kernel, K, and the basis functions, B_k, we can readily compute the gradient of the misfit function. The misfit kernel can also be thought of as a sum of *event kernels*, which we discuss next.

2.5.1 Event kernels

Tromp et al. (2005, Fig. 3) illustrated the construction of a data-independent banana-doughnut kernel based on adjoint methods. In this paper, the kernels we show are *misfit* kernels, whereby the adjoint source is constructed based in part on a set of measurements between data and synthetics.

The construction of misfit kernels based on cross-correlation traveltime measurements is outlined in *Tromp et al.* (2005, Section 4). For membrane waves, motion is restricted to the vertical direction, and the source functions and wavefields are scalar quantities. The source for the adjoint wavefield for a particular event is given by (*Tromp et al.*, 2005, eq. 57)

$$f^\dagger(x,y,t) = -\sum_{r=1}^{N_r} \Delta T_r \, \frac{1}{M_r} \, w_r(T-t) \, \partial_t s(x_r, y_r, T-t) \, \delta(x-x_r) \, \delta(y-y_r), \tag{2.33}$$

where r is the receiver index, N_r is the number of receivers, ΔT_r is the cross-correlation traveltime measurement over a time window $w_r(t)$, $s(x,y,t)$ is the forward wavefield determined by (2.29), (x_r, y_r) is the location of the receiver, T is the length of the time series, and M_r is a normalization factor. The key point is that the adjoint force comprises time-reversed velocity seismograms, input *at the location of the receivers* and weighted by the traveltime measurement associated with each receiver.

For a given earthquake (event), the interaction between the adjoint wavefield and the

forward wavefield gives rise to the membrane *event kernel*

$$K(x,y) = -2\mu(x,y) \int_0^T \left[\partial_x s^\dagger(x,y,T-t) \partial_x s(x,y,t) + \partial_y s^\dagger(x,y,T-t) \partial_y s(x,y,t) \right] \mathrm{d}t. \tag{2.34}$$

Note that the misfit between the data and synthetics is incorporated into the adjoint source (2.33), which gives rise to the adjoint wavefield $s^\dagger$. Equation (2.34) is obtained from the expression for an SH β-kernel in *Tromp et al.* (2005), which contains a product of the adjoint and regular deviatoric strain tensors. In the case of the SH (or membrane) waves, there are four nonzero components (two unique) of each deviatoric strain tensor, which leads to (2.34).

Figures 2.5 and 2.6 show the construction of an event kernel for a single source–receiver pair for a cross-correlation traveltime measurement. The source-receiver geometry and forward wavefield are shown in the left column of Figure 2.5. The synthetics are computed for a homogeneous reference model ($c = 3.50$ km/s), and the data are computed for a uniformly perturbed "target" model with $\delta \ln c = 0.1$, i.e., $c(1 + \delta \ln c) = 3.85$ km/s. The cross-correlation traveltime measurement at the receiver is $\Delta T = -9.72$ s, indicating a late arrival of the synthetics with respect to the data. The adjoint source function is constructed by time-reversing the synthetic velocity recorded at the receiver and multiplying by ΔT (Figure 2.6; eq. 2.33).

We now replace the homogeneous target model with the checkerboard target model in Figure 2.7a. Figure 2.8 shows the construction of an event kernel for this target model for multiple receivers, thereby incorporating multiple measurements. Just as in Figure 2.5, the event kernel that forms in Figure 2.8 highlights the regions of the current model that give rise to the (cross-correlation traveltime) discrepancies between the data and synthetics. However, in Figure 2.8 this is more obvious since the model used to generate the data is not simply a homogeneous perturbation but rather a large-scale checker pattern. The event kernel in Figure 2.8 looks qualitatively similar to the phase-speed model in Figure 2.7, except with the opposite sign, which is consistent with (2.11): for the variation of the misfit function to be negative, we invoke a fast, positive (blue) structural perturbation where the kernel is negative (red), and a slow, negative structural perturbation where the kernel is positive.

As shown in (2.33), the amplitude of the adjoint source at a particular receiver, r, is determined in part by the traveltime measurement ΔT_r. Changing the values of ΔT_r changes the weights of the corresponding individual banana-doughnut kernels that comprise the event kernel, something that is explicit in the classical sense (2.12). It is possible to incorporate some measure of weighting at the stage of constructing the adjoint source in order to account for uneven coverage (Figure 2.1), as demonstrated in *Takeuchi and Kobayashi* (2004). Another option is to weight the adjoint sources according to realistic uncertainties associated with each measurement (*Tarantola*, 1984): a measurement with a high uncertainty will have a small amplitude weight, and thus a relatively weak contribution to the event kernel.

2.5.2 Misfit kernels and damping

We define the *misfit kernel* as the sum of the event kernels for a particular model. Thus, the gradient of the misfit function, $\mathbf{g}$, is obtained as in (2.32) using the misfit kernel $K(\mathbf{x})$. Figure 2.9 shows the construction of a misfit kernel for 25 events. Note that features of each event kernel are very different, even for the simple checkerboard model in this example (Figure 2.7). Only after summing the event kernels does the pattern (Figure 2.9h) begin to resemble the model used to generate the data (Figure 2.9i).

We apply a smoothing operator to the misfit kernels in order to remove spurious amplitudes in the immediate vicinity of the sources and receivers (Figure 2.10). This is accomplished by convolving (in 2D) the unsmoothed misfit kernel with a Gaussian of the form

$$G(x,y) = \frac{4}{\pi\,\Gamma^2}\; e^{-4\,(x^2+y^2)/\Gamma^2}, \tag{2.35}$$

where Γ is the full-width of the Gaussian, defined such that at a (polar) distance $r = \Gamma/2$, the Gaussian has amplitude $G(r) = G(0)e^{-1}$; thus Γ is the scalelength of smoothing (Figure 2.10). The choice of Γ is somewhat analogous to the choice of damping parameter γ for the inversion of the Hessian (eq. 2.27), which involves a degree of subjectivity. In the adjoint method, subjectivity may be removed by selecting Γ according to the shortest wavelengths of the waves. It seems sensible to smooth the kernels using scalelengths somewhat less than the wavelengths of the seismic waves resolved in the numerical simulation.

There will exist short-scalelength features and fringes in kernels based on more complicated 2D or 3D models, such as the fringes shown in *Tromp et al.* (2005, Figure 9) for the P-SV wavefield or in *Zhou et al.* (2004, Figure 13b). The smoothing operation will tend to remove these sub-resolution features from the kernel. An alternative approach to smoothing the inversion is to add an explicit damping term to the misfit function (e.g., *Akçelik et al.*, 2002, 2003), as outlined in Section 2.10. This approach leads to an additional term in the expression for the gradient, which represents the desire to obtain a smooth model. We prefer to convolve the misfit kernel with a simple Gaussian that represents the resolution of the simulation, and this is the approach we will take in this paper.

2.5.3 Basis functions

As shown in (2.32), the calculation of the gradient of the misfit function requires a choice of model parameterization. Which basis functions should one use? In the classical tomographic example discussed in Section 2.3.3 we used $M = 286$ spherical spline basis functions to parameterize the model (see Figure 2.2). In adjoint tomography, where the wavefields and kernels are represented on discretized grids, we can use the basis functions embedded in the numerical method itself, for example for the SEM we use Lagrange polynomials (*Komatitsch and Tromp*, 1999). This has the advantage that no restrictions are placed on the wavelengths of the model, other than that they need to be resolvable by the waves used in the inversion. This approach increases the number of model parameters dramatically compared to a classical inversion, but because we do not need to invert a Hessian in the adjoint approach this is of no consequence.

Any smooth function $f(\mathbf{x})$, where $\mathbf{x} = (x, y)$, that is sufficiently resolved by the SEM mesh can be expressed in discrete form as

$$f(\mathbf{x}) = \sum_{k=1}^{N_{\text{glob}}} f_k L_k(\mathbf{x}), \tag{2.36}$$

where $k = 1, \ldots, N_{\text{glob}}$ is the index of the N_{glob} global node points, $f_k = f(\mathbf{x}_k)$ is the functional value at global node $\mathbf{x}_k$, and $L_k(\mathbf{x})$ is a global function defined by

$$L_k(\mathbf{x}) = \begin{cases} l_\alpha(\xi(x,y))\, l_\beta(\eta(x,y)) & \text{if } \mathbf{x}_k \in \Omega_e \text{ and } k|_{\Omega_e} = (\alpha, \beta), \\ 0 & \text{if } \mathbf{x}_k \notin \Omega_e. \end{cases} \tag{2.37}$$

Here l_α and l_β are Lagrange polynomials of degree α and β, respectively. We use degree 4 polynomials, i.e., 5 Gauss-Lobatto-Legendre points, in the 2D simulations presented in this paper. The invertible mapping from the reference square with points (ξ, η), with $-1 \leq \xi \leq 1$ and $-1 \leq \eta \leq 1$, to the deformed quadrilateral spectral-element Ω_e with points (x, y) may be written in the form $\xi = \xi(x, y)$, $\eta = \eta(x, y)$ (e.g., *Komatitsch and Vilotte*, 1998; *Komatitsch and Tromp*, 1999). Note that functions $L_k(\mathbf{x})$ corresponding to global grid points $\mathbf{x}_k$ located on the edges or corners of elements have nonzero contributions from all elements that share the global point. At the kth node,

$$L_k(\mathbf{x}_k) = 1, \tag{2.38}$$

in accordance with (2.36).

The functions $L_k(\mathbf{x})$ are orthogonal but not orthonormal. We may obtain a set of orthonormal basis functions $B_k(\mathbf{x})$ based on the definition

$$B_k(\mathbf{x}) = L_k(\mathbf{x})/A_k, \tag{2.39}$$

where A_k is the square-root-area associated with the kth node:

$$A_k^2 = \int_\Omega L_k^2(\mathbf{x})\, \mathrm{d}^2\mathbf{x}. \tag{2.40}$$

The B_k are orthonormal in the sense that

$$\int_\Omega B_k(\mathbf{x}) B_{k'}(\mathbf{x}) \mathrm{d}^2\mathbf{x} = \delta_{kk'}, \tag{2.41}$$

and any function can be expanded in terms of these basis functions. For example, we may expand the misfit kernel $K(\mathbf{x})$ in terms of the basis functions $B_k(\mathbf{x})$ as

$$K(\mathbf{x}) = \sum_{k=1}^{N_{\mathrm{glob}}} \tilde{K_k} B_k(\mathbf{x}). \tag{2.42}$$

The expansion coefficients $\tilde{K_k}$ are determined by

$$\begin{aligned}\tilde{K_k} &= \int_\Omega K(\mathbf{x}) B_k(\mathbf{x}) \, \mathrm{d}^2\mathbf{x} = \int_\Omega \sum_{k'} K_{k'} L_{k'}(\mathbf{x}) B_k(\mathbf{x}) \, \mathrm{d}^2\mathbf{x} = \sum_{k'} K_{k'} A_{k'} \int_\Omega B_{k'}(\mathbf{x}) B_k(\mathbf{x}) \, \mathrm{d}^2\mathbf{x} \\ &= K_k A_k \end{aligned} \tag{2.43}$$

where $K_k = K(\mathbf{x}_k)$ is the value of the misfit kernel at a global grid point, and we have used (2.36) and the orthonormality relation (2.41).

Now let us assume we have computed a misfit kernel $K(\mathbf{x})$. In discrete form, we can write $K(\mathbf{x}_k) = K_k$, since K is defined on the $N_{\mathrm{glob}} = 25921$ global nodes of the SEM mesh. Upon comparing (2.32) with (2.43), we see that, using the basis functions (2.39), the gradient of the misfit function is simply

$$g_k = K_k A_k. \tag{2.44}$$

This provides a trivial step from the discretized kernel to the gradient. Using the $M = N_{\mathrm{glob}}$ basis functions in (2.39), the model parameters (2.13) are therefore

$$\delta m_k = \delta \ln c_k A_k, \tag{2.45}$$

where $\delta \ln c_k$ is the discrete version of $\delta \ln c(\mathbf{x})$.

2.6 Optimization: iterative improvement of the model

In the previous section we showed how to compute the gradient of the misfit function by summing event kernels (Figure 2.9) and subsequently multiplying by the basis functions of the model (2.32). In this section we illustrate how iterative improvements to the model may be determined based on a nonlinear conjugate gradient algorithm (*Fletcher and Reeves*, 1964). We demonstrated this algorithm for a simple source inversion in *Tromp et al.* (2005, Section 8.1). In Section 2.6.2 we consider a 2D tomographic example.

2.6.1 Conjugate gradient algorithm

The algorithm we use may be summarized as follows: given an initial model $\mathbf{m}^0$, calculate $F(\mathbf{m}^0)$, $\mathbf{g}^0 = \partial F/\partial \mathbf{m}(\mathbf{m}^0)$, and set the initial conjugate gradient search direction equal to minus the initial gradient of the misfit function, $\mathbf{p}^0 = -\mathbf{g}^0$. If $||\mathbf{p}^0|| < \epsilon$, where ϵ is a suitably small number, then $\mathbf{m}^0$ is the model we seek to determine, otherwise:

1. Perform a line search to obtain the scalar ν^k that minimizes the function $\tilde{F}^k(\nu)$ where

$$\begin{aligned} \tilde{F}^k(\nu) &= F(\mathbf{m}^k + \nu \mathbf{p}^k) \\ \tilde{g}^k(\nu) &= \frac{\partial \tilde{F}^k}{\partial \nu} = \frac{\partial F}{\partial \mathbf{m}} \Big(\mathbf{m}^k + \nu \mathbf{p}^k \Big) \cdot \mathbf{p}^k \end{aligned}$$

   - Choose a test parameter $\nu_t^k = -2\tilde{F}^k(0)/\tilde{g}^k(0)$, based on quadratic extrapolation.
   - Calculate the test model $\mathbf{m}_t^k = \mathbf{m}^k + \nu_t^k \mathbf{p}^k$.
   - Calculate $F(\mathbf{m}_t^k)$ and, for cubic interpolation, $\mathbf{g}_t^k = \mathbf{g}(\mathbf{m}_t^k)$.
   - Interpolate the function $\tilde{F}^k(\nu)$ by a quadratic or cubic polynomial and obtain the ν^k that gives the (analytical) minimum value of this polynomial.

2. Update the model: $\mathbf{m}^{k+1} = \mathbf{m}^k + \nu^k \mathbf{p}^k$, then calculate $\mathbf{g}^{k+1} = \partial F/\partial \mathbf{m}(\mathbf{m}^{k+1})$.

3. Update the conjugate gradient search direction: $\mathbf{p}^{k+1} = -\mathbf{g}^{k+1} + \beta^{k+1}\mathbf{p}^k$, where $\beta^{k+1} = \mathbf{g}^{k+1} \cdot (\mathbf{g}^{k+1} - \mathbf{g}^k)/(\mathbf{g}^k \cdot \mathbf{g}^k)$.

4. If $||\mathbf{p}^{k+1}|| < \epsilon$, then $\mathbf{m}^{k+1}$ is the desired model; otherwise replace k with $k+1$ and restart from 1.

A key decision is the choice of the test parameter, ν_t^k, which determines how far one should go in the direction of the search direction (initially the negative gradient) to obtain the test model. We assume a quadratic form of the misfit function and determine ν_t^k based on this assumption. Computation of $F(\mathbf{m})$ (misfit) and $\mathbf{g}(\mathbf{m})$ (misfit kernel) are expensive in the tomographic problem, and thus we must limit the number of computations as much as possible. Some of these aspects are addressed in Section 2.6.3.

2.6.2 2D tomographic example

Using (2.6), we can define the average traveltime anomaly for a particular model:

$$\overline{\Delta T} = \sqrt{2\,F(\mathbf{m})/N}. \tag{2.46}$$

This gives some physical meaning to the F-values in the plots in this section. Figure 2.11a–i shows one cycle of the conjugate gradient algorithm for the 2D tomographic example. Part (a) shows the phase-speed model used to generate the data (the "target" model), and (b) shows the initial phase-speed model, $\mathbf{m}^0$, used to generate the initial synthetics. The phase speed of the initial model is $c = 3.50$ km/s, the period of the source in the simulations is $\tau = 20$ s, and thus the reference wavelength is approximately $\lambda = c\tau = 70$ km.

Figure 2.11c shows the (smoothed) gradient for this model. The gradient is represented by the slope $\tilde{g}^0(0)$ of a line passing through $[0,\ \tilde{F}(0)]$ (Figure 2.11d). Quadratic extrapolation with a parabolic minimum at $(\nu_t^0,\ 0)$ gives the ν-value for a new test model (Figure 2.11d, Section 2.11). Figure 2.11e shows the test model, $\mathbf{m}_t^0$, for which we compute the gradient via the process shown in Figures 2.7–2.10, but now the model is no longer homogeneous. The gradient, shown in Figure 2.11f, is then depicted as the slope of a line passing through $[\nu_t^0,\ \tilde{F}(\nu_t^0)]$ (Figure 2.11g).

Next, in Figure 2.11g we approximate $\tilde{F}^0(\nu)$ by a cubic polynomial, $P^0(\nu)$, passing through two points, $[0,\ \tilde{F}^0(0)]$ and $[\nu_t^0,\ \tilde{F}(\nu_t^0)]$, and having slopes at these points corresponding to the respective gradients. In other words, six values are needed to obtain an analytical minimum of the cubic function: the two models (represented by $\nu = 0$ and ν_t^0), the misfits of these models, and the derivatives at these points (see Section 2.11). The minimum, $[\nu^0,\ P^0(\nu^0)]$, indicates the expected value of the misfit for the updated model given by $\mathbf{m}^1 = \mathbf{m}^0 - \nu^0\mathbf{g}^0$, which is shown in Figure 2.11h and represented by the point

($\nu^0 = 1.2 \times 10^4$, 0) in Figure 2.11g. Figure 2.11i shows the decrease in the misfit function going from $\mathbf{m}^0$ to $\mathbf{m}^1$. The dashed curve is determined based on the nine iteration points in Figure 2.12.

Figure 2.11 thus constitutes one iteration of the conjugate gradient algorithm. The process is repeated, and the results are shown in Figure 2.12. Each iteration produces a model that looks qualitatively more similar to the target model shown in Figure 2.11a, and generates a lower value of the misfit function (2.6). We draw a best-fitting hyperbola to the $\log_{10}$ values to highlight the convergence.

We next use the seismologically more interesting Rayleigh wave phase-speed model in Figure 2.1. In comparison with Figure 2.11a, this model has variable scalelength and lower amplitude perturbations. The weaker perturbations result in a lower initial misfit, $F(\mathbf{m}^0) = 1182.0$ s^2. Figure 2.13 shows the recovery of an interior portion of the model, where path coverage is good. The basic features in the target phase-speed model (Figure 2.20a) are recovered by the third iteration (Figure 2.13d). The two sets of points in the Figure 2.13f are discussed in the next section. The model obtained after the first iteration, model $\mathbf{m}^1$ shown in Figure 2.13b, looks very similar to the model obtained based on a classical Hessian-based inversion with heavy damping shown in Figure 2.4g. This reflects the fact that in the conjugate gradient approach one is effectively working with an initial approximation to the Hessian that is the identity matrix.

2.6.3 Variations on the conjugate gradient algorithm

Based on the conjugate gradient algorithm outlined in Section 2.6.1, we require $4N_{\text{events}}$ numerical simulations for each iterative improvement of the model: synthetics for $\mathbf{m}^0$, the gradient for $\mathbf{m}^0$, synthetics for test model $\mathbf{m}_t^0$, and the gradient for $\mathbf{m}_t^0$. This information is used to compute the analytical minimum for a cubic polynomial. An alternative approach is to perform $3N_{\text{events}}$ numerical simulations per iteration by neglecting the gradient of the test model and using a quadratic polynomial to compute an analytic minimum (Section 2.11)

A comparison of these two approaches is shown in Figure 2.20f. The initial model for both cases has a misfit of $F(\mathbf{m}^0) = 1182.0$ s^2 for $N = 3300$ seismograms, corresponding to an average traveltime anomaly of $\overline{\Delta T}(\mathbf{m}^0) = 0.85$ s (eq. 2.46). Using $4N_{\text{events}}$ simulations with a cubic polynomial, we obtain a negligible advantage in terms of a better convergence of $F(\mathbf{m})$: for example, $F(\mathbf{m}^8_{\text{cubic}}) = 3.52$ s^2 whereas $F(\mathbf{m}^8_{\text{quad}}) = 3.75$ s^2 (Figure 2.20). To

the eye, the recovered models $\mathbf{m}^k_{\text{cubic}}$ and $\mathbf{m}^k_{\text{quad}}$ are indistinguishable.

An additional part of the conjugate gradient algorithm that can be adjusted is the selection of the test model, which we discuss in Section 2.11. Finally, we note that entrapment into local minima is common in the conjugate gradient method, as addressed in *Akçelik et al.* (2002, 2003). Such local minima may be avoided by using multiscale methods (*Bunks et al.*, 1995). Alternatively, by starting at longer periods, which constrain the long wavelength heterogeneity, and gradually moving to shorter periods, which constrain smaller scale structures, one can also try to avoid local minima.

2.7 Tomographic experiments

The greater the number of events used in the inversion, the better the recovery of the model. Figure 2.14 shows the recovery of the model in Figure 2.15i using 5, 15, and 25 sources, respectively.

Figure 2.15 examines the effect of the smoothing parameter, Γ, on the recovery of three different phase-speed models, each having a scalelength of structural heterogeneity that is proportional to the reference wavelength. Using a smaller Γ we resolve shorter-scalelength structures, whether they are in the target phase-speed model or not. When the scalelength of the smoothing exceeds that of the structure ($\Gamma > \Lambda$), the structure is smoothed out, as expected (Figure 2.15l).

The introduction of random errors into the cross-correlation traveltime measurements, ΔT_i, has essentially no impact on model recovery in our examples. For example, we denote a 50% error in the measurements by $\Delta T_i' = \Delta T_i\,(r + 0.5)$, where $r \in [0, 1]$ is a random number, ΔT_i is the "actual" measurement, and $\Delta T_i'$ is the randomized measurement used in the inversion. In terms of the adjoint method, the introduction of random errors has the effect of changing the amplitude of the various banana-doughnut kernels that comprise the event kernel. Because the coverage in this example is very good, several similar kernels are "stacked" in constructing the event kernel, and thus the random errors effectively cancel.

2.8 Source, structure, and joint inversions

The traveltime differences between data and synthetics may be due to an inaccurate structural model, inaccurate source models, or some combination of both. In this section we

illustrate the simultaneous inversion for structural and source parameters using adjoint methods and the conjugate gradient algorithm. We first describe and illustrate the basic source inversion and then address the joint inversion.

2.8.1 Basic source inversion

A perturbation of the point source (2.30) may be written in the form

$$\delta f(x,y,t) = -\dot{h}(t)\delta t_s\,\delta(x-x_s)\,\delta(y-y_s) + h(t)(\delta x_s\,\partial_{x_s} + \delta y_s\,\partial_{y_s})[\delta(x-x_s)\,\delta(y-y_s)], \quad (2.47)$$

where δt_s denotes a perturbation in the origin time, $(\delta x_s, \delta y_s)$ a perturbation in the source location, and $\dot{h}(t) = \partial h/\partial t = -\partial h/\partial t_s$.

Based on the theory outlined in *Tromp et al.* (2005, Section 8), a change in the traveltime misfit function (2.7) due to a change in the point source is given by

$$\delta F = \int_0^T \int_\Omega \delta f(x,y,t) s^\dagger(x,y,T-t)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t, \quad (2.48)$$

where $s^\dagger$ denotes the adjoint wavefield, whose sources are time-reversed, measurement-weighted seismograms, injected at the receivers, just as in the case of the structure inversions (2.33). (Here the traveltime measurement is affected by source perturbations only.) Upon substituting (2.47) we obtain

$$\delta F = -\,\delta t_s \int_0^T \dot{h}(t)\; s^\dagger(x_s,y_s,T-t)\,\mathrm{d}t + (\delta x_s\,\partial_{x_s} + \delta y_s\,\partial_{y_s}) \int_0^T h(t)\; s^\dagger(x_s,y_s,T-t)\,\mathrm{d}t. \quad (2.49)$$

We may express (2.49) in terms of the gradient as $\delta F = \mathbf{g}\cdot\delta\mathbf{m}$, where

$$\mathbf{m} = \begin{bmatrix} (x_s^k - x_s^0)/\lambda \\ (y_s^k - y_s^0)/\lambda \\ (t_s^k - t_s^0)/\tau \end{bmatrix}, \quad (2.50)$$

$$\mathbf{g} = \begin{bmatrix} \lambda \int_0^T h(t)\; \partial_{x_s} s^\dagger(x_s,y_s,T-t)\,\mathrm{d}t \\ \lambda \int_0^T h(t)\; \partial_{y_s} s^\dagger(x_s,y_s,T-t)\,\mathrm{d}t \\ -\tau \int_0^T \dot{h}(t)\; s^\dagger(x_s,y_s,T-t)\,\mathrm{d}t \end{bmatrix}. \quad (2.51)$$

Here $\mathbf{m}$ is a three-parameter nondimensionalized model vector describing the source. The source origin time t_s is scaled by the reference period τ, and the source coordinates are

scaled by the reference wavelength $\lambda = c\,\tau$, where c is the reference phase speed. The gradient vector, $\mathbf{g}$, depends on the model $\mathbf{m}$ through the adjoint wavefield $s^\dagger$: by perturbing the source, the measurement between data and synthetics changes, and thus the adjoint wavefield changes correspondingly.

In the experiments in Section 2.6, the sources for the data and synthetics were identical, whereas the structure was not. We now consider the effects of source perturbations, where the point sources for the initial synthetics are mislocated and initiate at an incorrect time.

Tromp et al. (2005, Fig. 12) demonstrated a two-parameter source inversion based on an adjoint method and the conjugate gradient algorithm. In that example, the two source parameters described the magnitude vector of the point source. In Figure 2.16 we illustrate a three-parameter source inversion for $\delta\mathbf{m} = (\delta x_s,\, \delta y_s,\, \delta t_s)$. The structural models for the data and synthetics are identical. We use the adjoint method to compute the gradient (2.51) of the misfit function (2.6). Using the conjugate gradient algorithm, we recover the source by the third iteration.

Finally, we emphasize that all of the equations in this section apply generally for any measurement, for example waveforms or amplitudes. The computed values for the expressions will differ, however, because the adjoint source $f^\dagger$ (and the corresponding adjoint wavefield $s^\dagger$) will vary for each measurement.

2.8.2 Joint inversions

In a joint inversion the sources and structure are initially different from the "target" sources and structure, and we seek to determine both. If we consider the three-parameter source inversion in Figure 2.16, then the model vector for the joint inversion is $\delta\mathbf{m} = [\delta\mathbf{m}_{\rm str}\,;\,\delta\mathbf{m}_{\rm src}]$ with dimension $N_{\rm structure} + 3N_{\rm event}$. The misfit function is given by (2.6). We adjust the gradient of the misfit function at each iteration according to a constant, J, computed from the initial gradient:

$$\mathbf{g}^k = \left[J\mathbf{g}^k_{\rm str}\,;\,\mathbf{g}^k_{\rm src}\right], \tag{2.52}$$

denoting a concatenation of the structure gradient $\mathbf{g}_{\text{str}}$ computed via (2.32) and the source gradient $\mathbf{g}_{\text{src}}$ computed via (2.51). The scaling factor J is given by

$$J = \|\mathbf{g}^0_{\text{src}}\|_2 \;/\; \|\mathbf{g}^0_{\text{str}}\|_2, \tag{2.53}$$

where $\|\cdot\|_2$ denotes the L2-norm of the enclosed vector. The motivation behind (2.52) is that we want the source parameters and structural parameters to have about the same contribution in the gradient in the conjugate gradient algorithm. The exact choice of J, e.g., L1- versus L2-norm, is not important. Note that the factor J is chosen once and for all based on the *initial* structural and source gradients. Also, the gradients ($\mathbf{g}^k$, $\mathbf{g}^k_{\text{str}}$, $\mathbf{g}^k_{\text{src}}$) are all with respect to the misfit function (2.6), evaluated at model $\mathbf{m}^k$.

Figure 2.17 compares a basic source inversion with a joint inversion. In the joint inversion the initial structural model is homogeneous, and the sources are mislocated randomly within 5 km of the target source and have an inaccurate origin time within the range ± 1 s. The two misfit curves in Figure 2.17d show that the joint inversion does almost as well as the basic structure inversion; in fact, it lags the misfit by only one or two iterations. In the final model only the sources on the edges of the grid contain location and timing errors (Figure 2.19c to Figure 2.17f), which is expected since there are few, if any, paths to constrain the structure.

Figure 2.18 shows the recovery of a single source during the joint inversion. It takes approximately 16 iterations to fully recover the source (instead of the 3 iterations in Figure 2.16 for the basic source inversion), although most of the source location is still recovered in the first few iterations. This increase is, of course, due to the gradual improvement of the structural parameters, which initially differ by up to 10% from the target structure.

In an inversion with real data, the initial model is bound to be deficient both in terms of structure and sources. Thus, a joint inversion is a logical approach. Figure 2.19 shows the consequences of neglecting either source or structure in the inversion. In Figure 2.19a–c we invert for structure and assume that the sources are accurate, when in fact they are perturbed as shown in (c). The misfit curve in (a) shows that the conjugate gradient approach appears to be working: the misfit decreases as the structure iterates to updated models. However, it is clear that Figure 2.19b does not represent the true structure, since we know the target model we are trying to recover, as well as its associated misfit curve for the basic structure inversion. This illustrates how (fixed) errors in the source parameters are

mapped into errors in the structural parameters. Figure 2.19d–f shows the opposite scenario: the structure is fixed and assumed to be accurate, and we allow the source parameters to be perturbed to reduce the traveltime misfit. The source parameters adjust themselves from Figure 2.19c (initial) to Figure 2.19f (final) while reducing the misfit.

2.9 Discussion

2.9.1 Three types of sensitivity kernels

We have designated three types of sensitivity kernels: "banana-doughnut kernels," "event kernels," and "misfit kernels." A banana-doughnut kernel (e.g., *Marquering et al.*, 1999) is a phase-specific (e.g., P) kernel for an individual source-receiver combination; for our purposes, the key point is that this kernel does not incorporate the measurement. Alternative names include "finite-frequency," "Born," and "sensitivity" kernel. An event kernel can be thought of as a sum of individual banana-doughnut kernels, such that each kernel in the sum is weighted by its corresponding measurement. Using the adjoint approach, however, the event kernel is not computed by summing banana-doughnut kernels, but rather in one single simulation through the interaction between the forward wavefield and an adjoint wavefield generated by simultaneous fictitious sources for all available arrivals at all available stations (Section 2.5.1). A misfit kernel is simply the sum of event kernels, and may be thought of as a graphical representation of the gradient of the misfit function. In classical tomography, the banana-doughnut kernels are used to compute the gradient and (approximate) Hessian of the misfit function for the Newton approach to the inverse problem. In adjoint tomography, only the misfit kernels are used in the inverse problem.

2.9.2 Classical tomography versus adjoint tomography

In this paper, "classical tomography" refers to Hessian-based inversions, whereby the Hessian is constructed from individual source-receiver paths, either in terms of rays or finite frequency kernels. The Hessian matrix, with a damping parameter γ, can be inverted to obtain structural models. We compute the traveltime anomalies, and thus F, via (2.6), and then compare these values with those obtained from gradient-derived models.

Figure 2.20 shows a comparison among models produced using classical tomography, $\mathbf{m}^{\text{ray}}$ (ray-based inversion) and $\mathbf{m}^{\text{ker}}$ (kernel-based inversion), and the model produced using

adjoint tomography, $\mathbf{m}^{16}$ (16 conjugate gradient iterations). All three models are only subtly different from the target model used to generate the data (Figure 2.20a). The misfit for each approach is summarized in Figure 2.20e. The misfit values for the classical models, $F(\mathbf{m}^{\text{ray}}) = 5.26$ s^2 and $F(\mathbf{m}^{\text{ker}}) = 4.90$ s^2, correspond to average traveltime anomalies of $\overline{\Delta T}(\mathbf{m}^{\text{ray}}) = 0.056$ s and $\overline{\Delta T}(\mathbf{m}^{\text{ker}}) = 0.055$ s (eq. 2.46), indicating that each recovered model explains almost all of the traveltime differences between a homogeneous model and the target model in Figure 2.1b. Two points regarding the two F-values are important: (1) they are essentially the same, which is expected, since the Hessian used in each inversion is very similar (Figure 2.3a); (2) they are met by the conjugate gradient approach by the seventh conjugate gradient iteration. In other words, $F(\mathbf{m}^{\text{ray}}) \approx F(\mathbf{m}^{\text{ker}}) \approx F(\mathbf{m}^{7})$; after seven conjugate gradient iterations, we recover a model equivalent to what could be recovered by having the (ray- or kernel-based) Hessian. It is important to note that $\mathbf{m}^{\text{ray}}$, $\mathbf{m}^{\text{ker}}$, and $\mathbf{m}^{1}$ are based on the homogeneous reference model $\mathbf{m}^{0}$, but for $k > 1$, the adjoint tomography models $\mathbf{m}^{k}$ are based on heterogeneous models.

Figure 2.20 might suggest that classical tomography "does pretty well" in comparison with adjoint tomography. This is more or less true for the simple examples in this paper. However, seismic tomography is transitioning from simple 1D reference models to fully 3D reference models. The calculation of a Hessian for 3D reference models is generally not an option, and thus one must resort to iterative, gradient-based algorithms. The results in this paper illustrate that for the problems considered here, such iterative techniques work quite efficiently and converge quickly.

The main advantages of the adjoint tomography approach are fivefold. First, all the complexities that are considered in the forward problem (e.g., *Komatitsch and Tromp*, 2002a,b) can be considered in the inversion. For example, in this paper we have shown finite-frequency sensitivity kernels based on heterogeneous models. But one could also consider fully anisotropic Earth models with 21 elastic parameters for essentially the same numerical cost as an isotropic simulation involving just two parameters. Second, the style of tomography — traveltime, amplitude, waveform — is determined by the choice of the misfit function (*Tromp et al.*, 2005). Given the choice of measurement, one simply determines the associated adjoint source that gives rise to the corresponding kernel. Third, any time segment where the data and synthetics match reasonably well is suitable for a measurement. One does not need to label a particular phase, like P or SS, because the adjoint simulation

will reveal how this particular measurement (or time window) "sees" the Earth model, and the resulting sensitivity kernel will reflect this view. Fourth, the model parameterization is trivial (2.43) and requires only a conservative level of smoothing to remove numerical artifacts in the kernels near the sources and receivers (Section 2.5.2). Furthermore, structure can only be introduced in regions where the kernel (or gradient) is nonzero. This is in contrast to classical tomography, where both the selection of basis functions and the choice of damping involve a certain degree of undesired subjectivity. Finally, the approach scales linearly with the number of earthquakes but is independent of the number of receivers and the number of arrivals that are used in the inversion.

With southern California in mind, say we have $N_{\text{events}} = 150$ earthquakes, $N_{\text{receivers}} = 150$ SCSN stations, $N_{\text{comp}} = 3$ components per seismogram, and $N_{\text{picks}} = 4$ time-windowed measurements per component, for a total of $N_{\text{events}} N_{\text{receivers}} N_{\text{comp}} N_{\text{picks}} = 270,000$ measurements. An adjoint approach would require $2N_{\text{events}} = 300$ simulations to compute one misfit kernel. A complete 7-iteration conjugate gradient inversion based on cubic interpolation would require $7 \times 3N_{\text{events}} = 3150$ total simulations. By comparison, a Hessian-based inversion would require individual kernels for the 270,000 measurements, which, for 3D models, is neither computationally feasible nor practical.

2.9.3 Feasibility of 3D–3D tomography

This paper is a step toward "3D–3D tomography," denoting 3D heterogeneity within the reference models and a 3D physical domain for the model, from which we compute finite-frequency sensitivity kernels. (Based on this labeling, the classical tomographic examples in this paper are 0D–2D, whereas the adjoint tomographic examples are 2D–2D.) Presently our SEM codes are set up to compute 3D–3D sensitivity kernels on both regional and global scales (*Liu and Tromp*, 2006, 2008). In this paper we have highlighted some aspects of the inversion process that will be key to limiting the number of wavefield simulations required in the inversion.

Let us estimate the computational cost of a regional-scale tomographic inversion. As discussed, for 150 earthquakes we require 3150 total simulations for a 7-iteration inversion. Each simulation takes approximately 35 minutes on 72 nodes (144 processors). Thus we can perform 40 runs per day on 72 nodes, and more than 500 runs per day on a 1000-node machine. Therefore, on this kind of hardware the whole inversion can theoretically be

completed in about one week.

To avoid reaching a local minimum in the optimization procedure, we intend to start by using longer-period waveforms, which existing 3D models fit reasonably well, and work our way toward shorter periods. As we improve the model and increase the frequency contents of the waveforms, we expect to not only improve the fit to the current data used to constrain the model, but also to steadily increase the number of picks that is used in the inversion, i.e., more and more parts of the seismograms are expected to be used and matched in the iterative inversion process. Unlike *Akçelik et al.* (2003), our emphasis will be on matching targeted, frequency-dependent body-wave traveltimes and surface-wave phase anomalies, rather than entire waveforms. Waveform tomography is largely controlled by amplitude differences, which are notoriously difficult to fit in seismology. Traveltime or phase, on the other hand, is a robust measure of misfit that has been used for decades to constrain local, regional and global Earth models. From our perspective, the progression from ray-based traveltime tomography to finite-frequency "banana-doughnut" tomography to frequency-dependent adjoint tomography is very natural.

2.10 Appendix A: Regularization

Here we review the fact that stabilizing the Hessian matrix (as in eq. 2.5) via damping is equivalent to adding a regularization term R to the misfit function (2.6):

$$F_R(\mathbf{m}) = F(\mathbf{m}) + R(\mathbf{m}), \tag{2.54}$$

whose gradient is, using (2.20),

$$\frac{\partial F_R}{\partial m_k} = \frac{\partial F}{\partial m_k} + \frac{\partial R}{\partial m_k} = -\sum_{i=1}^{N} G_{ik}\,\Delta T_i + \frac{\partial R}{\partial m_k}. \tag{2.55}$$

There are many options for regularization. For illustrative purposes, we consider regularization according to the wave-speed model itself:

$$R = \tfrac{1}{2}\gamma^2 \int_V (\delta \ln c)^2 \,\mathrm{d}^3\mathbf{x}, \tag{2.56}$$

where γ is the damping parameter. Substituting (2.13), and then differentiating with respect

to the kth model parameter, we obtain

$$R = \tfrac{1}{2}\gamma^2 \sum_{k=1}^{M} \delta m_k \sum_{k'=1}^{M} \delta m_{k'} D_{kk'}, \tag{2.57}$$

$$\frac{\partial R}{\partial m_k} = \gamma^2 \sum_{k'=1}^{M} \delta m_{k'} D_{kk'}, \tag{2.58}$$

where the $M \times M$ damping matrix $\mathbf{D}$ is given by

$$D_{kk'} = \int_V B_k \, B_{k'} \, \mathrm{d}^3\mathbf{x}. \tag{2.59}$$

If the basis functions are orthonormal, then $\mathbf{D} = \mathbf{I}$, the identity matrix. Substituting (2.58) into (2.55), we obtain

$$\frac{\partial F_R}{\partial \mathbf{m}} = -\mathbf{G}^T\mathbf{d} + \gamma^2 \mathbf{D}\delta\mathbf{m}, \tag{2.60}$$

where $\mathbf{D}$ is, for example, (2.59) or (2.63). Substituting this for $\mathbf{g}(\mathbf{m})$ into (2.5), with $\mathbf{H} = \mathbf{G}^T\mathbf{G}$, we obtain

$$\left(\mathbf{G}^T\mathbf{G} + \gamma^2\mathbf{D}\right)\delta\mathbf{m} = \mathbf{G}^T\mathbf{d}, \tag{2.61}$$

which leads to (2.28). Equation (2.61) is known as "Tikhonov" regularization or "ridge regression," and is based on minimizing an L2-norm measure of $\mathbf{D}\,\delta\mathbf{m}$ (e.g., *Hansen*, 1998, Ch. 5). (Typically these two labels refer to the case $\mathbf{D} = \mathbf{I}$.)

Instead, if we regularize using the gradient of the wave-speed model (e.g., *Akçelik et al.*, 2003), we obtain

$$R = \tfrac{1}{2}\gamma^2 \int_V \boldsymbol{\nabla}(\delta \ln c) \cdot \boldsymbol{\nabla}(\delta \ln c) \, \mathrm{d}^3\mathbf{x}, \tag{2.62}$$

then the damping matrix is

$$D_{kk'} = \int_V (\boldsymbol{\nabla} B_k) \cdot (\boldsymbol{\nabla} B_{k'}) \, \mathrm{d}^3\mathbf{x}. \tag{2.63}$$

Regularization according to the *roughness* of the model (e.g., *Zhou et al.*, 2005) leads to

$$R = \tfrac{1}{2}\gamma^2 \int_V \left(\nabla^2 \delta \ln c\right)^2 \, \mathrm{d}^3\mathbf{x}, \tag{2.64}$$

$$D_{kk'} = \int_V \left(\nabla^2 B_k\right)\left(\nabla^2 B_{k'}\right) \mathrm{d}^3\mathbf{x}. \tag{2.65}$$

Different norms or constraints, as well as combinations of constraints (resulting in multiple damping parameters), may be used in forming $R(\mathbf{m})$. For example, *Akçelik et al.* (2002) advocated the use of L1-based, "total variation" regularization, which avoids smoothing of sharp gradients in material properties. *Akçelik et al.* (2003) applied L1 regularization for the structure gradients and L2 regularization for the source gradients.

2.11 Appendix B: Details of the conjugate gradient algorithm

The computation of the misfit value $F(\mathbf{m})$ and gradient $\mathbf{g}$ takes $2N_{\text{events}}$ simulations. Because each simulation is expensive, it is important to limit the number of simulations in the inverse problem, which we approach using a conjugate gradient algorithm (Section 2.6.1). Two possible areas to aid in this are the selection of the trial step ν_t and the choice of the polynomial (quadratic or cubic) to use in the interpolation. In this section we have omitted the superscript k on quantities to avoid clutter.

2.11.1 Selection of the trial step

The trial step, or test parameter, ν_t, determines how far away from the current model to go in the search direction in order to obtain a test model (and, possibly, test gradient). Given the misfit value, $F(\mathbf{m})$, and the gradient, $\mathbf{g}(\mathbf{m})$, for the current model, the user is faced with determining how far to step in the search direction away from the current model to obtain a test model, for which an additional misfit value will be computed. The gradient vector $\mathbf{g}$ is represented in the conjugate gradient algorithm as a slope, $\tilde{g}(0)$, and the misfit function in the search direction by $\tilde{F}(\nu)$. In the algorithm, we select the test parameter by interpolating $\tilde{F}(\nu)$ using a quadratic polynomial, $Q(\nu)$:

$$Q(\nu) = a\nu^2 + b\nu + c, \tag{2.66}$$

where a, b, and c are determined using the value (r_1) and slope (g_1) for the current model, and a test model location such that the value (r_2) and slope (g_2) of $Q(\nu)$ at ν_t are both zero (see Figure 2.11d). The four values are given by

$$\begin{aligned} r_1 &\equiv Q(0) = \tilde{F}(0) = F(\mathbf{m}), \\ g_1 &\equiv Q'(0) = \tilde{g}(0), \\ r_2 &\equiv Q(\nu_t) = 0, \\ g_2 &\equiv Q'(\nu_t) = 0. \end{aligned}$$

These equations can be used to determine the coefficients of $Q(\nu)$:

$$\begin{aligned} a &= -g_1/(2\nu_t) = g_1^2/(4r_1), \\ b &= g_1, \\ c &= r_1, \end{aligned}$$

as well as the test parameter

$$\nu_t = \frac{-2r_1}{g_1} = \frac{-2F(\mathbf{m})}{\tilde{g}(0)}, \tag{2.67}$$

which is the value used in the algorithm discussed in Section 2.6.1.

The "test model parabola" $Q(\nu)$ is chosen such that its vertex lies on $F = 0$; however, one could choose the vertex at some $F > 0$ that is determined based on the change in misfit from a previous step. The quadratic extrapolation through $F = 0$ is perhaps too conservative, and computational savings — in the form of better convergence — could be obtained by exploring the choice of the initial step.

2.11.2 Quadratic versus cubic interpolation

As discussed in Section 2.6.3, the tomographer is faced with the choice of using a second- or third-order polynomial, $P(\nu)$, in the interpolation scheme within the conjugate gradient algorithm. Here we outline the formulas required to compute an analytical minimum, $\nu_{\min}$, using each interpolation scheme.

With an order-3 polynomial, one must have four quantities in addition to the test parameter ν_t: the misfit and gradient for the current model, $F(\mathbf{m})$ and $\mathbf{g}$, and the misfit

and gradient for the test model, $F(\mathbf{m}_t)$ and $\mathbf{g}_t$. These values are converted into scalar values for an interpolating polynomial $P(\nu)$:

$$\begin{aligned}
\nu_1 &= 0, \\
\nu_2 &= \nu_t, \\
p_1 &\equiv P(\nu_1) = F(\mathbf{m}), \\
g_1 &\equiv P'(\nu_1) = \tilde{g}(0), \\
p_2 &\equiv P(\nu_2) = F(\mathbf{m}_t), \\
g_2 &\equiv P'(\nu_2) = \tilde{g}(\nu_t).
\end{aligned}$$

The cubic polynomial can be written in terms of these quantities as

$$P(\nu) = a\,(\nu - \nu_1)^3 + b\,(\nu - \nu_1)^2 + c\,(\nu - \nu_1) + d, \tag{2.68}$$

where

$$\begin{aligned}
a &= [-2(p_2 - p_1) + (g_1 + g_2)(\nu_2 - \nu_1)] \;/\; (\nu_2 - \nu_1)^3\,, \\
b &= [3(p_2 - p_1) - (2g_1 + g_2)(\nu_2 - \nu_1)] \;/\; (\nu_2 - \nu_1)^2\,, \\
c &= g_1, \\
d &= p_1.
\end{aligned}$$

An analytical minimum of $P(\nu)$ can be obtained when $|c| > 0$:

$$\nu_{\min} = \begin{cases} \nu_1 + [-b + (b^2 - 3ac)^{1/2}] \,/\, (3a) & a \neq 0 \text{ and } b^2 - 3ac > 0 \\ -c/(2b) & a = 0 \text{ and } b \neq 0;\ b^2 - 3ac < 0. \end{cases} \tag{2.69}$$

With an order-2 polynomial, the gradient of the test model — $\mathbf{g}_t$ or g_2 — is not required. In this case, the quadratic polynomial can be written in terms of (2.68) as

$$P(\nu) = a\,(\nu - \nu_1)^2 + b\,(\nu - \nu_1) + c, \tag{2.70}$$

where

$$\begin{aligned} a &= [(p_2 - p_1) - g_1(\nu_2 - \nu_1)] \;/\; \left(\nu_2^2 - \nu_1^2\right), \\ b &= g_1, \\ c &= p_1 - a\nu_1^2 - b\nu_1. \end{aligned}$$

The analytical minimum of $P(\nu)$ is simply

$$\nu_{\min} = -b/(2a). \tag{2.71}$$

Based on our experiments, the quadratic interpolation is preferred over the cubic interpolation, since it costs $3N_{\text{events}}$ per iteration (versus $4N_{\text{events}}$) and performs only slightly worse (Figure 2.13f).

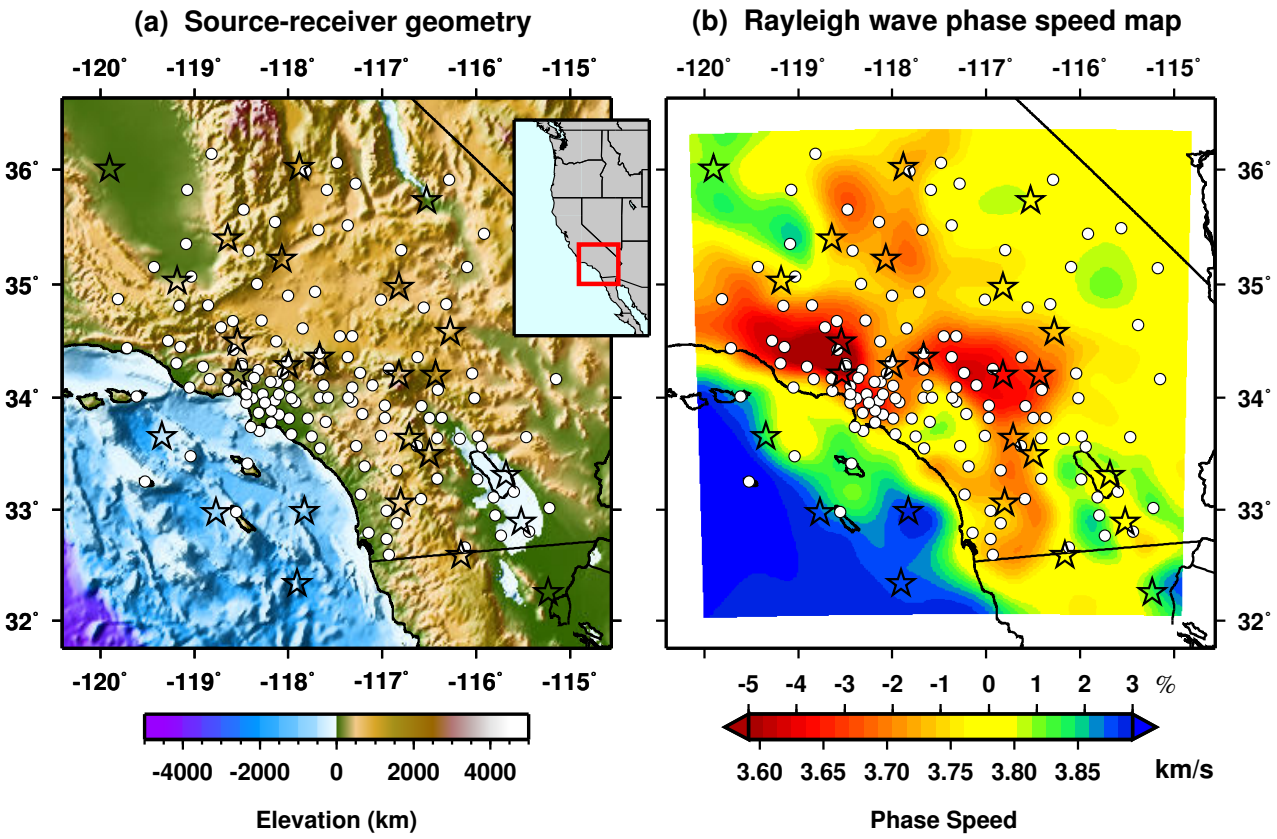


Figure 2.1: Source–receiver geometry for the numerical experiments in this study. The ☆ symbols denote the locations of 25 earthquakes (each has a $M \geq 4.0$ and occurred between 1990 and 2005); the ∘ symbols denote the locations of 132 broadband receivers in the Southern California Seismic Network (SCSN). The earthquakes are selected to obtain relatively uniform coverage; all SCSN receivers in the area are included. (a) Topography and bathymetry in the region. (b) Phase speed map for 20 s Rayleigh waves, based on the regional model of *Hauksson* (2000), modified with the Moho map of *Zhu and Kanamori* (2000). This phase speed map is used to generate synthetic data used in some of the inversion experiments.

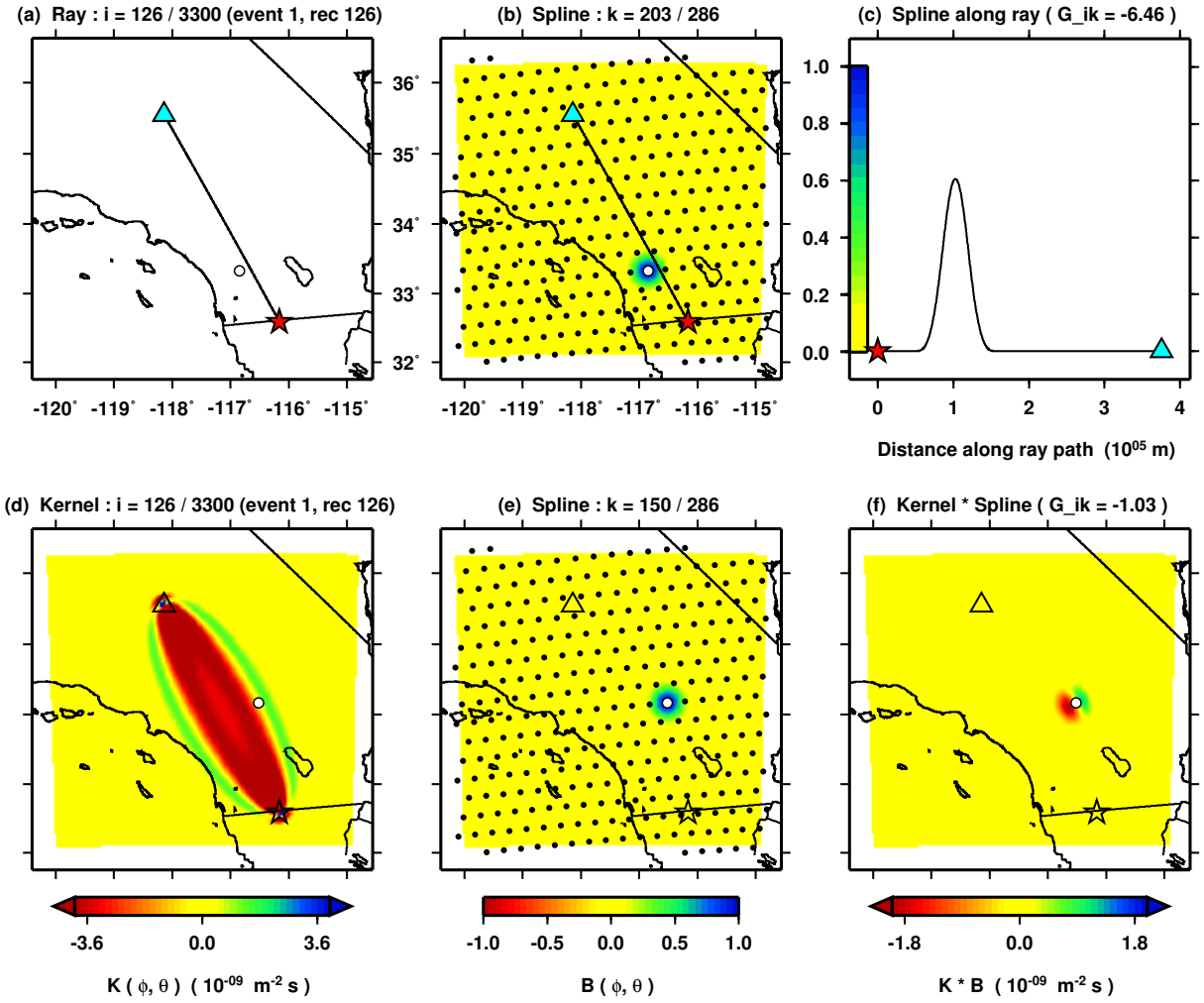


Figure 2.2: Example computation for an element, G_{ik}, of the design matrix $\mathbf{G}$, using rays (a-c) and finite-frequency kernels (d-f). The row index i is the source-receiver combination, the column index k is the basis function index. The source is denoted by the ☆, the receiver is denoted by the △, and the ∘ shows the center-point of the spherical spline in (b) or (e). (a) Ray path for event number 1 and receiver number 126 (Figure 2.1), corresponding to the $i = 126$ index of the $N = 3300$ ray paths. (b) $B_{203}(\mathbf{x})$, the spherical spline basis function for index $k = 203$. Also shown are the center-points of the $M = 286$ spherical splines. (c) Spline B_{203} evaluated along the ray path. The value of the phase speed for the reference model is constant, so $G_{ik} = (-1/c) \int_{\text{ray}_i} B_k \, \mathrm{d}s$ (eq. 2.15). In this example $G_{ik} = -1/(3780\,\mathrm{m\,s^{-1}})\ (2.45 \times 10^4\,\mathrm{m}) = -6.46$ s. (d) Cross-correlation traveltime sensitivity kernel for event number 1 and receiver number 126 (Figure 2.1), corresponding to the $i = 126$ index of the $N = 3300$ kernels. (e) $B_{150}(\mathbf{x})$, the spherical spline basis function for index $k = 150$. Also shown are the center-points of the $M = 286$ spherical splines. (f) The function $K_{126}(\mathbf{x})\, B_{150}(\mathbf{x})$. The integral of this function gives the value $G_{ik} = \int_\Omega K_{126}\, B_{150}\, \mathrm{d}^2\mathbf{x} = -1.03$ s. (See Section 2.2.)

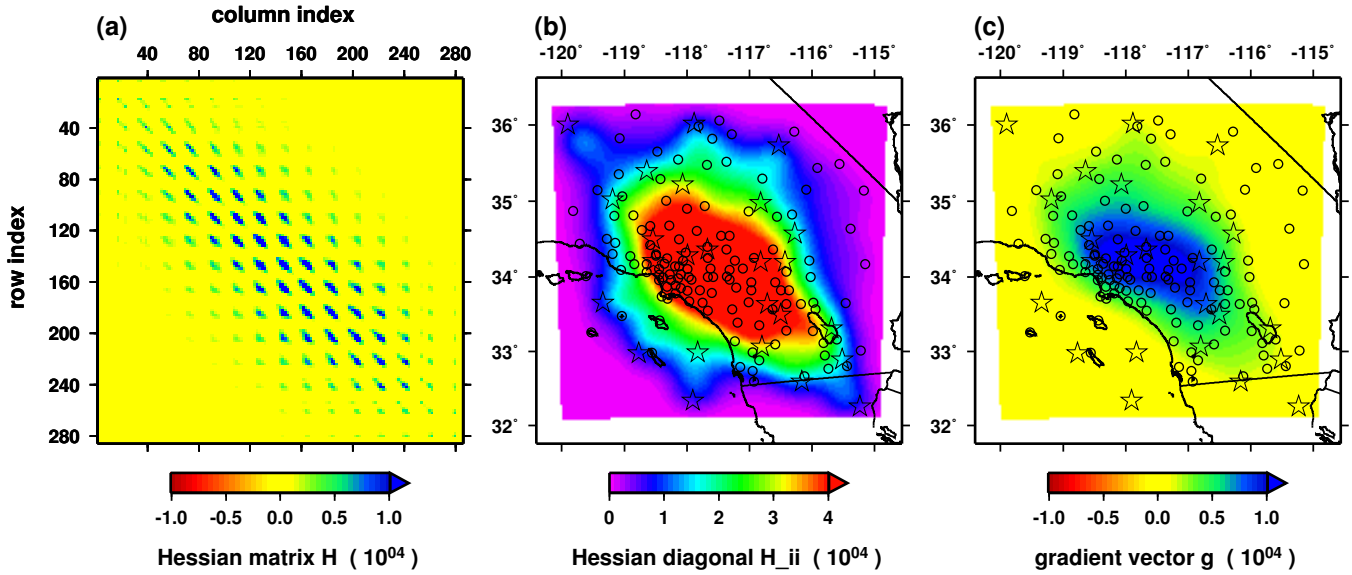


Figure 2.3: The Hessian matrix and gradient vector for a classical tomography inversion. (a) The Hessian matrix, $\tilde{\mathbf{H}} = \mathbf{G}^T\mathbf{G}$, for the source-receiver geometry shown in Figure 2.1, using finite-frequency kernels based on cross-correlation traveltime measurements. Each element of $\mathbf{G}$ is constructed as shown in Figure 2.2d-f. The Hessian matrix computed using rays, as shown in Figure 2.2a-c, is indistinguishable from the one shown in this figure. In practice, a damping matrix is added to the Hessian to stabilize the inversion (2.27). (b) The diagonal elements of the Hessian matrix, $\tilde{H}_{ii}$, expanded in the spherical spline basis functions to illuminate the spatial pattern (e.g., *Zhou et al.*, 2005, Figure 3). This map is a proxy for spatial coverage and depends on the source-receiver geometry, the basis functions, and the sensitivity kernels. (c) The gradient vector, $\mathbf{g} = -\mathbf{G}^T\mathbf{d}$, expanded in the spherical spline basis functions, whereby $\mathbf{d}$ includes cross-correlation traveltime measurements between data computed for the target phase-speed model in Figure 2.1b and synthetics computed for a homogeneous phase-speed model ($c = 3.78$ km/s). The ☆ symbols denote the sources, and the ∘ symbols denote the receivers.

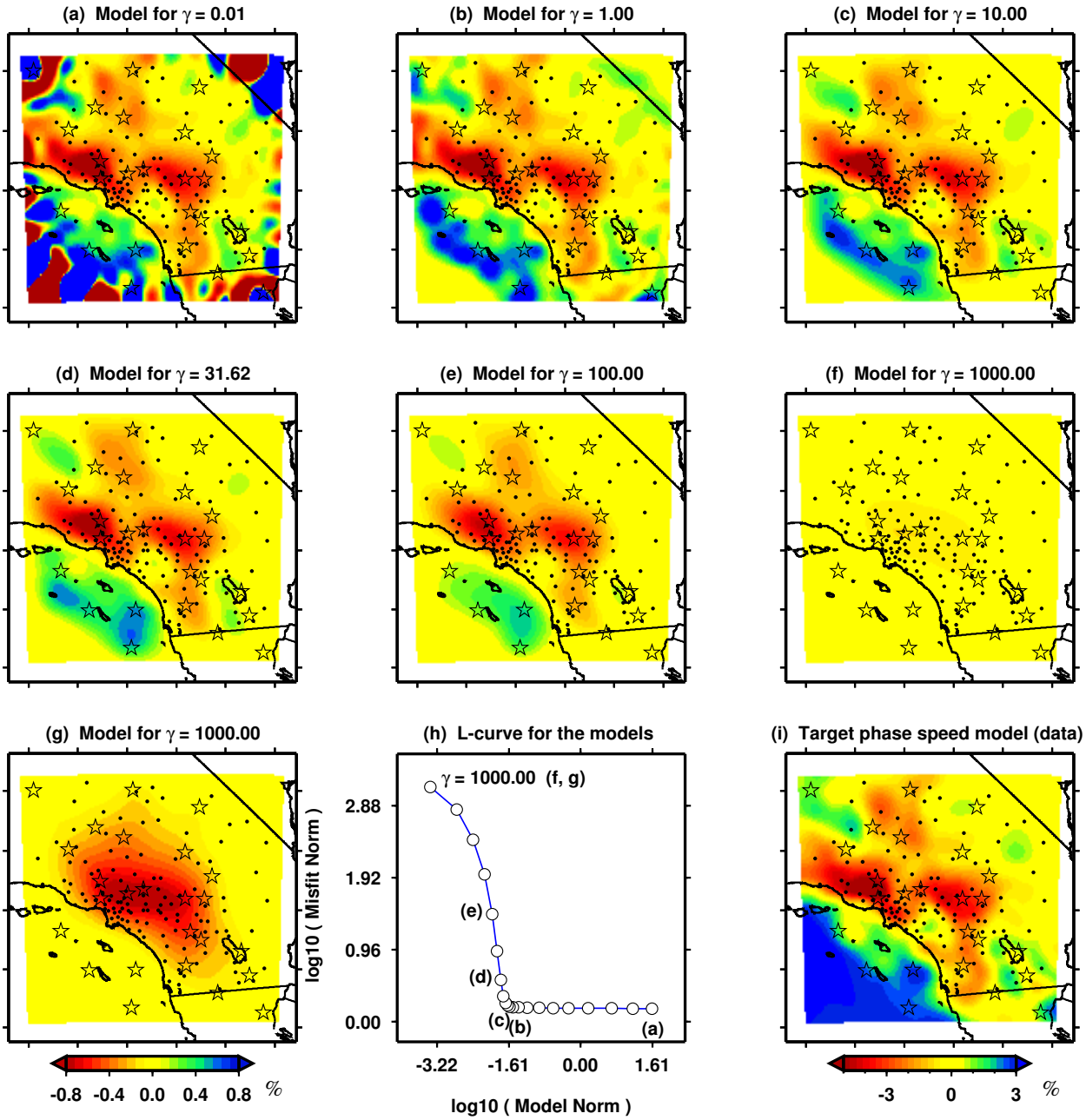


Figure 2.4: Model recovery and damping in classical tomography, illustrated for an inversion using 3300 banana-doughnut kernels. Each model is computed via $\mathbf{m} = -\tilde{\mathbf{H}}_\gamma^{-1}\mathbf{g}$, where $\tilde{\mathbf{H}}_\gamma = \mathbf{G}^T\mathbf{G} + \gamma^2\mathbf{I}$ is the Hessian matrix with damping parameter γ, and $\mathbf{g} = -\mathbf{G}^T\mathbf{d}$ is the gradient vector (Figure 2.3c). The undamped Hessian matrix, $\tilde{\mathbf{H}}_0$, is shown in Figure 2.3a-b. (a)–(f) Recovered phase-speed models for various values of γ. The color scale for each model is shown below (i). (g) Same as (f), but with a more saturated color scale to show its resemblance to the gradient (Figure 2.3c). (h) L-curve illustrating the trade-off between misfit norm and model norm, that is, $\|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2$ versus $\|\mathbf{m}\|_2$. Note that this measure of misfit is not the same as $\mathbf{d}^T\mathbf{d}$, the misfit function in (2.6). The γ values for the model-points are spaced by uniform $\log_{10}$ increments. (i) Target phase-speed model used to generate the data (Figure 2.1b). The ☆ symbols denote the sources, and the • symbols denote the receivers. See Section 2.3.3.

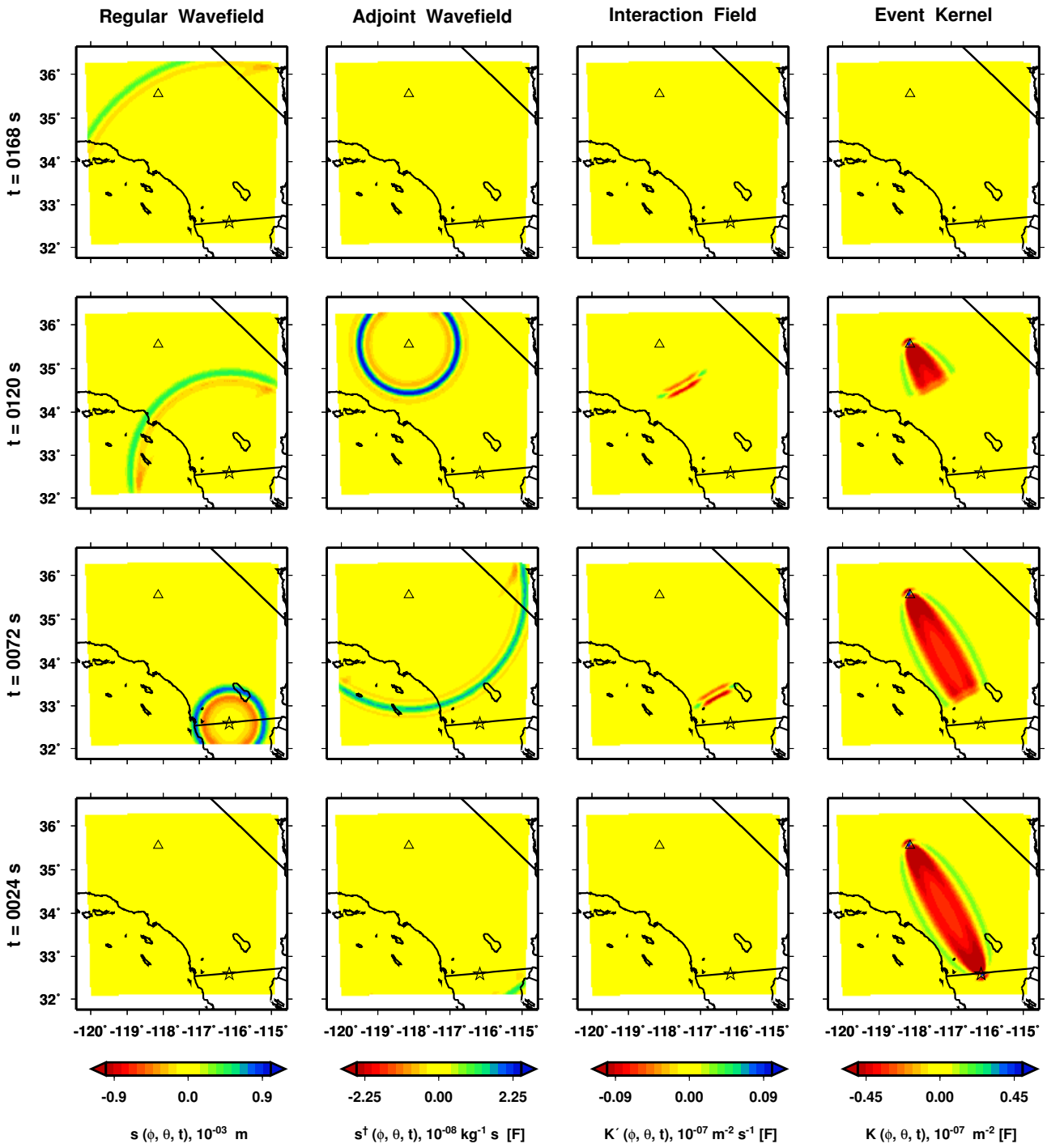


Figure 2.5: Sequence of interactions between the regular and adjoint wavefields during the construction of a traveltime cross-correlation event kernel $K(\mathbf{x})$. The ☆ symbol denotes the source, and the △ symbol denotes the receiver. Each row represents the time-step indicated on the left. In this case, with only a single receiver and a uniform model perturbation, the event kernel resembles a banana-doughnut kernel $K_i(\mathbf{x})$. The event kernel is constructed via the interaction between the forward wavefield (first column) and the adjoint wavefield (second column). The interaction field (third column) is the instantaneous product of the two wavefields, which is integrated to form the event kernel (fourth column). The event kernel shows the region of the current model that gives rise to the discrepancy between the data and the synthetics. The regular source function and adjoint source function are shown in Figure 2.6.

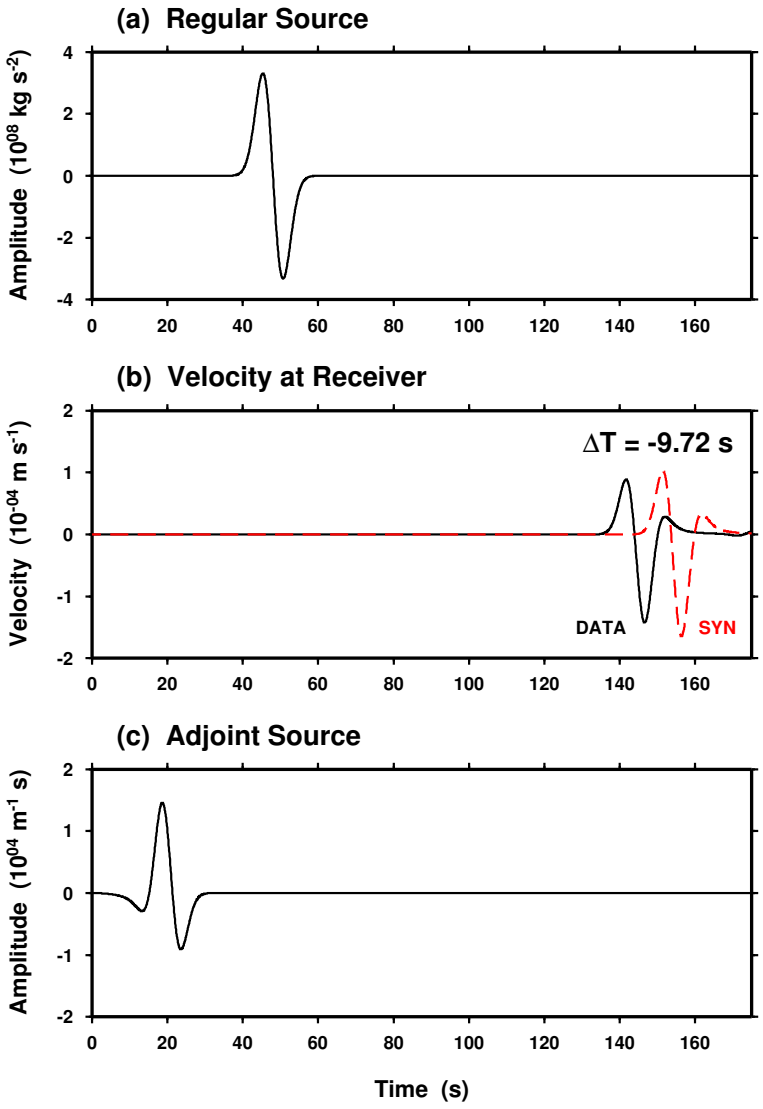


Figure 2.6: Construction of an adjoint source function used in calculating the membrane surface-wave event kernel in Figure 2.5. The traveltime sign convention is shown in (2.8), such that $\Delta T < 0$ represents a delay of the synthetics with respect to the data. The duration of the simulation is $T = 175$ s.

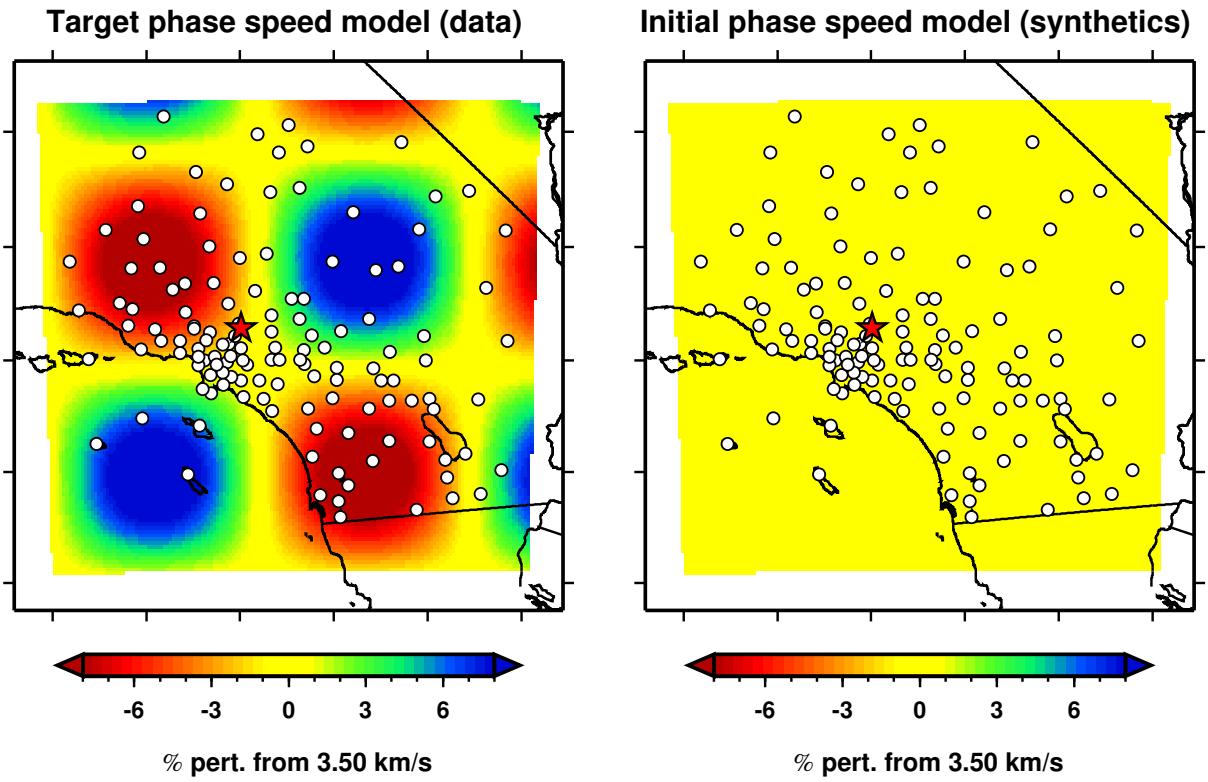


Figure 2.7: Experimental setup for the event kernel shown in Figure 2.8. The data are computed using the target phase-speed model, and the synthetics are computed using the initial model. The minimum and maximum percent perturbations in the target model are $\pm 10\%$. The red star is the event location, and the circles denote the 132 receivers. For plotting purposes, the gridpoints are converted to longitude-latitude points, which results in the nonrectangular appearance of the boundary of the grid.

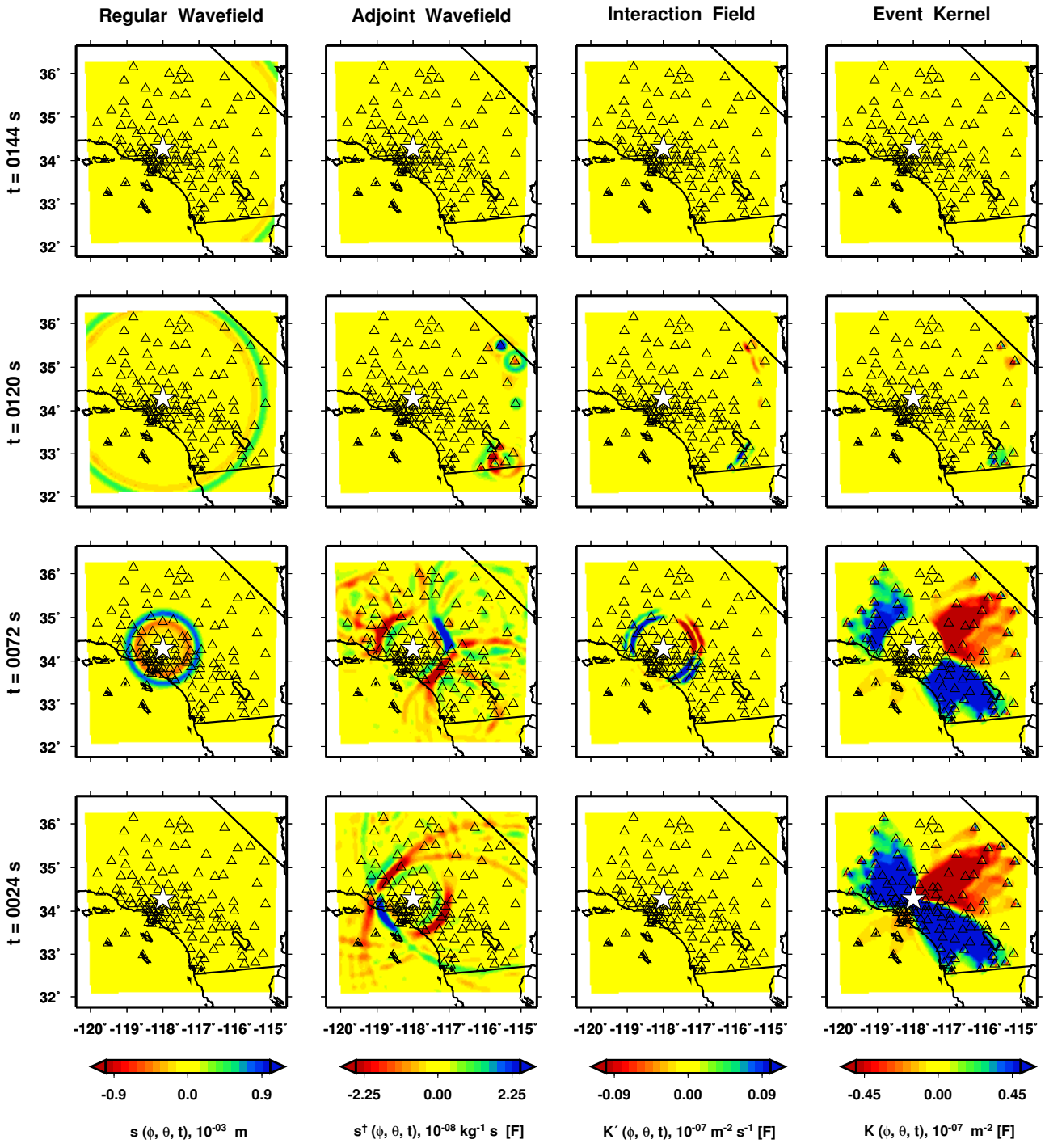


Figure 2.8: Formation of an event kernel for multiple receivers. The phase-speed models used to generate the data and synthetics are shown in Figure 2.7. See Figure 2.5 for details. In comparison with Figure 2.5, here the event is in a different location, there are 132 receivers instead of one, and the data are generated from a checkerboard model, not a uniformly perturbed model.

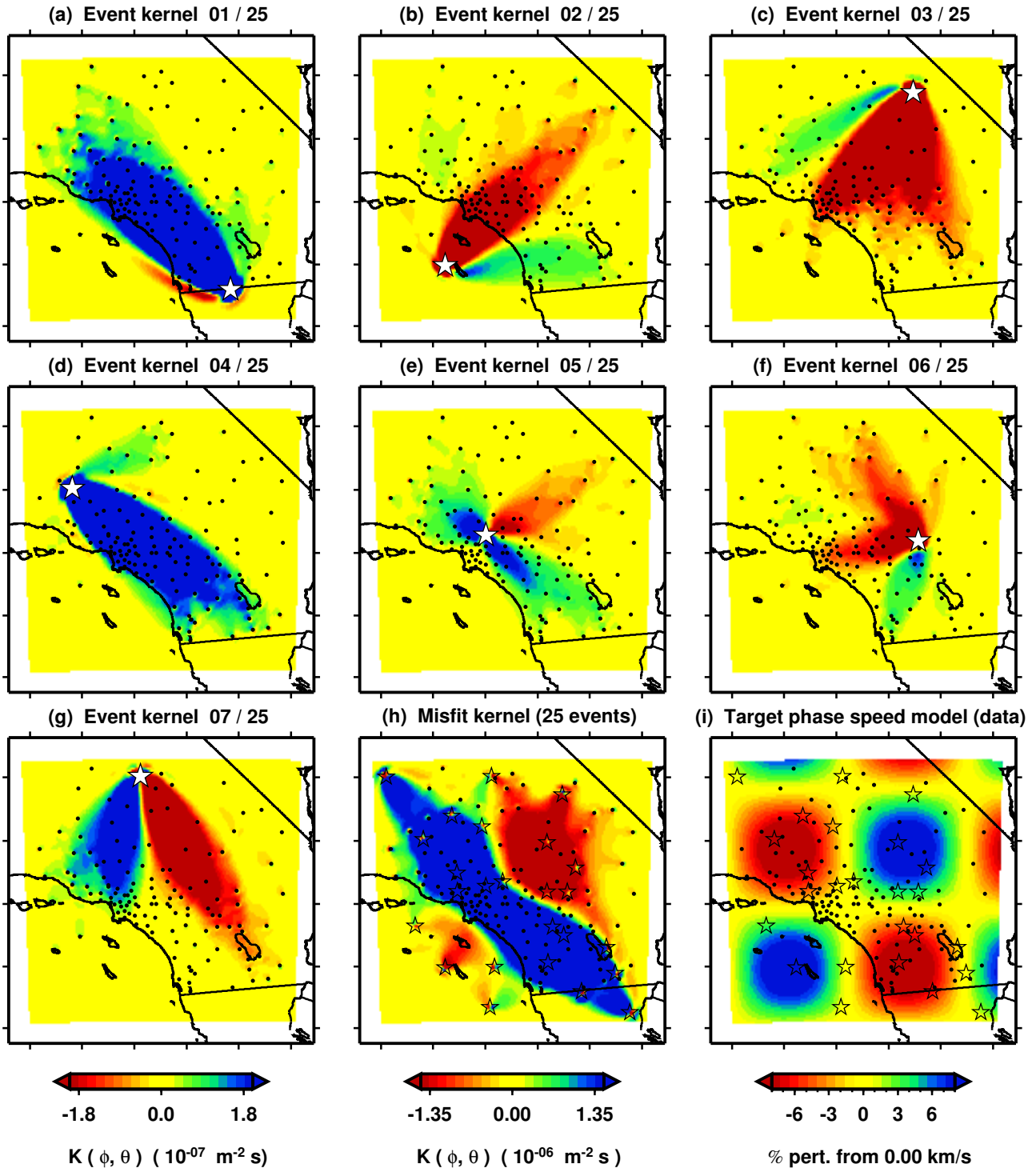


Figure 2.9: Construction of a misfit kernel. (a)–(g) Individual event kernels, each constructed via the method shown in Figure 2.8 (which shows Event 5). The color scale for each event kernel is shown beneath (g). (h) The misfit kernel is simply the sum of the 25 event kernels. (i) The source-receiver geometry and target phase-speed model. There are a total of $N = 25 \times 132 = 3300$ measurements that are used in constructing the misfit kernel. (See Section 2.5.)

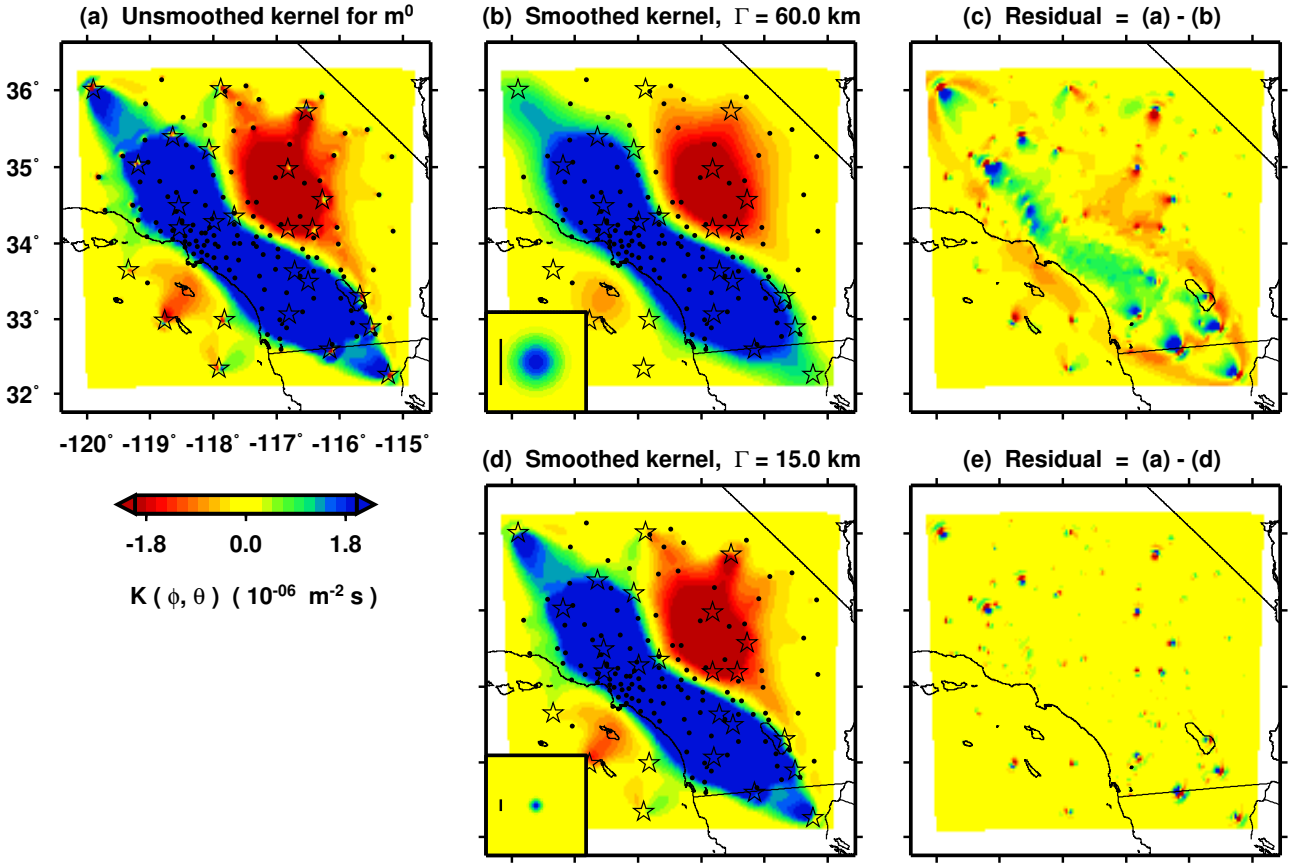


Figure 2.10: Smoothing the misfit kernel. (a) Unsmoothed misfit kernel (Fig. 2.9h). (b)–(c) Smoothed misfit kernel, with residual, obtained via convolution of a Gaussian function (bottom left inset) with (a). The parameter $\Gamma = 60$ km controls the width of the Gaussian and, thus, the degree of smoothing; its value is plotted as a line next to the Gaussian. (d)–(e) Same as (b)–(c), but for less smoothing ($\Gamma = 15$ km). Note that the source and receiver labels are not plotted in the residual plots. (See Section 2.5.2.)

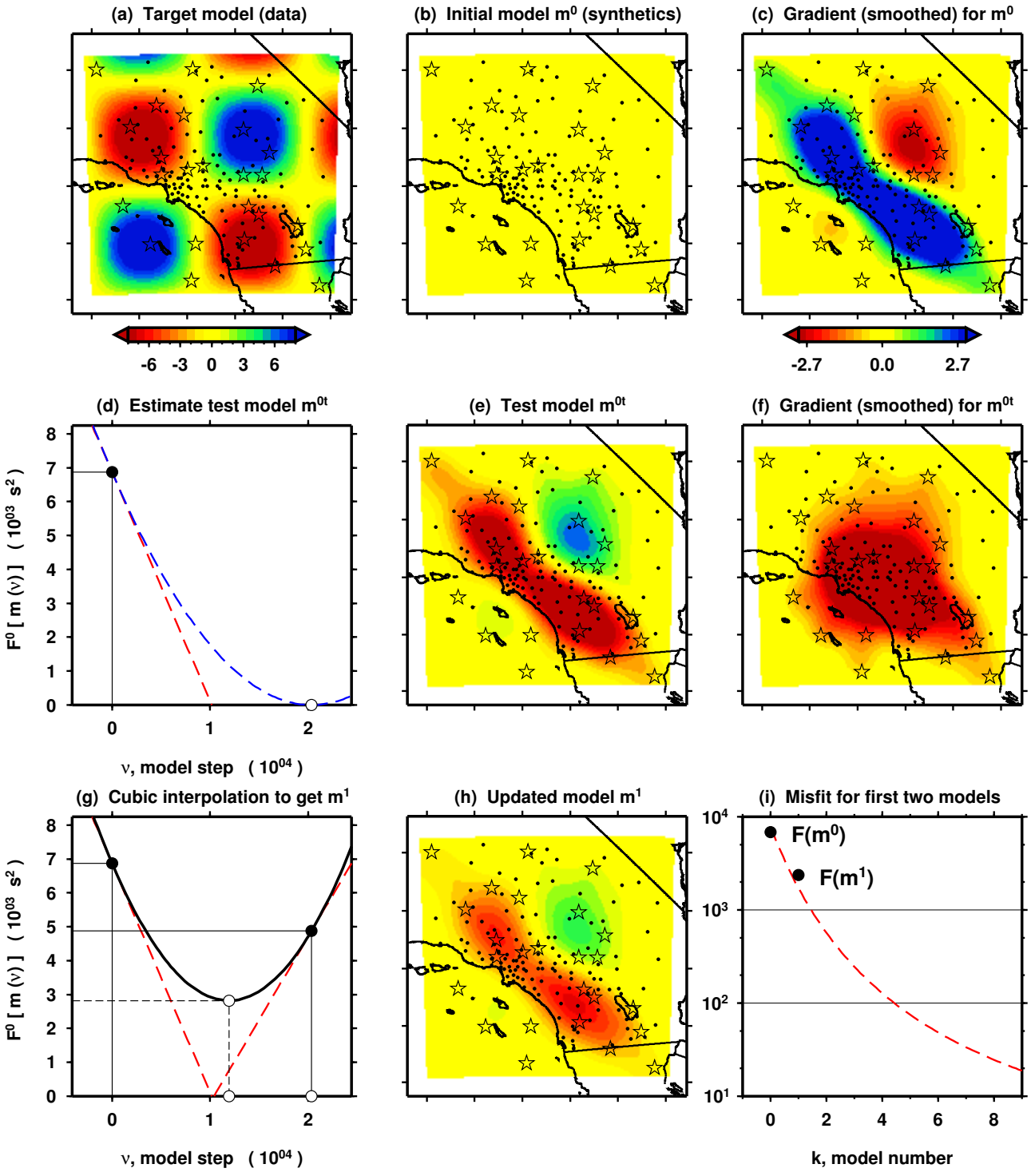


Figure 2.11: (Caption on following page)

Figure 2.11 (*previous page*): The conjugate gradient algorithm applied to a 2D tomographic example. An extended explanation can be found in Section 2.6.2. The algorithm is repeated to obtain the models in Figure 2.12. (a) Target phase-speed model used to generate the data. (b) Phase-speed model used to generate the initial synthetics. The period of the source is $\tau \approx 20$ s, the reference wave speed is $c = 3.50$ km/s, and thus the reference wavelength is $\lambda \approx 70$ km. (c) Misfit kernel — corresponding to the gradient of the misfit function — constructed as illustrated in Figures 2.8–2.10, with smoothing parameter $\Gamma = 30$ km. This kernel highlights the regions of model (b) that need to be improved to reduce the misfit between data and synthetics. (d) Representation of the misfit of the initial model (b) and the initial gradient (c) in the conjugate gradient algorithm. The misfit is denoted by the $\bullet$, and the gradient is denoted by the red dashed line. The white circle indicates the test model obtained by quadratic extrapolation of the gradient through $F = 0$. (e) Test model $\mathbf{m}_t^0$ corresponding to the white circle in (d). (f) Gradient associated with the test model in (e). (g) Cubic interpolation of two misfit values, $F(\mathbf{m}^0)$ and $F(\mathbf{m}_t^0)$, and two gradients, shown in (c) and (f). The analytical minimum provides ν^0, the distance away from $\mathbf{m}^0$ (b) in the direction of (c) that is taken to obtain the first updated model, $\mathbf{m}^1$. (h) First updated model, $\mathbf{m}^1$, corresponding to the white circles in (g) with $\nu^0 = 1.24 \times 10^4$. (i) Misfit values for the first two models. The red dashed curve is taken from Figure 2.12f.

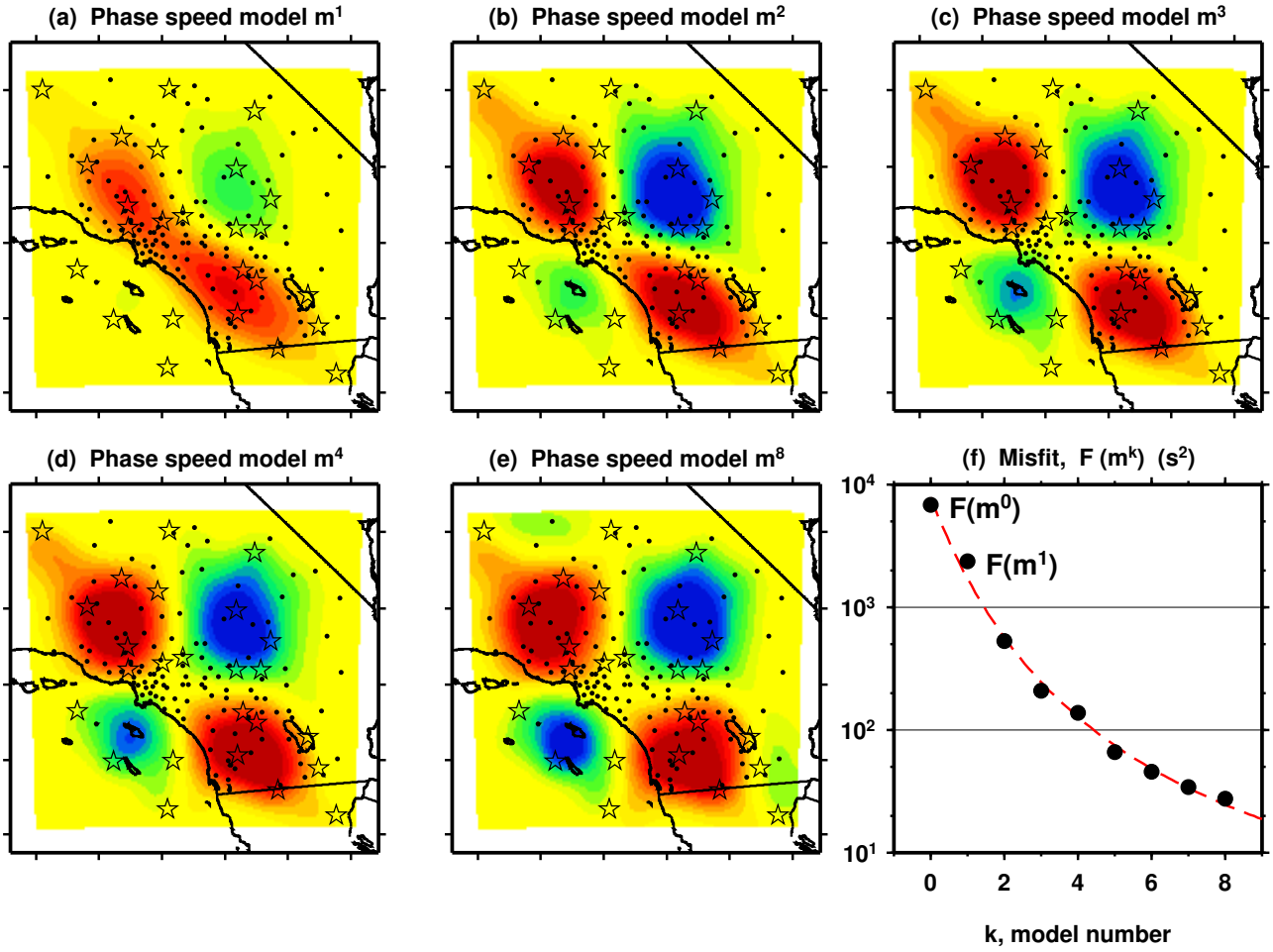


Figure 2.12: Iterative improvement of the reference phase-speed model using the conjugate gradient algorithm illustrated in Figure 2.11. An extended explanation can be found in Section 2.6.2. The first iteration in Figure 2.11 produces $\mathbf{m}^1$ (a), which becomes the current model, from which we obtain $\mathbf{m}^2$ (b), and so on. The red dashed hyperbolic curve in (f) is drawn to accentuate the reduction in misfit.

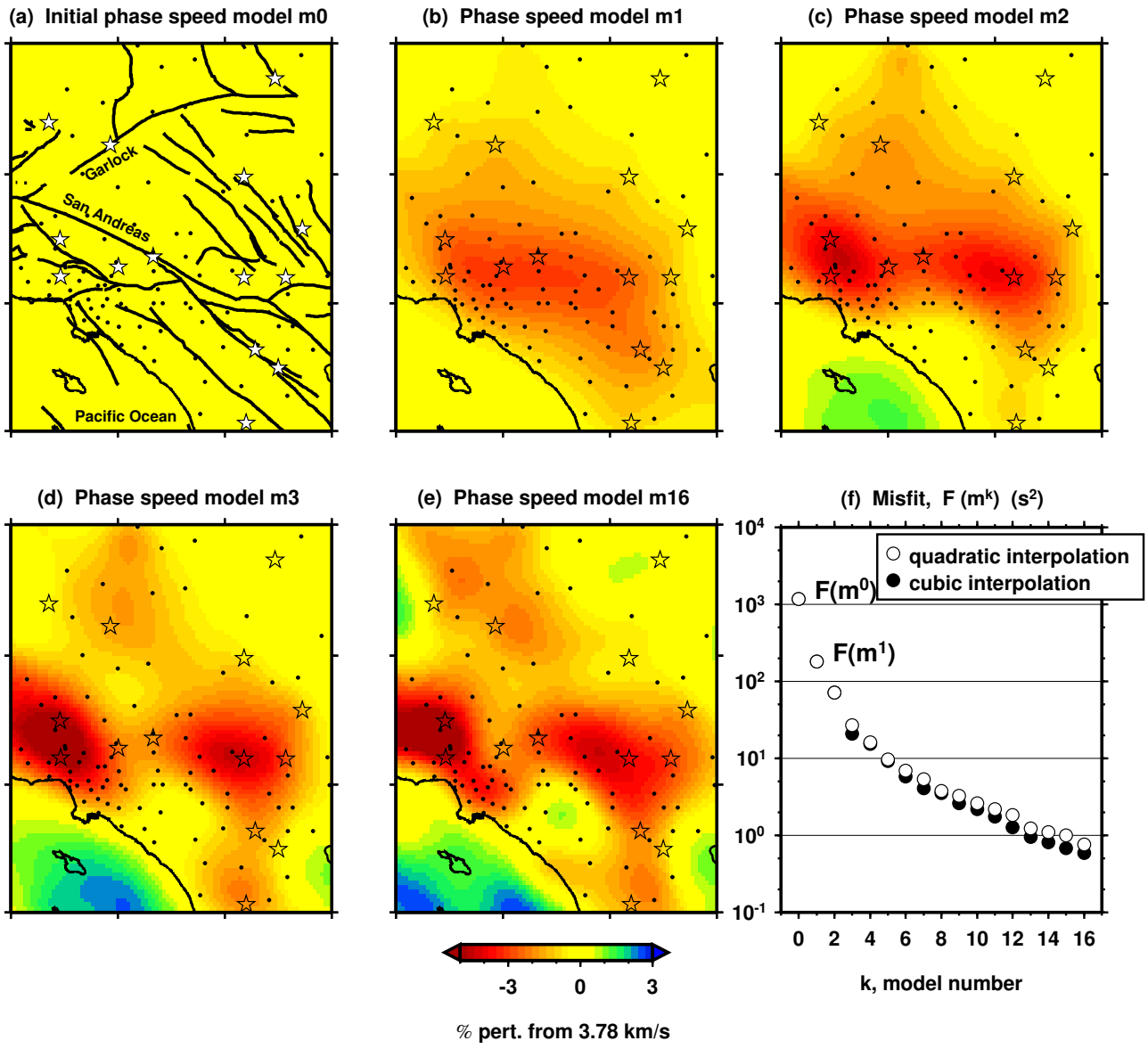


Figure 2.13: Adjoint tomography recovery of a Rayleigh wave phase-speed model (Figure 2.20a). Here we show an interior portion of the southern California region with sufficient path coverage. The color scale for each model is shown below (e). (a) Initial phase-speed model $\mathbf{m}^0$. Faults of *Jennings* (1994) are drawn only for scale. (b)–(e) Iterations $\mathbf{m}^1$, $\mathbf{m}^2$, $\mathbf{m}^3$, and $\mathbf{m}^{16}$. (f) Reduction in the misfit function (2.6) using cubic interpolation (•) versus quadratic interpolation (○) in the conjugate gradient algorithm (Section 2.6.3).

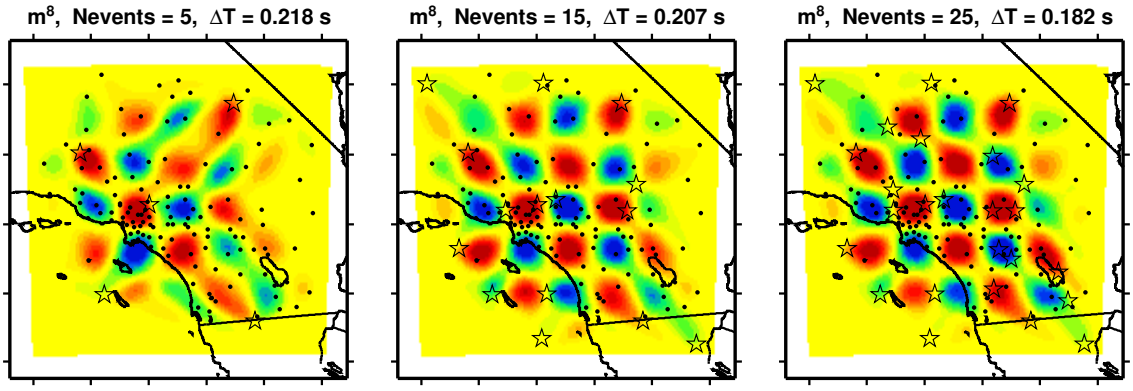


Figure 2.14: Effect of the number of events on the recovery of the phase-speed models. Data are generated from the phase-speed model in Figure 2.15i. The average traveltime anomaly, $\overline{\Delta T}$, is computed from the misfit function value, $F(\mathbf{m}^8)$, using (2.46). As expected, $\overline{\Delta T}$ decreases as we increase the number of events.

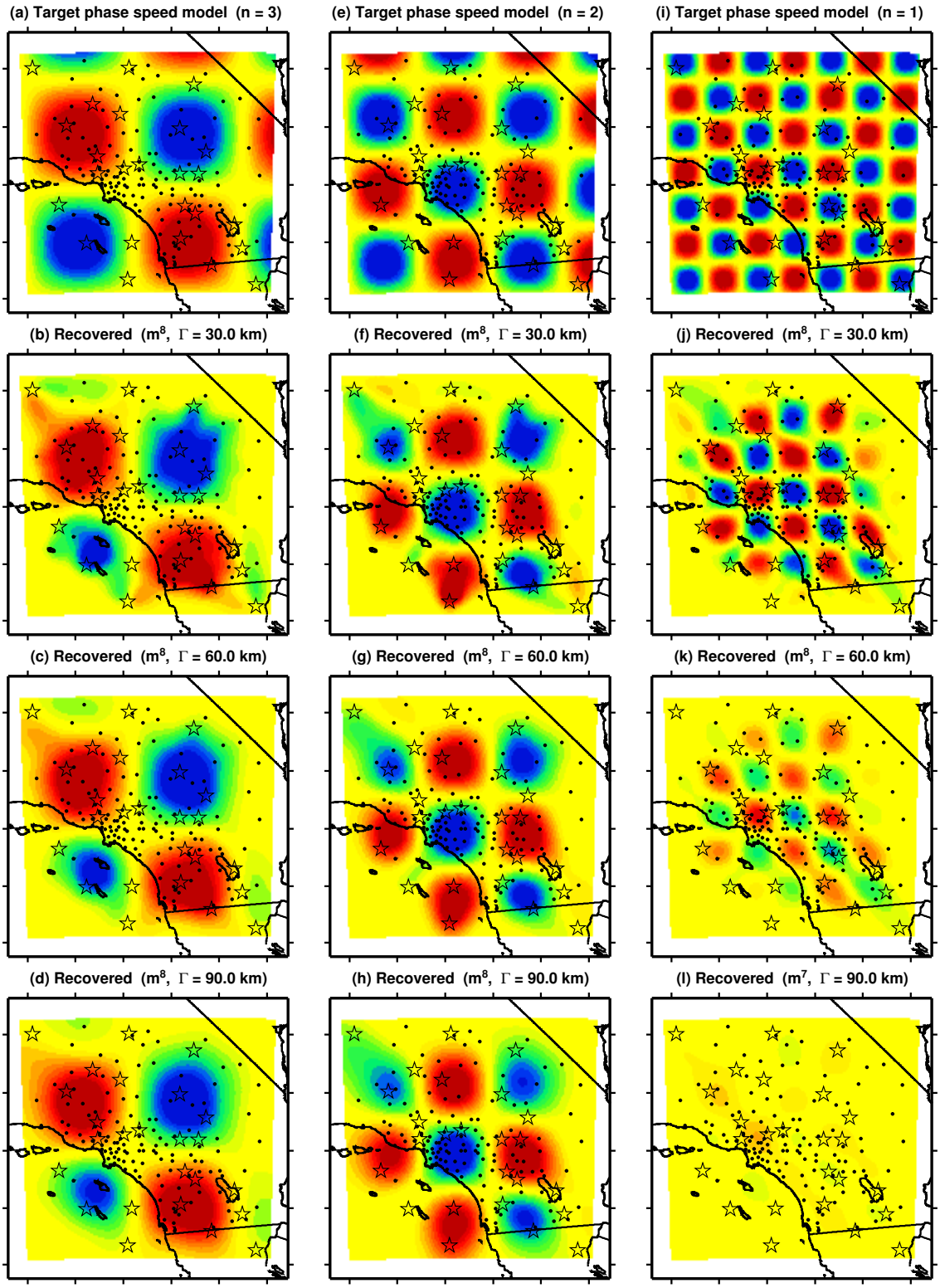


Figure 2.15: Effect of the degree of smoothing and scalelength of heterogeneity on the recovery of the phase-speed models. The factor n is given by $\Lambda = n\lambda$, where Λ is the scalelength of heterogeneity and $\lambda = c\tau = 70$ km is the reference wavelength. The smoothing parameter, Γ, is constant for each row. (See Figure 2.10 and Section 2.7.)

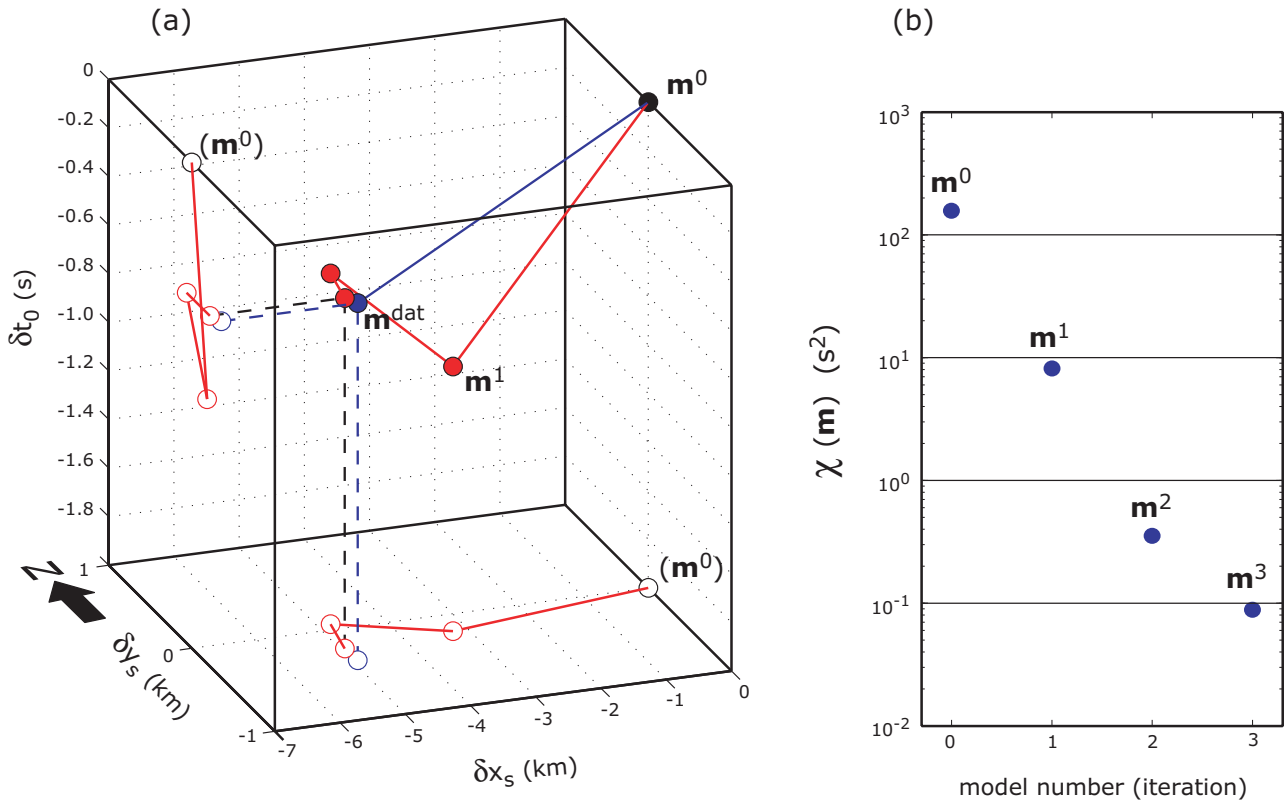


Figure 2.16: Basic source inversion: source recovery using an unperturbed (fixed) wave-speed structure. The model vector, $\mathbf{m} = (\delta x_s, \delta y_s, \delta t_s)$, represents the source with respect to the initial model, where (x_s, y_s) is the location and t_s is the origin time. The data are generated using the target source model $\mathbf{m}^{\mathrm{dat}}$. The initial source model for the synthetics, $\mathbf{m}^0$, initiates 0.53 s late with respect to the data and is mislocated by 4.93 km at an azimuth of N85.8°E with respect to the data. The initial source parameters are randomly chosen from a mislocation "patch" with a 5 km radius and a timing error range of ± 1 s. (a) Iterative improvement of the source model toward the target source model. White circles show the projections of the model points onto horizontal and vertical planes, respectively; these are shown to aid in the perspective. (b) Reduction in the traveltime cross-correlation misfit (2.6) for the source models shown in (a). (See also Figure 2.18.)

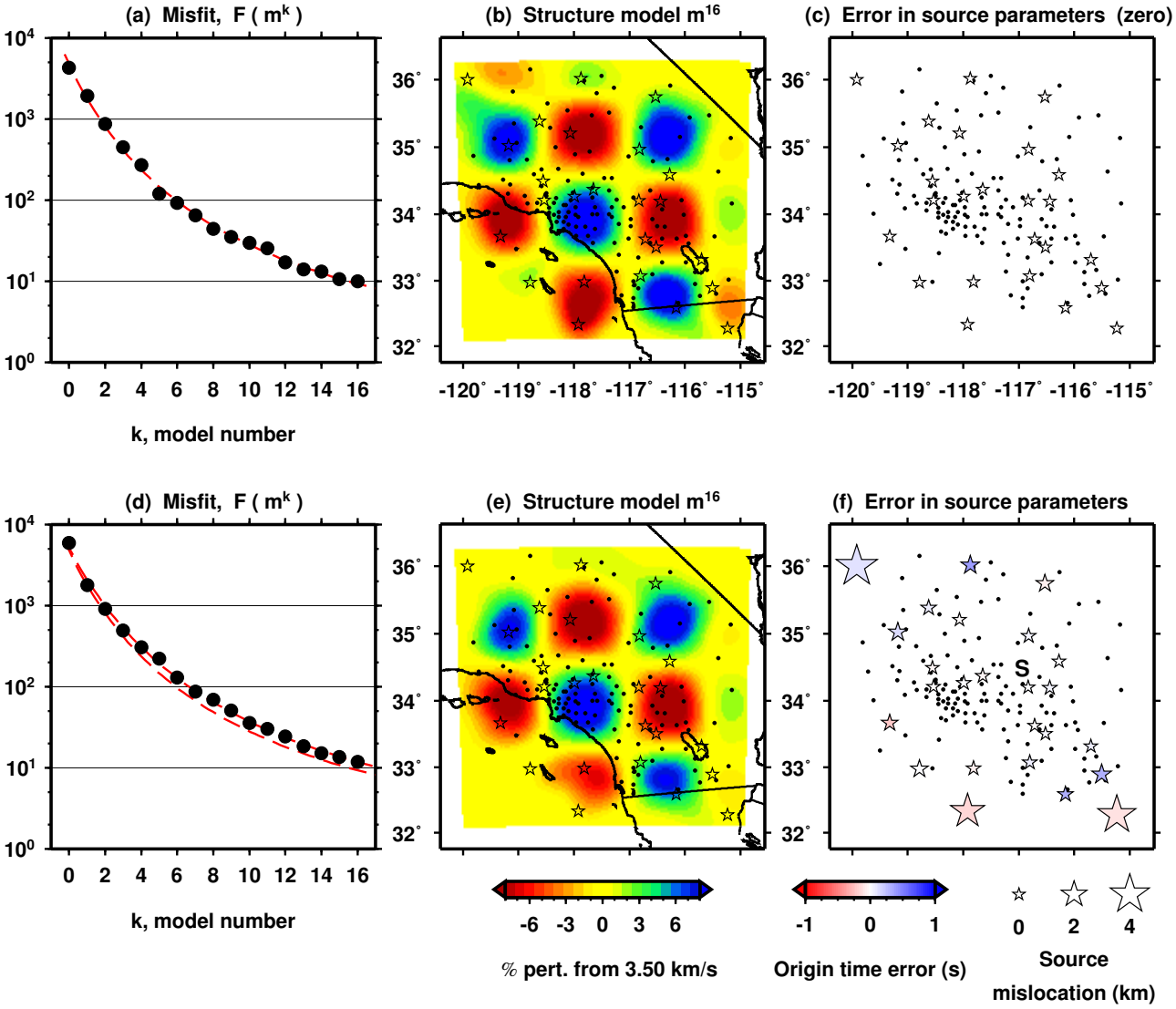


Figure 2.17: Joint inversion for source and structural parameters. The initial structural model is homogeneous. The traveltime cross-correlation misfit function values in (a) and (d) are computed from (2.6). The data are generated using Figure 2.15e. (a) Reduction in misfit for a basic structure inversion, i.e., one in which the structure of the initial model differs from that of the data, but the sources are always identical to those that generated the data. (b) Recovered model $\mathbf{m}^{16}$. Color scale is shown in (e). (c) Error in recovered source parameters. In the basic source inversion, the sources for the data and synthetics are identical and hence there is no error. Key is shown in (f). (d) Reduction in misfit for a joint inversion. The lower dashed curve is the basic structure inversion in (a). (e) Recovered model $\mathbf{m}^{16}$. Subtle differences from (b) can be seen near the edges. (f) Error in recovered source parameters. The initial error in the source parameters is shown in Figure 2.19c. Sources near the edges have the largest remaining error. The recovery of the source parameters for the event labeled S is shown in Figure 2.18.

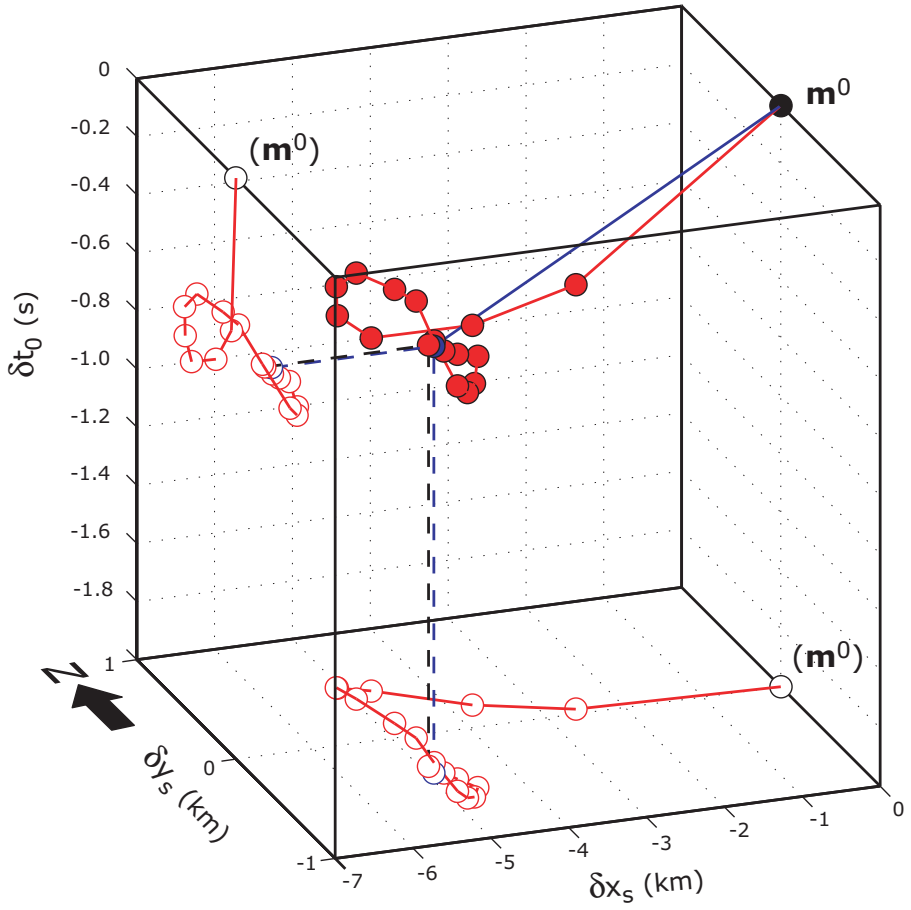


Figure 2.18: Source recovery of a particular event during the joint inversion shown in Figure 2.17d-f. The source location is denoted by the S in Figure 2.17f. By the sixteenth iteration, the source is nearly identical to the source used to generate the synthetics. The recovered structure is shown in Figure 2.17e. Compare with Figure 2.16a, which is the same source perturbation, (4.93 km, −0.53 s), but for a basic source inversion.

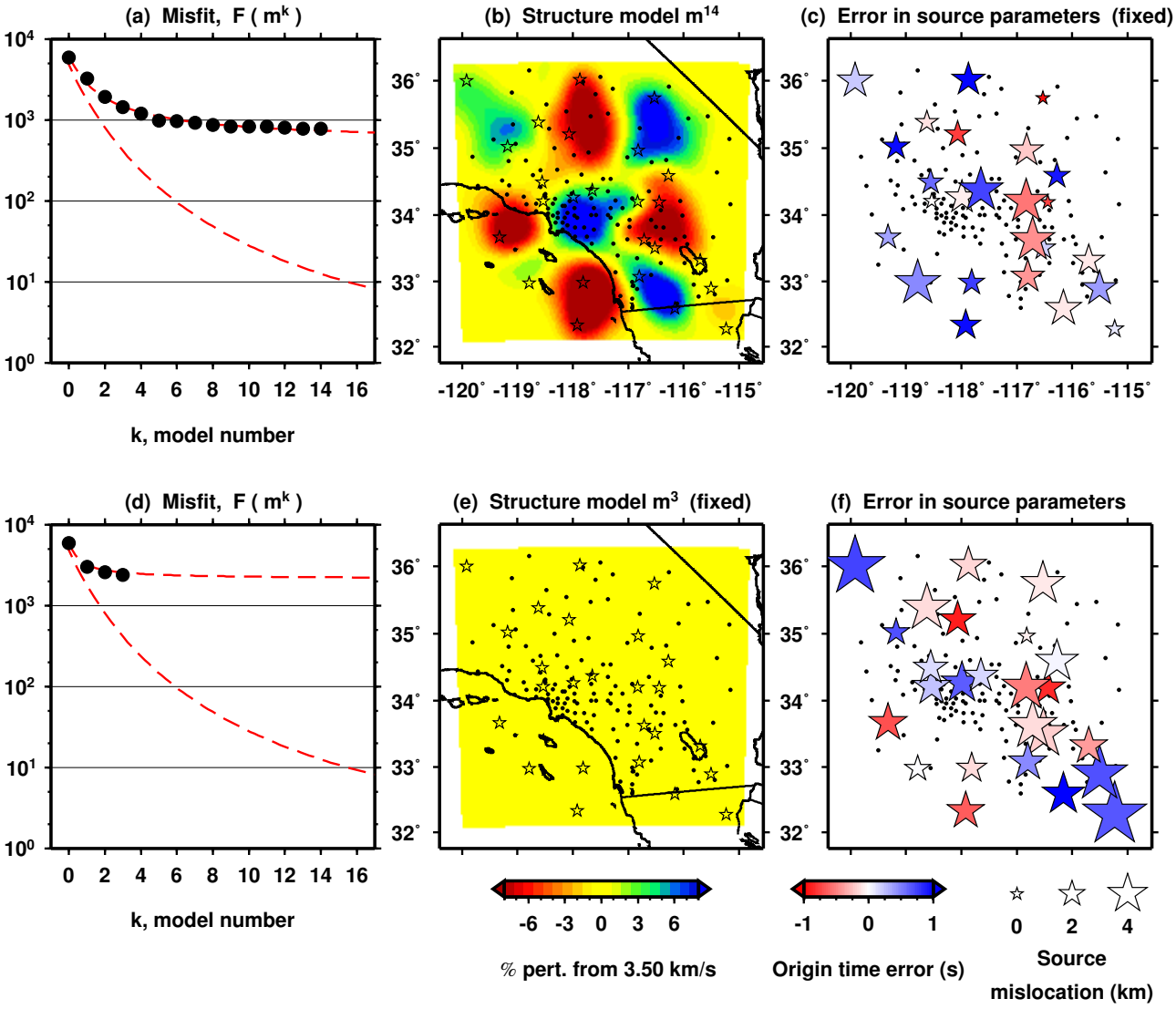


Figure 2.19: Mapping source errors onto structure and vice versa. The initial source and structural model parameters are different from the target source and structural parameters. The traveltime cross-correlation misfit function values in (a) and (d) are computed from (2.6); the number of values is <17, because the stopping criterion for the conjugate gradient algorithm was reached. The data are generated using Figure 2.15e. (a) Reduction in misfit for a structure inversion, whereby the source errors are fixed. The lower dashed curve is the basic structure inversion in Figure 2.17a. (b) Recovered model $\mathbf{m}^{14}$. Color scale is shown in (e). Note the discrepancy with Figure 2.17b. (c) Error in source parameters used in the inversion. Key is shown in (f). (d) Reduction in misfit for a source inversion, whereby the structure errors are fixed. The lower dashed curve is the basic structure inversion in Figure 2.17a. (e) Structure model used in the inversion. (f) Error in recovered source parameters. The initial error in the source parameters is shown in (c).

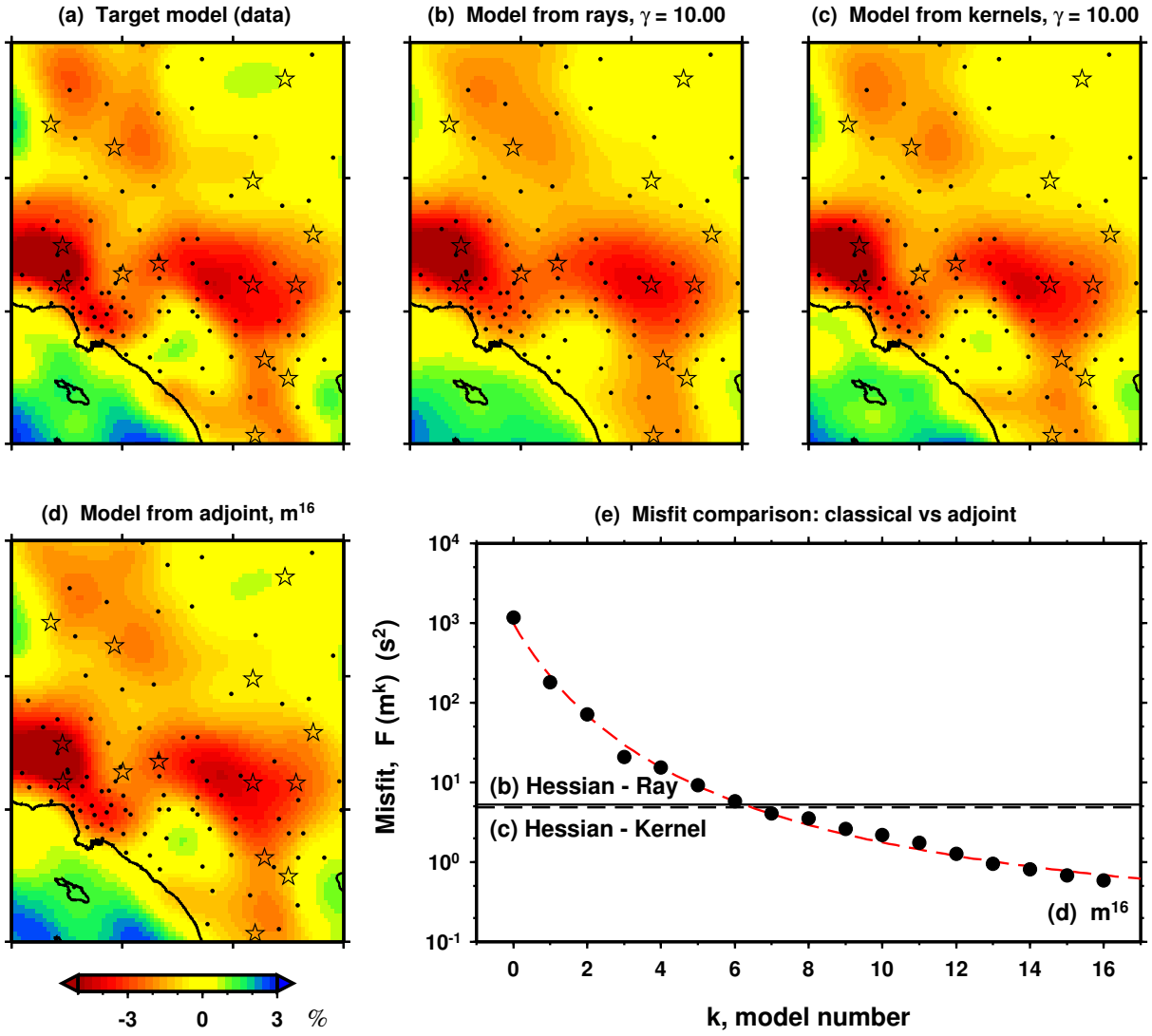


Figure 2.20: Comparison of recovered phase-speed models for classical and adjoint tomography. (a) Target model used to generate the data; expanded version is shown in Figure 2.1b. (b) Recovered model $\mathbf{m}^{\text{ray}}$ using classical, ray-based inversion. The damping parameter γ is defined in (2.27). (c) Recovered model $\mathbf{m}^{\text{ker}}$ using classical, kernel-based inversion (Figure 2.4c). (d) Recovered model $\mathbf{m}^{16}$ using the adjoint method and a conjugate gradient algorithm (Figure 2.13e). (e) Misfit comparison for the three approaches (eq. 2.6). The horizontal lines denote the misfit computed for the ray- and kernel-based models shown in (b) and (c). (See Section 2.9.2 for details.)

Chapter 3

Construction of finite-frequency kernels using adjoint methods

Note

This chapter contains excerpts from "Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels," by Jeroen Tromp, Carl Tape, and Qinya Liu. My primary contribution to this study was to adapt a 2D SEM wave propagation code to construct finite-frequency kernels. In a series of numerical experiements, I illustrated the formation of finite-frequency sensitivity kernels via the interaction between a forward wavefield ($\mathbf{s}$), propagating from source to receiver, with an adjoint wavefield ($\mathbf{s}^\dagger$), propagating from receiver to source. In this chapter, I have included some additional figures to complement those in *Tromp et al.* (2005).

3.1 Kernel Gallery

Expressions for seisitivity kernels for a α-β-ρ parameterization of compressional wave speed (α), shear wave speed (β), and density (ρ) are given by *Tromp et al.* (2005, Eq. 51):

$$\begin{aligned}
\bar{K}_{\rho(\alpha\beta)} &= \bar{K}_{\rho(\kappa\mu)} + \bar{K}_{\kappa(\mu\rho)} + \bar{K}_{\mu(\kappa\rho)}, \\
\bar{K}_{\beta(\alpha\rho)} &= 2\left(\bar{K}_{\mu(\kappa\rho)} - \frac{4}{3}\frac{\mu}{\kappa}\bar{K}_{\kappa(\mu\rho)}\right), \\
\bar{K}_{\alpha(\beta\rho)} &= 2\left(1 + \frac{4}{3}\frac{\mu}{\kappa}\right)\bar{K}_{\kappa(\mu\rho)} \, .
\end{aligned} \tag{3.1}$$

Appendix B shows similar expressions for parameterizations in κ-μ-ρ and c-β-ρ.

We use a two-dimensional (2D) elastic wave propagation code to illustrate the construction of sensitivity kernels using the adjoint methodology discussed in this paper. Each kernel is based upon the interaction between a regular wavefield $\mathbf{s}$ and an adjoint wavefield $\mathbf{s}^\dagger$. Changing the adjoint source $\mathbf{f}^\dagger$ results in a different adjoint field $\mathbf{s}^\dagger$ and, hence, different kernels. For example, we can use the residuals between the data and the synthetics as the waveform adjoint source to construct misfit kernels, or we can use the synthetic velocity field as the traveltime adjoint source to construct banana-doughnut kernels. In this section we present examples of finite-frequency traveltime kernels.

3.1.1 Model setup

We simulate 2D elastic wave propagation using a spectral-element method, which combines the flexible spatial parameterization of finite-element methods with the accuracy of pseudospectral methods (e.g., *Komatitsch and Tromp*, 1999). The source-receiver geometry and the various SH and P-SV body-wave arrivals are illustrated in Figure 3.1. The top boundary is a free surface, whereas the remaining three boundaries are absorbing to mimic a half space. The model extends 200 km in width and 80 km in depth, and is homogeneous with density $\rho = 2600$ kg m^{-3}, bulk modulus $\kappa = 5.20 \times 10^{10}$ Pa, and shear modulus $\mu = 2.66 \times 10^{10}$ Pa; these values correspond to a compressional wave speed of $\alpha = 5800$ m s^{-1} and a shear wave speed of $\beta = 3199$ m s^{-1}. We use a simple one-way treatment for the implementation of the absorbing boundary conditions (*Komatitsch and Tromp*, 1999). For pedagogical reasons, both the source and the receiver are located at a depth of 40 km to generate direct and surface reflected waves, leading to a variety of interesting phases and associated kernels.

The source-time function used in the simulations is a Gaussian of the form

$$h(t) = (-2\alpha^3/\sqrt{\pi})\ (t - t_0)\ \exp[-\alpha^2(t - t_0)^2], \tag{3.2}$$

where $t_0 = 8.0$ s, $\alpha = 2\tau_0/\tau$, $\tau_0 = 2.628$ s, and τ is the duration of $h(t)$ (e.g., Figure 3.2a). The source duration is $\tau = 4.0$ s in each example, with the exception of Figure 1.1, where we also used $\tau = 8.0$ s. In each simulation the source is applied in the x and y directions to generate both P-SV and SH motions (which are of course completely decoupled). Changing the orientation of the source results in different sensitivity kernels.

3.1.2 Banana-doughnut kernels

Banana-doughnut traveltime kernels are constructed by using the time-reversed velocity field at one particular receiver as the adjoint source. Kernels calculated in this manner may be compared with the finite-frequency kernels presented in recent studies using ray-based methods (e.g., *Hung et al.*, 2000). As discussed earlier, the construction of each kernel is based on the interaction between the time-reversed regular field and the adjoint field; hence the "interaction field" can be thought of as propagating from the receiver to the source in reverse time from $t = T$ to $t = 0$.

SH waves

We begin with the simplest case, the SH wavefield. The experimental setup is depicted in Figure 3.1. Because both the source and the receiver are located at depth, there are two possible arrivals, which we label S and SS. The source-time function used to generate the regular wavefield is shown in Figure 3.2a, and the associated seismogram with distinct S and SS arrivals is displayed in Figure 3.2b. Figure 3.3 illustrates the construction of the $\bar{K}_{\beta(\alpha\rho)}$ kernel from the interaction between the regular field $\mathbf{s}$ and the S adjoint field $\mathbf{s}^\dagger$, whose source is shown in Figure 3.2d. Keep in mind that for increasing time t the regular field propagates from the source to the receiver, whereas the adjoint field propagates from the receiver to the source. Marching backward in time from $t = T$, the traveltime adjoint source (located at the receiver) "turns on" at the precise moment that the regular S wavefield passes over it (between Figure 3.3a and 3.3b). At each moment in time the two wavefields are combined via (3.1) to form the "interaction" field, which is integrated to construct the kernel. In other words, the interaction field represents the time-dependent integrand in the kernel definition. Once the regular source is "extinguished," no further contributions are made toward $\bar{K}_{\beta(\alpha\rho)}$. Note that $\bar{K}_{\beta(\alpha\rho)}$ is cigar-shaped rather than banana-shaped because the model is homogeneous, and there is no doughnut hole because we are dealing with 2D rather than 3D kernels. We refer to this example as $\mathrm{SH_S}$, where SH designates participation of only the y-component of the wavefields, and the subscript S denotes the phase that is being reversed. The pulse is tapered within the time window w_r using a Welch window (*Press et al.*, 1994).

Figures 3.4 and 3.5 show the effect of reversing different time windows of the synthetic

velocity field. Reversing the SS pulse gives a kernel in the shape of a "folded-over cigar" (Figure 3.5f). The ellipse surrounding the source and receiver represents SS scatterers with comparable traveltimes to the SS wave reflected at the surface. Reversing the entire waveform (i.e., both pulses) illuminates the sensitivity regions of both S and SS (Figure 3.5gh).

Figure 3.6 shows all six kernels for the SH_S scenario. These kernels are constructed simultaneously via the process illustrated in Figure 3.3 for $\bar{K}_{\beta(\alpha\rho)}$. Notice that the relative amplitudes of the kernels are consistent with what is expected from the relationships in (3.1). For example, since $\bar{K}_{\kappa(\mu\rho)} = 0$ and $\bar{K}_{\mu(\kappa\rho)} \approx -\bar{K}_{\rho(\kappa\mu)}$, we see that $\bar{K}_{\rho(\alpha\beta)} = \bar{K}_{\mu(\kappa\rho)} + \bar{K}_{\kappa(\mu\rho)} + \bar{K}_{\rho(\kappa\mu)}$ is very weak. Note that for SH waves we have $\bar{K}_{\beta(\alpha\rho)} = 2\,\bar{K}_{\mu(\kappa\rho)}$.

Figure 1.1 (Chapter 1) illustrates the effect of changing the source duration, τ in (3.2), on the kernels. We see that the width of the kernel shrinks at higher frequencies. We expect this since in the limit of infinite frequency the kernel should collapse onto the ray path. Note that the amplitude of the kernel increases with increasing frequency. This frequency dependence was illustrated by *Hung et al.* (2000) using a different technique to construct the kernels. Cross sections of the kernels (Figure 1.1d) help to highlight the Fresnel zones. In the case of the SH_S β-kernel, the broad, low-sensitivity red zone represents the first Fresnel zone, whereas the sidelobes defined by the narrow, high-sensitivity green zone correspond to the second Fresnel zone (e.g., *Hung et al.*, 2000).

P-SV waves

The P-SV wavefield is more complicated than the SH wavefield (Figure 3.1), and even in the homogeneous case Rayleigh waves arise through interactions at the free surface. Figure 3.7 illustrates the construction of the P-SV adjoint source for the PS+SP arrival, and Figures 3.8 and 3.9 show the corresponding formation of the $\bar{K}_{\rho(\alpha\beta)}$, $\bar{K}_{\alpha(\beta\rho)}$, and $\bar{K}_{\beta(\alpha\rho)}$ kernels.

Notice how the interactions SP$\sim$P† and PS$\sim$S† form the right portion of the sensitivity kernel, e.g., at $t = 32.0$ s (Figure 3.8b), whereas the left portion results from the interactions P$\sim$SP† and S$\sim$PS†, e.g., at $t = 16.0$ s (Figure 3.8d). This can be deduced by matching up the portions of the regular and adjoint wavefields that are contributing to the interaction field. These interactions "paint" the resultant sensitivity kernel.

Figure 3.10 shows all nine kernels for the P-SV_{PS+SP} scenario. We have included the c-β-ρ parameterization, where c is bulk sound speed, in addition to the κ-μ-ρ and α-β-ρ cases. These expressions are derived in Appendix B.

Figure 3.11 shows the effect of reversing four distinct time windows of the P-SV synthetic velocity field: the P, PP, PS+SP, and SS arrivals. In Figure 3.11b we see that the α-kernel for P-SV$_{\rm P}$ is wider than the β-kernel for SH$_{\rm S}$ (Figure 3.5c) for the same source period. This is due to the relatively longer wavelengths of the P waves: $\lambda_\alpha = \alpha T > \lambda_\beta = \beta T$. Figures 3.11c-f illustrate examples of α and β kernels for the P-SV$_{\rm PP}$, P-SV$_{\rm PS+SP}$, and P-SV$_{\rm SS}$ scenarios. Note that, as expected, the α-kernel for the SS wave (Figure 3.11e) is insignificant relative to the β-kernel (Figure 3.11f).

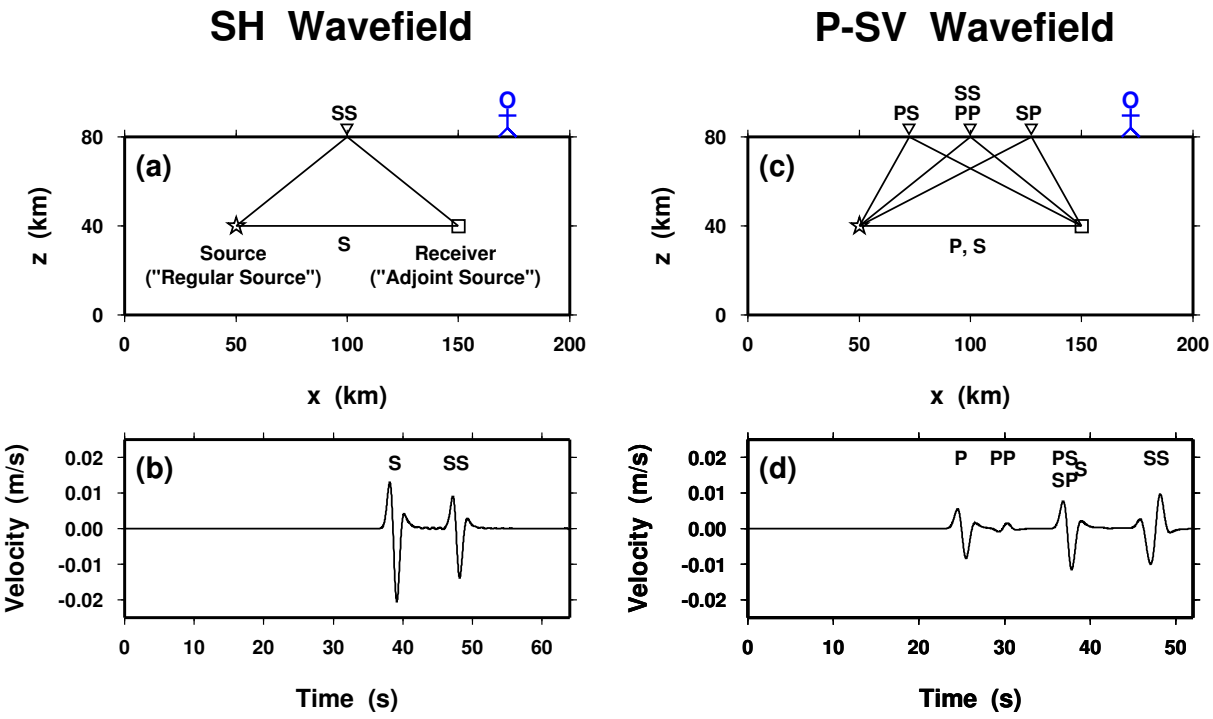


Figure 3.1: Sketch of the 2D model dimensions and the source-receiver geometry (after *Tromp et al.*, 2005, Figure 1). The solid line denotes a free surface, whereas dashed lines are absorbing boundaries. The source is indicated by the $\star$ and the receiver by the $\square$. Left: The two possible ray paths for the SH wavefield are labeled S and SS. The $\bigtriangledown$ denotes the SS bounce point. Right: The possible body-wave ray paths for the P-SV wavefield. The ray paths are based on a homogeneous model.

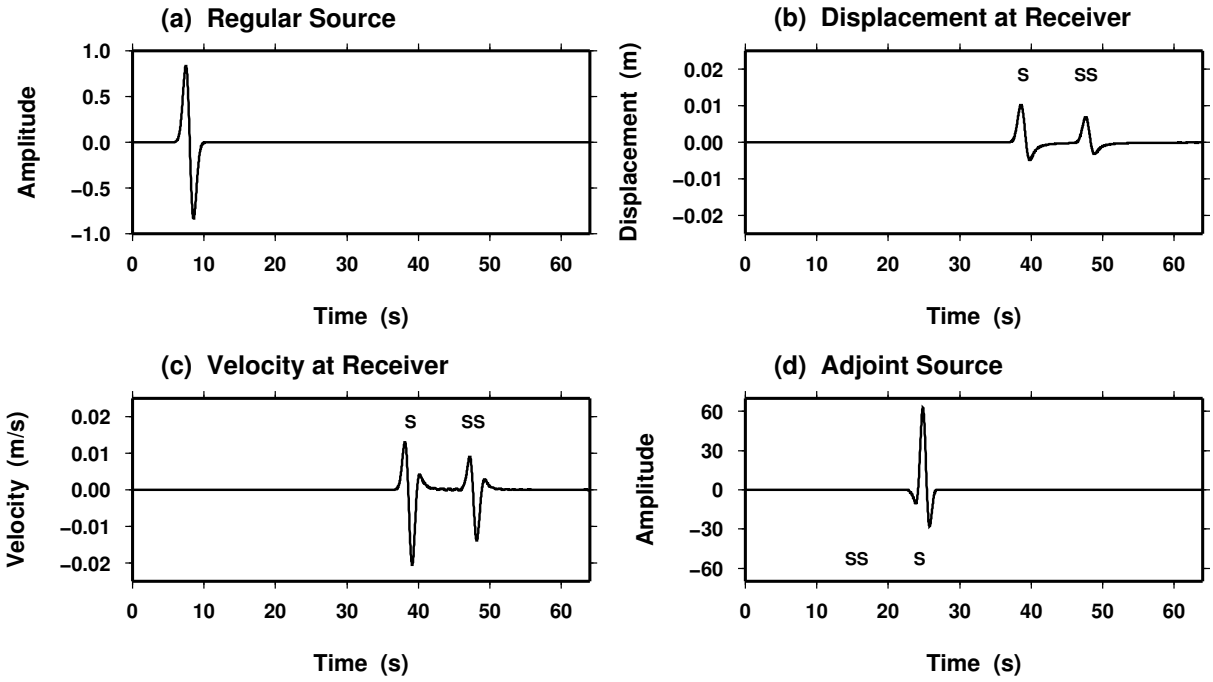


Figure 3.2: Construction of the adjoint source-time function used in calculating SH banana-doughnut kernels (*Tromp et al.*, 2005, Figure 2). All traces represent the y-component. (a) Source for the regular wavefield. (b) Regular seismogram recorded at the receiver. (c) Velocity seismogram at the receiver. (d) Source for the adjoint wavefield constructed by time-reversing (c) and Welch tapering the S arrival. Note that this includes the normalization factor $M_{\mathrm{T}} \leq 0$ defined in *Tromp et al.* (2005).

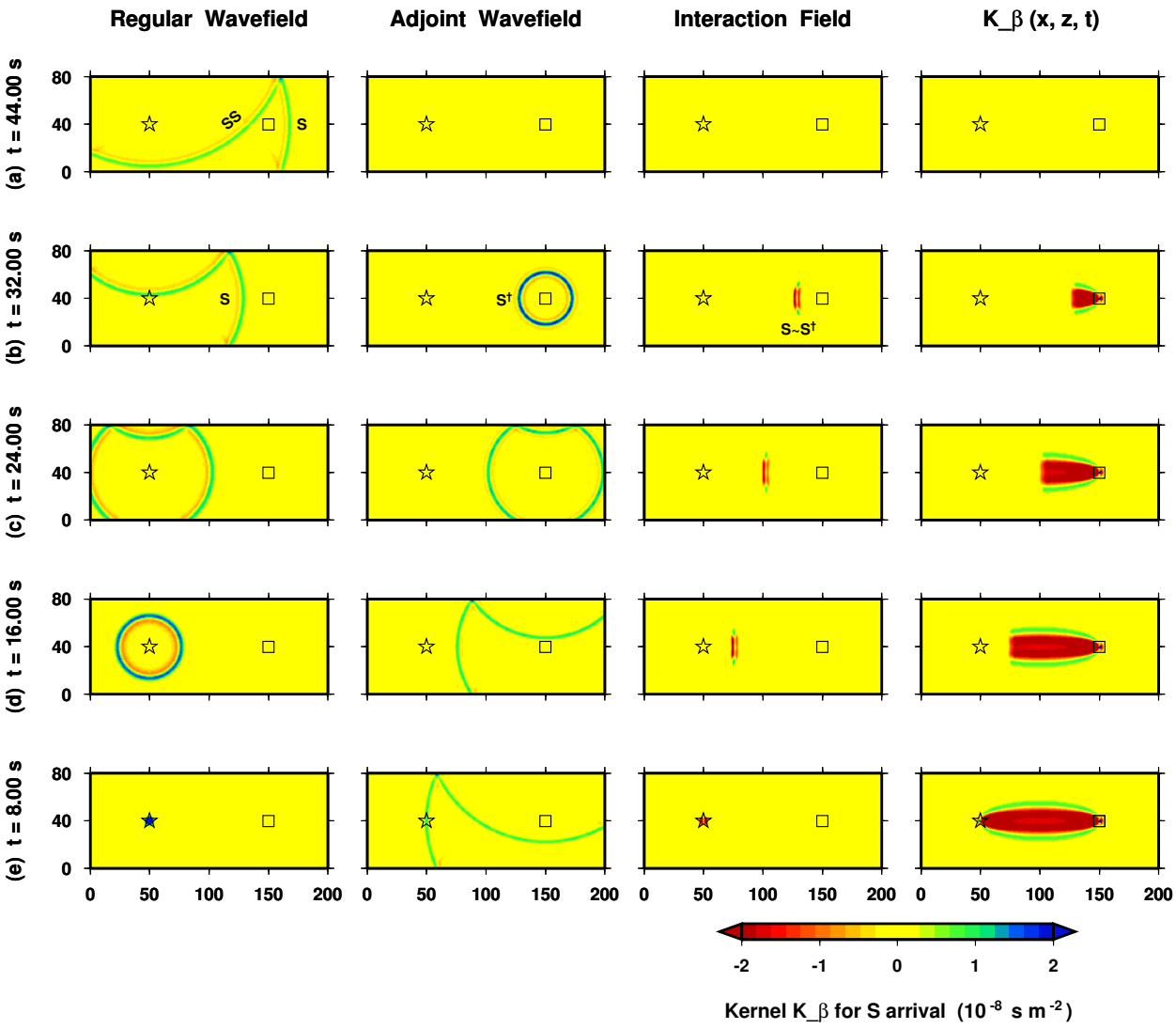


Figure 3.3: Sequence of interactions between the regular and adjoint SH wavefields during the construction of the banana-doughnut kernel $\bar{K}_{\beta(\alpha\rho)}$ (*Tromp et al.*, 2005, Figure 3). This particular $\bar{K}_{\beta(\alpha\rho)}$ kernel is for SH_S, i.e., the SH β–kernel obtained by time-reversing the S arrival. The regular and adjoint sources are shown in Figure 3.2; the model is a homogeneous half space. Each row represents an instantaneous interaction between the regular and adjoint fields. From the left column to the right column are shown the regular field, the adjoint field, the interaction field, and the instantaneous sensitivity to shear velocity perturbations, $\bar{K}_{\beta(\alpha\rho)}$. The $\bar{K}_{\beta(\alpha\rho)}$ kernel is constructed by integrating the interaction field, shown in the third column, over time. (a) At this point in time there is no interaction between the regular field and the adjoint field, since the S wave has yet to reach the receiver. (b) Adjoint wavefield "lights up" as the regular wavefield S phase passes over the receiver (traveling toward the source in reverse time). The label $S{\sim}S^{\dagger}$ indicates interaction between the regular and adjoint S waves, respectively. (c)–(d) The sensitivity kernel forms via the interaction between the regular and adjoint wavefields. (e) Time of regular source initiation, before which no interaction occurs. The source is labeled by the $\star$ and the receiver by the $\square$.

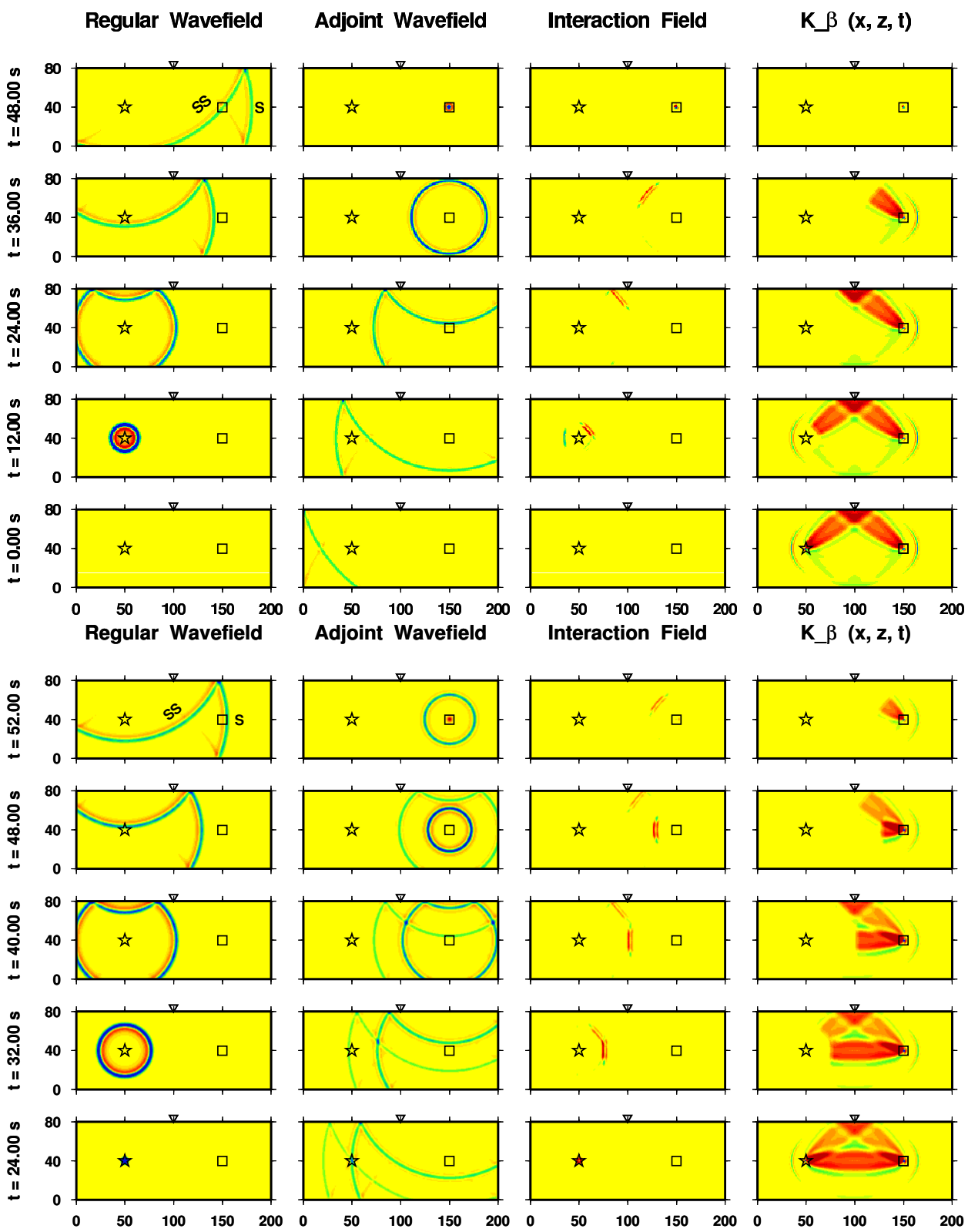


Figure 3.4: Same as Figure 3.3, but here we have time-reversed the SS arrival (top) and the entire record (bottom).

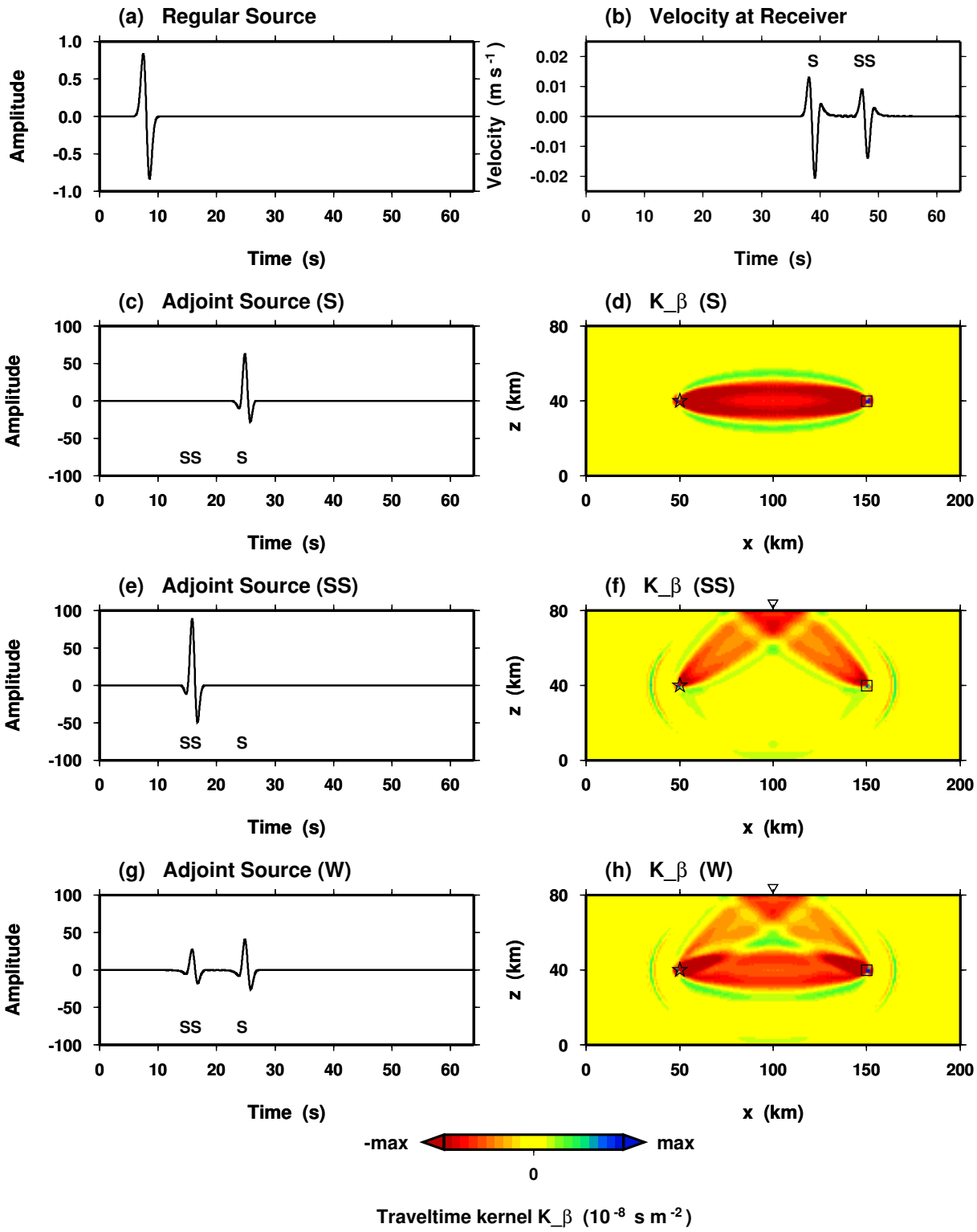


Figure 3.5: The effect of time window selection on sensitivity kernels, using $\bar{K}_{\beta(\alpha\rho)}$ (SH) as an example (after *Tromp et al.*, 2005, Figure 5). (a) Source for the regular wavefield. (b) Velocity recorded at the receiver showing the arrivals S and SS. (c) Adjoint source for $\mathrm{SH_S}$, constructed by time-reversing S in (b) and normalizing by M_T defined in *Tromp et al.* (2005). (d) $\bar{K}_{\beta(\alpha\rho)}$ for reversing S only. (e) Adjoint source for $\mathrm{SH_{SS}}$, constructed by time-reversing SS in (b) and normalizing by M_T. (f) $\bar{K}_{\beta(\alpha\rho)}$ for reversing SS only. Each point on the ellipse represents a scattering point for a path with a comparable traveltime to the SS path. The SS bounce point is labeled by the $\bigtriangledown$, the source by the $\star$, and the receiver by the $\square$.

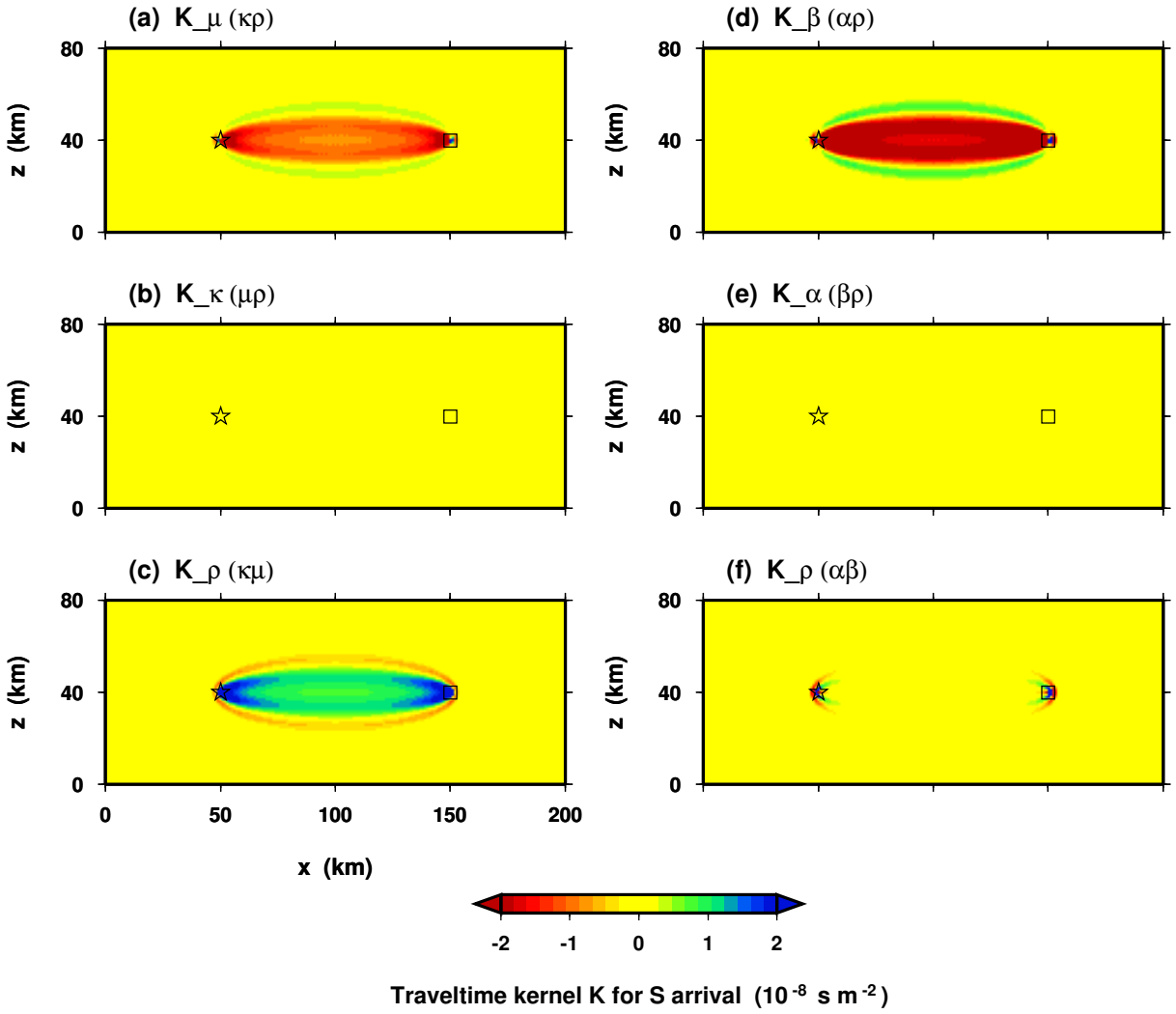


Figure 3.6: The six SH_S banana-doughnut kernels. Each kernel is constructed simultaneously as shown in Figure 3.3 for $\bar{K}_{\beta(\alpha\rho)}$. Note that $\bar{K}_{\alpha(\beta\rho)} = \bar{K}_{\kappa(\mu\rho)} = 0$, $\bar{K}_{\beta(\alpha\rho)} = 2\bar{K}_{\mu(\kappa\rho)}$, $\bar{K}_{\mu(\kappa\rho)} \approx -\bar{K}_{\rho(\kappa\mu)}$, and $\bar{K}_{\rho(\alpha\beta)} \approx 0$ for SH propagation.

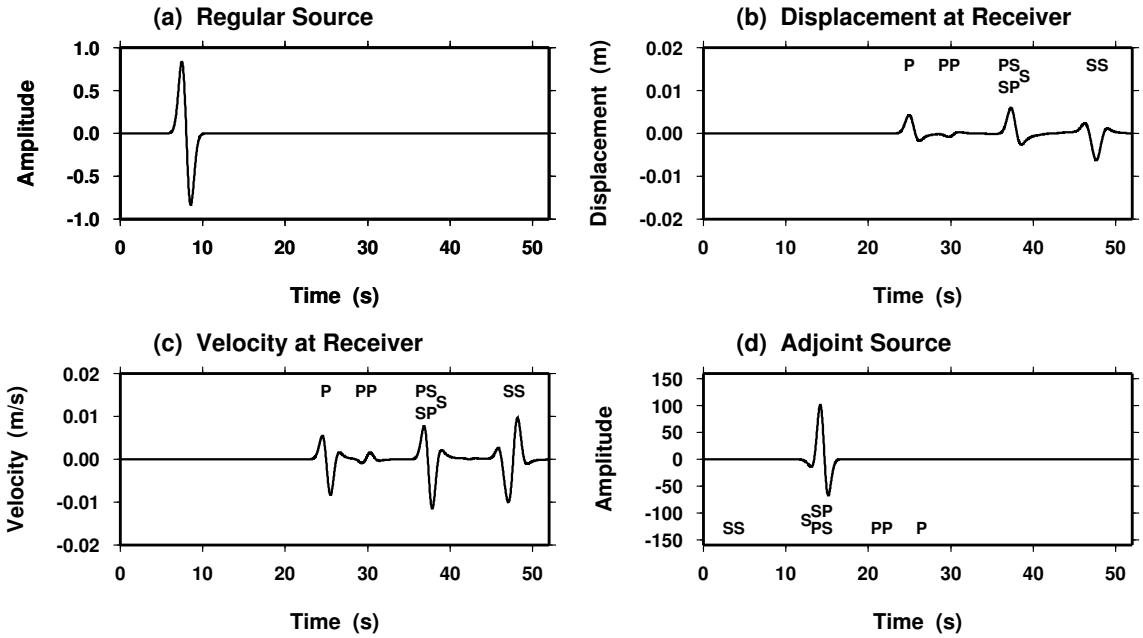


Figure 3.7: Construction of the adjoint source-time function used in calculating P-SV banana-doughnut kernels. (a) Source-time function responsible for the regular wavefield (x-component; the z-component is zero). (b) Regular seismogram (x-component). (c) Velocity seismogram (x-component). (d) Source-time function for the adjoint source constructed by time-reversing (c) and Welch tapering the PS+SP arrival (x-component). Note that this includes the normalization factor M_{T} defined in *Tromp et al.* (2005). This is the source-time function used in Figure 3.8.

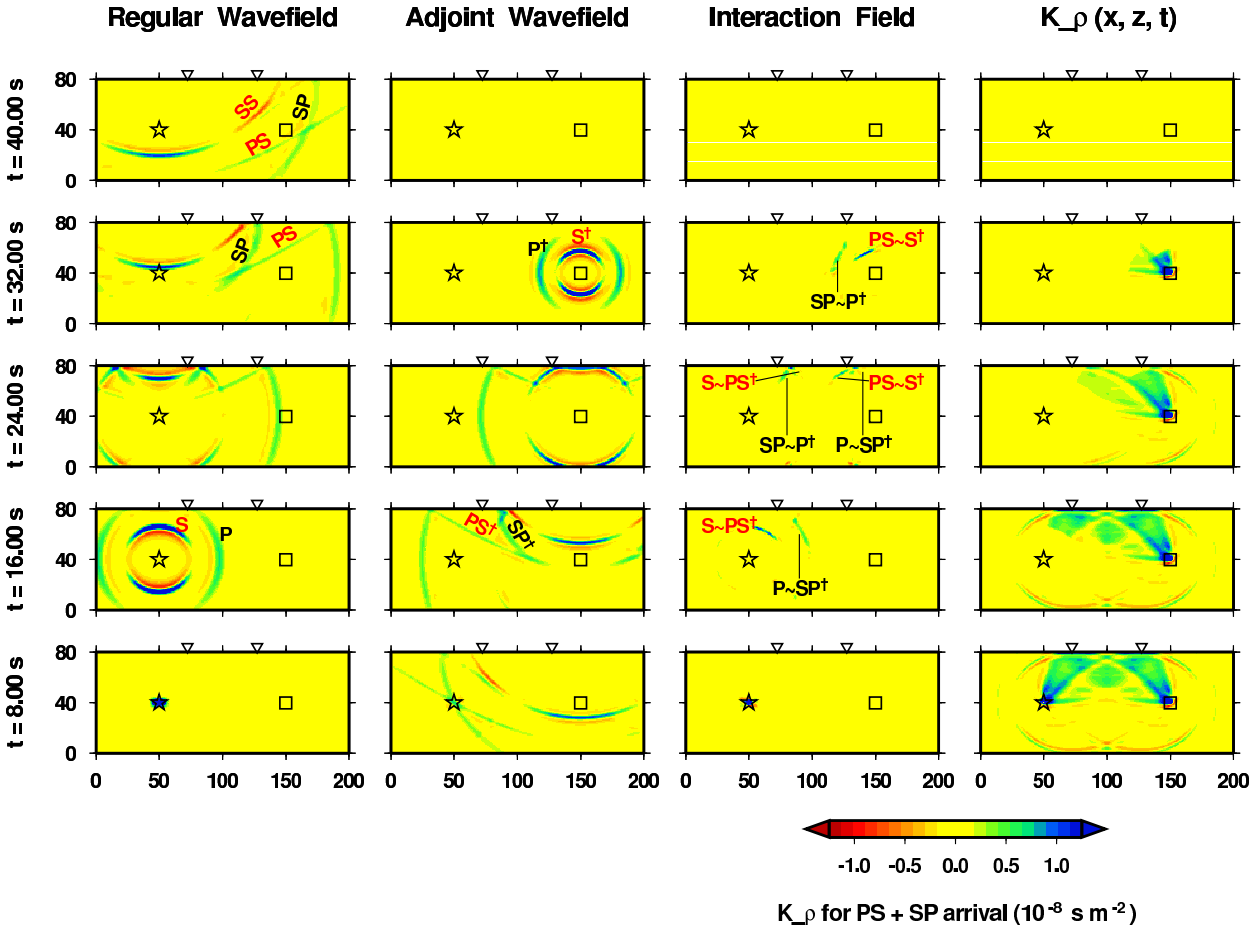


Figure 3.8: Sequence of interactions between the regular and adjoint P-SV wavefields to produce the banana-doughnut kernel $\bar{K}_{\rho(\alpha\beta)}$. This particular $\bar{K}_{\rho(\alpha\beta)}$ is for P-SV$_{\mathrm{PS+SP}}$, i.e., the P-SV ρ-kernel obtained by time-reversing the PS+SP arrival. Given the geometry in Figure 3.1, the SP and PS phases arrive simultaneously, at nearly the same time as the S arrival (Figure 3.7b). The x-z grid in each snapshot is 200 km in width and 80 km in depth. Wavefield snapshots capture the x-component of displacement. We use labels $\star$ for the source, $\square$ for the receiver, and $\bigtriangledown$ for the PS (right) and SP (left) bounce points. See Section 3.1.2 for details, and compare with Figure 3.3. (a) No interaction between the regular and adjoint fields, since the PS+SP phase has yet to reach the receiver. (b) Adjoint wavefield "lights up" as the regular wavefield PS+SP phase, depicted by the X-shaped crossing of the two green wavefields, passes over the receiver (traveling toward the source). (c)–(d) Sensitivity kernel forms via the interaction between the regular and adjoint fields. (e) Time of regular source initiation, before which no interaction occurs.

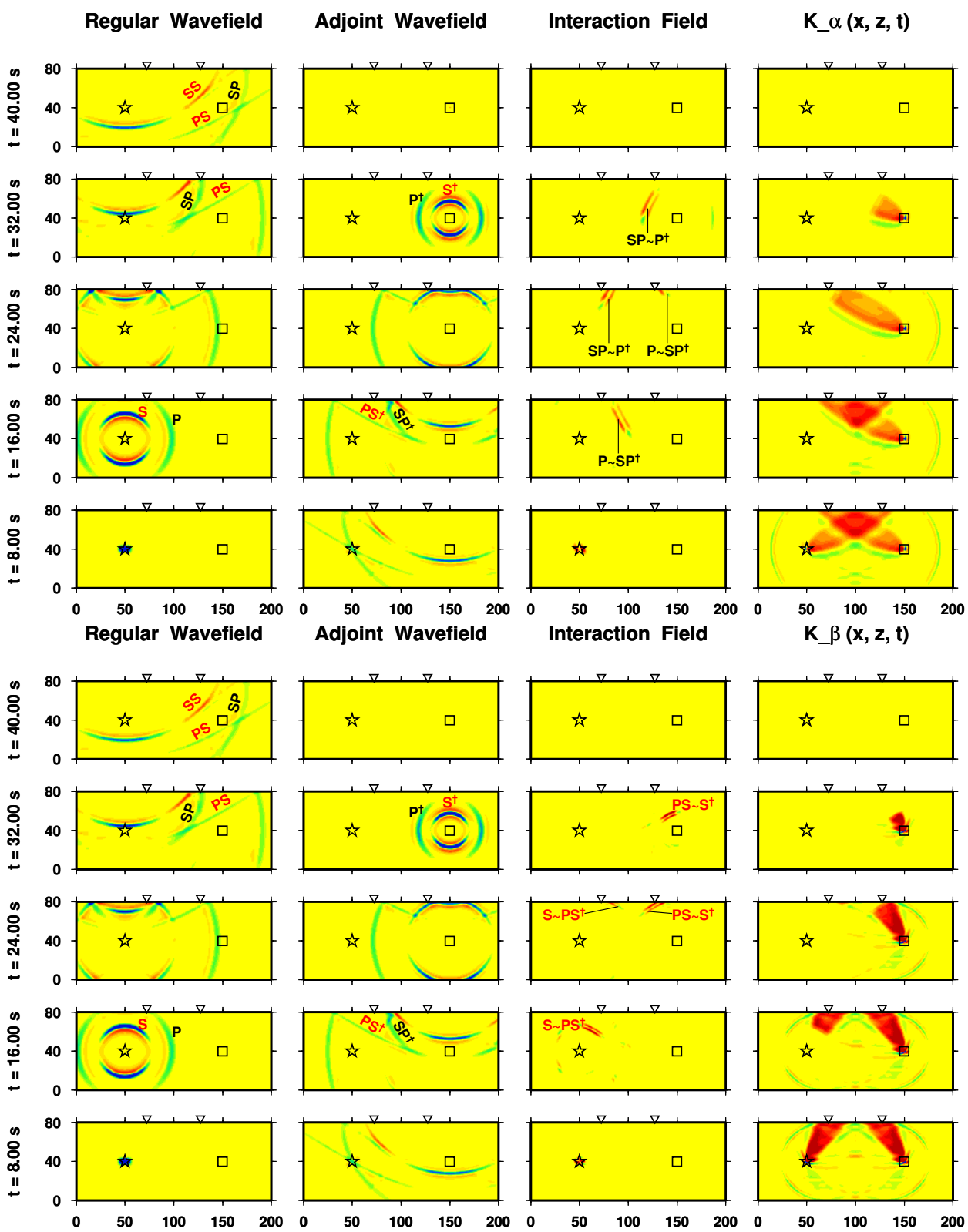


Figure 3.9: Same forward and adjoint wavefields as in Figure 3.8, but here we show the formation of the α-kernel (top) and the β-kernel (bottom).

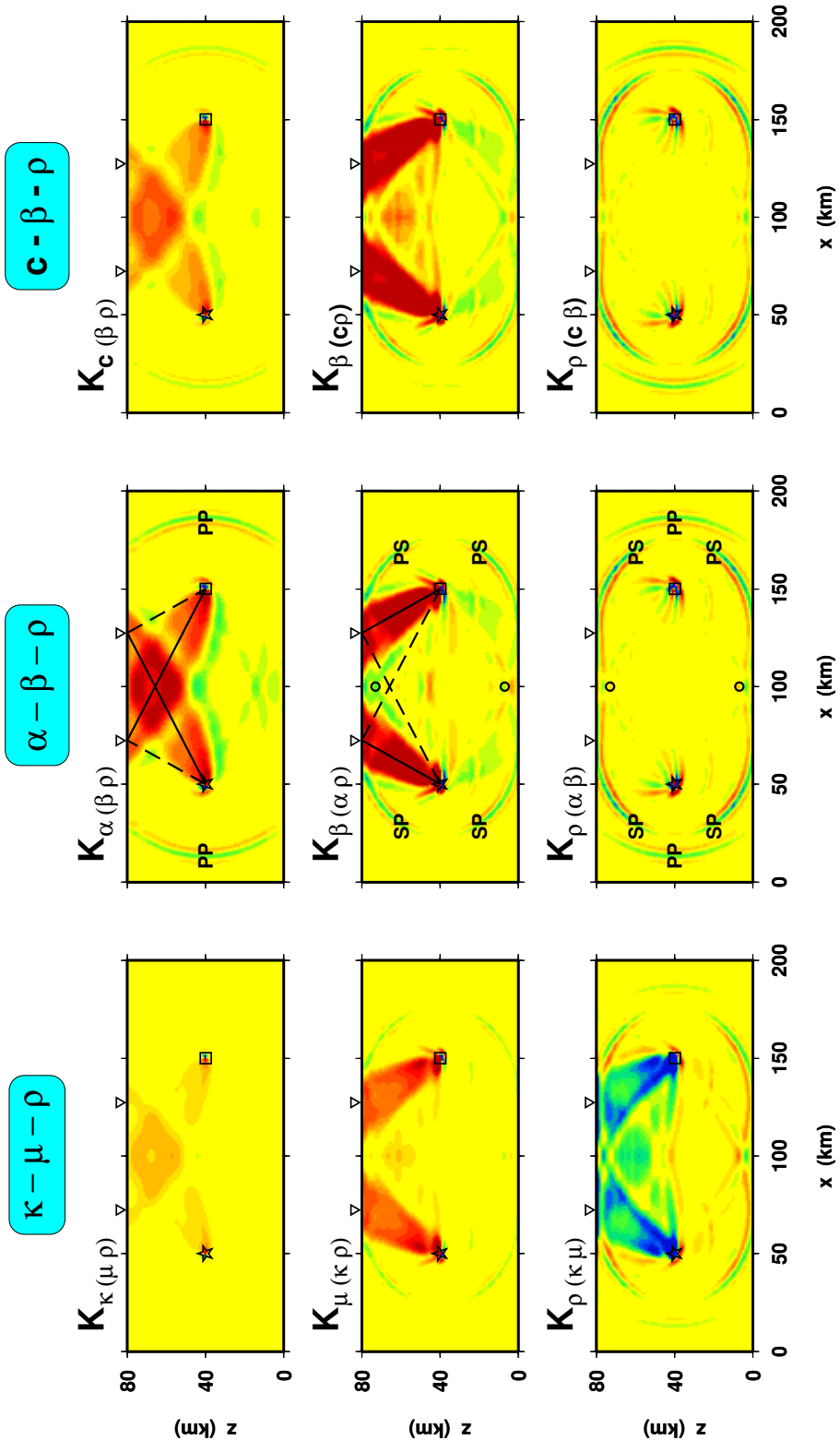


Figure 3.10: P-SV$_{\mathrm{PS+SP}}$ banana-doughnut kernels for three different model parameterizations (see Appendix B). Each kernel is constructed simultaneously, as shown in Figure 3.8. Notice that the predominant shape of the $\bar{K}_{\alpha(\beta\rho)}$ kernel is that of two adjacent, folded-over "cigars," the right one for PS and the left for SP. The labels $\bigtriangledown$ denotes the PS (right) and SP (left) bounce points.

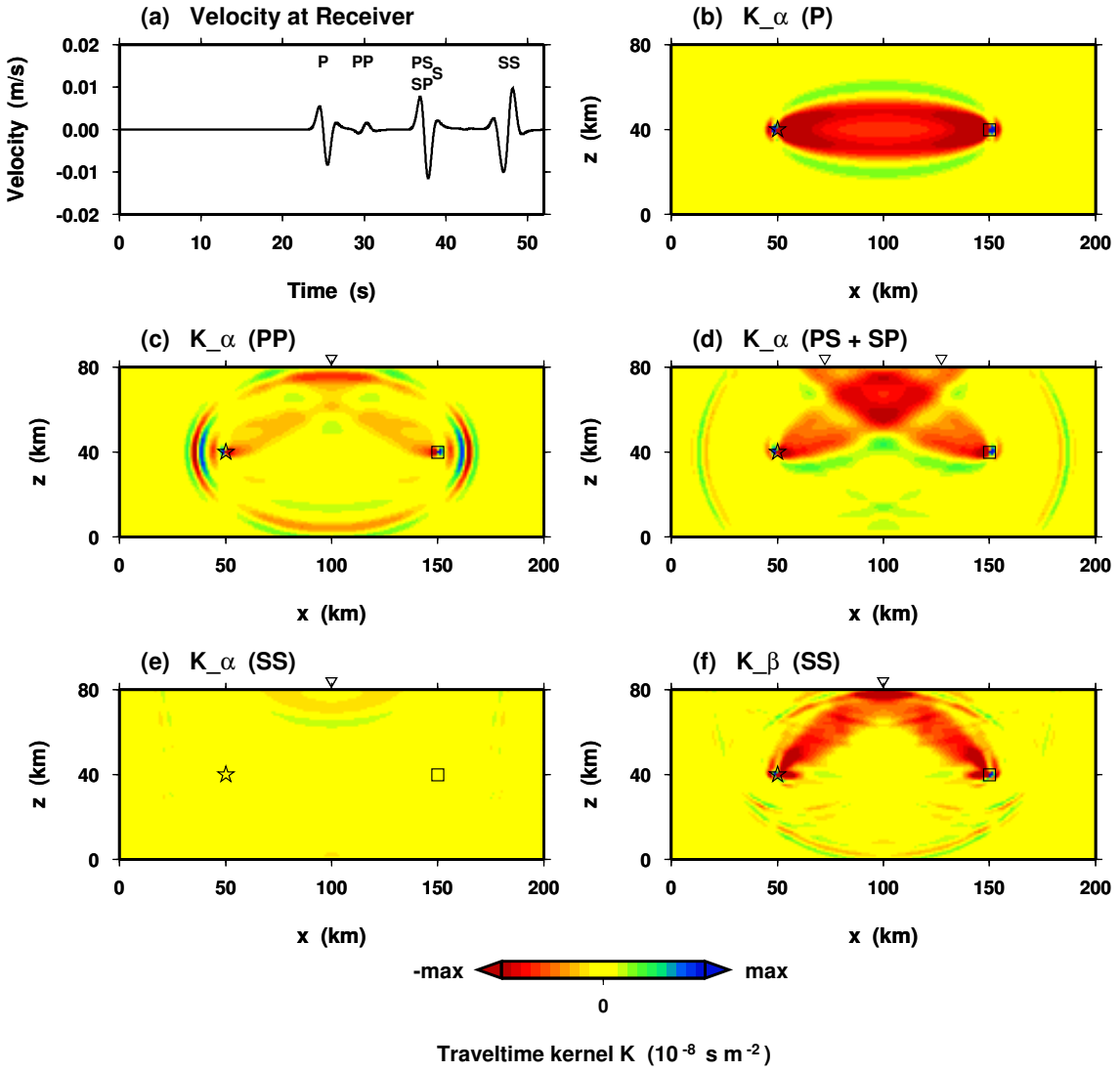


Figure 3.11: The effect of time-window selection on sensitivity kernels, using $\bar{K}_{\alpha(\beta\rho)}$ as an example. See Figure 3.1 for labeling and Section 3.1.2 for details. (a) Velocity recorded at the receiver (x-component) showing the consecutive arrivals of P, PP, PS+SP, and SS. The S phase is expected to arrive nearly simultaneous with PS+SP, but is insignificant on this component. In (b)–(f) we Welch taper one of the pulses and time-reverse it as the adjoint source via the method explained in Figure 3.7. The color scale varies for each plot according to the value "max." (b) $\bar{K}_{\alpha(\beta\rho)}$ for reversing the P arrival (P-SV$_{\mathrm{P}}$, max = 1.0). (c) $\bar{K}_{\alpha(\beta\rho)}$ for reversing the PP arrival (P-SV$_{\mathrm{PP}}$, max = 5.0). (d) $\bar{K}_{\alpha(\beta\rho)}$ for reversing the PS+SP arrival (P-SV$_{\mathrm{PS+SP}}$, max = 1.5). (e) $\bar{K}_{\alpha(\beta\rho)}$ for reversing the SS arrival (P-SV$_{\mathrm{SS}}$, max = 2.75). (f) $\bar{K}_{\beta(\alpha\rho)}$ for reversing the SS arrival (P-SV$_{\mathrm{SS}}$, max = 2.75).

Chapter 4

Adjoint tomography based on source subspace projection

Note

This chapter contains excerpts from a paper in preparation by Carl Tape, Malcolm Sambridge, and Jeroen Tromp. Each author is an equal contributor to the paper. Sambridge proposed using the subspace of sources. The concept was further developed by Tromp within the theoretical framework of *Tromp et al.* (2005). My primary contribution was to implement and test the source subspace projection method in comparison with a conjugate gradient algorithm. I will also focus on joint source-structure inversions using the source subspace method, in comparison with the conjugate gradient results in *Tape et al.* (2007).

4.1 Introduction

In adjoint tomography, for a given model m, one generally has access to the value of the objective function, $F(\mathsf{m})$, and its Fréchet derivative, $\partial F/\partial \mathsf{m}$, but not its second derivative or Hessian, $\partial^2 F/\partial \mathsf{m} \partial \mathsf{m}$. From an inverse theory perspective, this implies that one has to resort to conjugate-gradient based methods to determine the minimum of the objective function, rather than more rapidly converging and thus more desirable Gauss-Newton methods. Numerically, the gradient may be obtained based upon just two simulations for each earthquake: one calculation for the current model and a second, 'adjoint', calculation that uses time-reversed signals at the receivers as simultaneous, fictitious sources (e.g.,

Tarantola, 1984, 1986; *Akçelik et al.*, 2002, 2003; *Tromp et al.*, 2005; *Tape et al.*, 2007). The calculation of the gradient is independent of the number of receivers, components, and picks.

To increase the convergence rate of nonlinear inversion algorithms, *Sambridge et al.* (1991) proposed an improvement to the conjugate gradient algorithm advocated by *Tarantola* (1986) for the nonlinear inversion of seismic reflection data. The *Sambridge et al.* (1991) approach involves decomposing the gradient of the misfit function in terms of parts that correspond to a particular parameter type, e.g., separating the contributions to the gradient due to density, bulk-sound wave speed, shear wave speed, source location, and source mechanism. Collectively, these contributions to the gradient define a small subspace, and the algorithm proceeds by minimizing the objective function within this subspace. By solving a linearized problem within this subspace, at each iteration one only needs to invert a small matrix, which is the projection of the full Hessian onto the subspace.

In this article we introduce an alternative to the *Sambridge et al.* (1991) algorithm, which involves a projection onto the subspace spanned by the model parameters. Instead, we will consider a strategy that involves projecting the gradient and Hessian of the objective function onto the subspace spanned by the earthquakes; hence the phrase 'source subspace projection'. By performing projections in the data space the new approach differs from all earlier applications of subspace methods in seismology, which carried out projections in the model space (e.g. *Kennett et al.*, 1988; *Sambridge*, 1990; *Rawlinson et al.*, 2001). The resulting source-projected Hessian is still manageable, having the dimension of the number of earthquakes, which will be in the hundreds to thousands.

We compare the convergence rate of the classical conjugate gradient method with that of the source subspace projection algorithm by repeating some of the 2D experiments presented by *Tape et al.* (2007). We demonstrate that the source subspace projection algorithm involves only minor modifications of the classical conjugate gradient method, but that the source subspace projection algorithm converges two to three times faster. The conjugate gradient approach involves the determination of a trial model in the (conjugate) gradient direction, followed by quadratic or cubic interpolation to determine the minimum of the misfit function in the search direction. The calculation and associated storage and evaluation of this trial model is avoided in the source subspace projection algorithm, thus saving considerable compute time, I/O, and storage.

4.2 Classical least-squares solutions

To set the stage, and to introduce the necessary notation, we begin by considering the classical least-squares solution to an inverse problem (*Tarantola*, 2005). Let m_0 denote a reference *a priori* model and m a new model; these are M-dimensional vectors. The prior $M \times M$ symmetric, positive-definite model covariance matrix is denoted by $\mathsf{C}_{\mathsf{m}_0}$. The N-dimensional data vector is denoted by d, and the associated $N \times N$ symmetric, positive-definite data covariance matrix is denoted by C_d. The prediction for the current model is represented by the N-dimensional vector $\mathsf{g}(\mathsf{m})$.

Following *Tarantola* (2005), consider the *a posteriori* probability density function in the model space:

$$\sigma_\mathsf{m} = \text{const. } \exp[-F(\mathsf{m})], \tag{4.1}$$

where

$$2F(\mathsf{m}) = [\mathsf{g}(\mathsf{m}) - \mathsf{d}]^\mathsf{T} \mathsf{C}_\mathsf{d}^{-1} [\mathsf{g}(\mathsf{m}) - \mathsf{d}] + (\mathsf{m} - \mathsf{m}_0)^\mathsf{T} \mathsf{C}_{\mathsf{m}_0}^{-1} (\mathsf{m} - \mathsf{m}_0). \tag{4.2}$$

If the function $\mathsf{g}(\mathsf{m})$ can be linearized around m_0, we may write

$$\mathsf{g}(\mathsf{m}) \approx \mathsf{g}(\mathsf{m}_0) + \mathsf{G}(\mathsf{m} - \mathsf{m}_0), \tag{4.3}$$

where G denotes the $N \times M$ partial derivative matrix

$$\mathsf{G} = \frac{\partial \mathsf{g}}{\partial \mathsf{m}}. \tag{4.4}$$

Now let us introduce the notation

$$\Delta\mathsf{m} = \mathsf{m} - \mathsf{m}_0, \tag{4.5}$$

$$\Delta\mathsf{d} = \mathsf{d} - \mathsf{g}(\mathsf{m}_0). \tag{4.6}$$

Then to first order in $\Delta\mathsf{m}$ and $\Delta\mathsf{d}$ (4.2) becomes

$$2F \approx (\mathsf{G}\Delta\mathsf{m} - \Delta\mathsf{d})^\mathsf{T} \mathsf{C}_\mathsf{d}^{-1} (\mathsf{G}\Delta\mathsf{m} - \Delta\mathsf{d}) + \Delta\mathsf{m}^\mathsf{T} \mathsf{C}_{\mathsf{m}_0}^{-1} \Delta\mathsf{m}, \tag{4.7}$$

and thus the *a posteriori* probability density function is approximately Gaussian, such that (*Tarantola*, 2005)

$$\Delta \mathsf{m} = (\mathsf{G}^\mathsf{T}\mathsf{C}_\mathsf{d}^{-1}\mathsf{G} + \mathsf{C}_{\mathsf{m}_0}^{-1})^{-1}\mathsf{G}^\mathsf{T}\mathsf{C}_\mathsf{d}^{-1}\Delta \mathsf{d} = \mathsf{C}_{\mathsf{m}_0}\mathsf{G}^\mathsf{T}(\mathsf{G}\mathsf{C}_{\mathsf{m}_0}\mathsf{G}^\mathsf{T} + \mathsf{C}_\mathsf{d})^{-1}\Delta \mathsf{d}. \tag{4.8}$$

Using a quasi-Newton method, we may use the iterative algorithm (*Tarantola*, 2005)

$$\begin{aligned} \mathsf{m}_{n+1} &= \mathsf{m}_n + \lambda_n(\mathsf{G}_n^\mathsf{T}\mathsf{C}_\mathsf{d}^{-1}\mathsf{G}_n + \mathsf{C}_{\mathsf{m}_0}^{-1})^{-1}(\mathsf{G}_n^\mathsf{T}\mathsf{C}_\mathsf{d}^{-1}\Delta \mathsf{d}_n - \mathsf{C}_{\mathsf{m}_0}^{-1}\Delta \mathsf{m}_n) \\ &= \mathsf{m}_n - \lambda_n\Delta \mathsf{m}_n + \lambda_n\mathsf{C}_{\mathsf{m}_0}\mathsf{G}_n^\mathsf{T}(\mathsf{G}_n\mathsf{C}_{\mathsf{m}_0}\mathsf{G}_n^\mathsf{T} + \mathsf{C}_\mathsf{d})^{-1}(\Delta \mathsf{d}_n + \mathsf{G}_n\Delta \mathsf{m}_n), \end{aligned} \tag{4.9}$$

where $\lambda_n \approx 1$, $\Delta \mathsf{d}_n = \mathsf{d} - \mathsf{g}(\mathsf{m}_n)$, and $\Delta \mathsf{m}_n = \mathsf{m}_n - \mathsf{m}_0$.

4.3 Source subspace projection method

In the source subspace projection approach, we project the problem onto the subspace spanned by the sources as follows. We partition the data vector as follows:

$$\Delta \mathsf{d} = (\Delta d_i,\ i = 1, N) = \{(\Delta d_{sp},\ p = 1, N_s)\,,\ s = 1, S\}\,, \tag{4.10}$$

where S is the total number of sources, N_s is the number of measurements per source, and Δd_{sp} is the pth measurement for source s. Then the total number of data, N, is defined in terms of the number of sources, S, and the number of picks per source, N_s, by

$$N = \sum_{s=1}^{S} N_s. \tag{4.11}$$

Because the data covariance matrix C_d is symmetric and positive-definite we can define its square root, which will be denoted by $\mathsf{C}_\mathsf{d}^{1/2}$, and its inverse by $\mathsf{C}_\mathsf{d}^{-1/2}$. We will assume that the data covariance matrix C_d is block diagonal, with S symmetric positive-definite blocks $\mathsf{C}_{\mathsf{d}s}$ of size $N_s \times N_s$. We define a set of S orthonormal N-dimensional vectors

$$\mathsf{p}_s^\mathsf{T} = (0 \cdots 0\, \Delta\overline{d}_{s1} \cdots \Delta\overline{d}_{sN_s}\, 0 \cdots 0), \tag{4.12}$$

where the N_s-dimensional vector $\Delta\bar{\mathsf{d}}_s^{\mathsf{T}} = (\Delta\bar{d}_{s1} \cdots \Delta\bar{d}_{sN_s})$ is determined by

$$\Delta\bar{\mathsf{d}}_s = \mathsf{C}_{\mathsf{d}s}^{-1/2}\Delta\mathsf{d}_s. \tag{4.13}$$

If we further assume that the data covariance matrix $\mathsf{C}_{\mathsf{d}s}$ associated with source s is diagonal with elements σ_{sp}^2, which implies that $\mathsf{C}_{\mathsf{d}s}^{1/2}$ is also diagonal with elements σ_{sp}, then (4.13) implies

$$\Delta\bar{d}_{sp} = \Delta d_{sp}/\sigma_{sp}. \tag{4.14}$$

It is easily shown that

$$\mathsf{p}_s^{\mathsf{T}}\mathsf{p}_{s'} = \delta_{ss'}\sum_{p=1}^{N_s}(\Delta d_{sp}/\sigma_{sp})^2. \tag{4.15}$$

We now define the $S \times N$ projection operator P by

$$\mathsf{P}^{\mathsf{T}} = (\mathsf{p}_1 \cdots \mathsf{p}_S). \tag{4.16}$$

We will see in what follows that this choice of projection operator fits beautifully with the adjoint approach to calculating the gradient of an objective function.

In the source subspace projection method, we consider the *a posteriori* model space probability density function

$$\tilde{\sigma}_{\mathsf{m}} = \text{const. } \exp[-\tilde{F}(\tilde{\mathsf{m}})], \tag{4.17}$$

where

$$2\tilde{F} \approx [\mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}(\mathsf{G}\Delta\tilde{\mathsf{m}} - \Delta\mathsf{d})]^{\mathsf{T}}[\mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}(\mathsf{G}\Delta\tilde{\mathsf{m}} - \Delta\mathsf{d})] + \Delta\tilde{\mathsf{m}}^{\mathsf{T}}\mathsf{C}_{\mathsf{m}_0}^{-1}\Delta\tilde{\mathsf{m}}. \tag{4.18}$$

Note that in comparison to the classical expression (4.2) this amounts to using an inverse data covariance matrix $\mathsf{C}_{\mathsf{d}}^{-1/2}\mathsf{P}^{\mathsf{T}}\mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}$, rather than $\mathsf{C}_{\mathsf{d}}^{-1}$. In terms of the $S \times M$ 'projected gradient'

$$\tilde{\mathsf{G}} \equiv \mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}\mathsf{G}, \tag{4.19}$$

and the 'projected data vector'

$$\Delta\tilde{\mathsf{d}} = \mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}\Delta\mathsf{d}, \tag{4.20}$$

we have

$$2\tilde{F} \approx (\tilde{\mathsf{G}}\Delta\tilde{\mathsf{m}} - \Delta\tilde{\mathsf{d}})^{\mathsf{T}}(\tilde{\mathsf{G}}\Delta\tilde{\mathsf{m}} - \Delta\tilde{\mathsf{d}}) + \Delta\tilde{\mathsf{m}}^{\mathsf{T}}\mathsf{C}_{\mathsf{m}_0}^{-1}\Delta\tilde{\mathsf{m}}. \tag{4.21}$$

Again the *a posteriori* probability density function is approximately Gaussian, such that

$$\Delta\tilde{\mathsf{m}} = (\tilde{\mathsf{G}}^{\mathsf{T}}\tilde{\mathsf{G}} + \mathsf{C}_{\mathsf{m}_0}^{-1})^{-1}\tilde{\mathsf{G}}^{\mathsf{T}}\Delta\tilde{\mathsf{d}} = \mathsf{C}_{\mathsf{m}_0}\tilde{\mathsf{G}}^{\mathsf{T}}(\tilde{\mathsf{G}}\mathsf{C}_{\mathsf{m}_0}\tilde{\mathsf{G}}^{\mathsf{T}} + \mathsf{I})^{-1}\Delta\tilde{\mathsf{d}}. \tag{4.22}$$

Note that compared to (4.8) the projected data covariance matrix

$$\tilde{\mathsf{C}}_{\mathsf{d}} = (\mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2})\mathsf{C}_{\mathsf{d}}(\mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2})^{\mathsf{T}} = \mathsf{I}, \tag{4.23}$$

has become the $S \times S$ identity matrix, because the data covariance matrix C_{d} is absorbed in the definition (4.19) of G. Note also that the model update (4.22) only requires the inversion of a positive-definite $S \times S$ matrix.

Using a quasi-Newton method, we may use the iterative algorithm (*Tarantola*, 2005)

$$\begin{aligned}\tilde{\mathsf{m}}_{n+1} &= \tilde{\mathsf{m}}_n + \lambda_n(\tilde{\mathsf{G}}^{\mathsf{T}}\tilde{\mathsf{G}} + \mathsf{C}_{\mathsf{m}_0}^{-1})^{-1}(\tilde{\mathsf{G}}^{\mathsf{T}}\Delta\tilde{\mathsf{d}}_n - \mathsf{C}_{\mathsf{m}_0}^{-1}\Delta\tilde{\mathsf{m}}_n) \\ &= \tilde{\mathsf{m}}_n - \lambda_n\Delta\tilde{\mathsf{m}}_n + \lambda_n\mathsf{C}_{\mathsf{m}_0}\tilde{\mathsf{G}}_n^{\mathsf{T}}(\tilde{\mathsf{G}}_n\mathsf{C}_{\mathsf{m}_0}\tilde{\mathsf{G}}_n^{\mathsf{T}} + \mathsf{I})^{-1}(\Delta\tilde{\mathsf{d}}_n + \tilde{\mathsf{G}}_n\Delta\tilde{\mathsf{m}}_n),\end{aligned} \tag{4.24}$$

where $\lambda_n \approx 1$, $\Delta\tilde{\mathsf{d}}_n = \mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}\Delta\mathsf{d}_n$, and $\Delta\tilde{\mathsf{m}}_n = \tilde{\mathsf{m}}_n - \mathsf{m}_0$, to determine successive model updates. Because $(\tilde{\mathsf{G}}^{\mathsf{T}}\tilde{\mathsf{G}} + \mathsf{C}_{\mathsf{m}_0}^{-1})$ is an $M \times M$ matrix and $(\tilde{\mathsf{G}}_n\mathsf{C}_{\mathsf{m}_0}\tilde{\mathsf{G}}_n^{\mathsf{T}} + \mathsf{I})$ a generally much smaller $S \times S$ matrix, in practice we use the second equality in (4.24).

4.3.1 Significance of the source-projected gradient

In this section we investigate the significance of the source-projected partial derivative matrix $\tilde{\mathsf{G}}$ given by (4.19). To make the connection between this gradient and adjoint methods, let us consider a specific problem involving N cross-correlation traveltime anomalies ΔT_i, $i = 1, \ldots, N$, with associated standard deviations σ_i, $i = 1, \ldots, N$. Let us further assume that we are dealing with a structural inversion, and that the model m is expanded in M

basis functions $B_k(\mathbf{x})$, $k = 1, \ldots, M$, such that

$$\mathsf{m}(\mathbf{x}) = \sum_{k=1}^{M} m_k B_k(\mathbf{x}). \tag{4.25}$$

In this case the partial derivative matrix G has elements

$$G_{ik} = \frac{\partial T_i}{\partial m_k} = \int_V K_i(\mathbf{x}) B_k(\mathbf{x}) \, \mathrm{d}^3\mathbf{x}, \tag{4.26}$$

where V denotes the model volume and $K_i(\mathbf{x})$ the finite-frequency sensitivity kernel associated with observation i (e.g., *Dahlen et al.*, 2000; *Tromp et al.*, 2005).

In the particular case of cross-correlation traveltime anomalies, the source subspace projection operator P is given by (4.16), where

$$\Delta\overline{d}_{sp} = \Delta T_{sp}/\sigma_{sp}. \tag{4.27}$$

It is now straightforward to show that the source-projected data vector $\Delta\tilde{\mathsf{d}}$ (4.20) has elements

$$(\Delta\tilde{\mathsf{d}})_s = \sum_{p=1}^{N_s} (\Delta T_{sp}/\sigma_{sp})^2, \tag{4.28}$$

and that the elements of the source-projected gradient $\tilde{\mathsf{G}}$ (4.19) are given by

$$(\tilde{\mathsf{G}})_{sk} = (\mathsf{P}\mathsf{C}_{\mathsf{d}}^{-1/2}\mathsf{G})_{sk} = \int_V \left[\sum_{p=1}^{N_s} (\Delta T_{sp}/\sigma_{sp}^2) K_{sp}(\mathbf{x}) \right] B_k(\mathbf{x}) \, \mathrm{d}^3\mathbf{x} = -\int_V K_s(\mathbf{x}) B_k(\mathbf{x}) \, \mathrm{d}^3\mathbf{x}, \tag{4.29}$$

where we have defined the event kernel (*Tromp et al.*, 2005; *Tape et al.*, 2007)

$$K_s(\mathbf{x}) = -\sum_{p=1}^{N_s} (\Delta T_{sp}/\sigma_{sp}^2) K_{sp}(\mathbf{x}). \tag{4.30}$$

These kernels are calculated based upon the interaction between the regular wavefield $\mathbf{s}$ and

an adjoint wavefield $\mathbf{s}^\dagger$ that is generated by the adjoint source

$$\mathbf{f}_s^\dagger(\mathbf{x},t) = -\sum_{p=1}^{N_s}(\Delta T_{sp}/\sigma_{sp}^2)\frac{1}{M_{sp}}\,\partial_t \mathbf{s}_{sp}(T-t)]\delta(\mathbf{x}-\mathbf{x}_{sp}), \tag{4.31}$$

where M_{sp} is a normalization factor, and $\mathbf{x}_{sp}$ denotes the receiver location associated with source s and pick p. The calculation of the event kernels involves only two numerical simulations per earthquake.

4.3.2 Comparison with the conjugate gradient method

In a traditional conjugate gradient method, the first model update is in the opposite direction of the gradient of the cross-correlation traveltime misfit function (e.g., *Tarantola*, 2005; *Tape et al.*, 2007), i.e.,

$$\Delta m_k \approx -\nu \sum_{k'=1}^{M}(\mathsf{C_m})_{kk'}\sum_{s=1}^{S}\int_V K_s(\mathbf{x})B_{k'}(\mathbf{x})\,\mathrm{d}^3\mathbf{x}, \tag{4.32}$$

where the scalar ν determines the step length and thus the location of the trial model. Note how the 'metric' $\mathsf{C_m}$ turns the dual vector $\hat{\gamma}_{k'} = \int_V K_s(\mathbf{x})B_{k'}(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$, i.e., the gradient, into a vector: $\gamma = \mathsf{C_m}\hat{\gamma}$ (see e.g., *Tarantola*, 2005). Upon comparing this expression with the source subspace projection result (4.22), i.e.,

$$\Delta\tilde{m}_k \approx \sum_{k'=1}^{M}\sum_{s=1}^{S}(\mathsf{C_m})_{kk'}(\tilde{\mathsf{G}})_{sk'}\Delta\mu_s = -(\mathsf{C_m})_{kk'}\sum_{s=1}^{S}\Delta\mu_s\int_V K_s(\mathbf{x})B_{k'}(\mathbf{x})\,\mathrm{d}^3\mathbf{x}, \tag{4.33}$$

where the S-dimensional vector $\Delta\mu$ is determined by

$$\Delta\mu = (\tilde{\mathsf{G}}\mathsf{C}_{\mathsf{m}_0}\tilde{\mathsf{G}}^{\mathsf{T}} + \mathsf{I})^{-1}\Delta\tilde{\mathsf{d}}, \tag{4.34}$$

we see how the source subspace projection method 'preconditions' the model update by combining the event Fréchet derivatives $\int_V K_s(\mathbf{x})B_{k'}(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$ with weights $\Delta\mu_s$.

4.4 2D synthetic experiments

In Figures 4.1–4.3 we compare the source-subspace (SS) inversion with a conjugate-gradient (CG) inversion. In Figure 4.2, $\mathbf{m}_{01}$ for SS (d) is much closer to the target model that the CG version (a). This can be seen visually, as well as in the misfit values in (h). A key distinction is that the CG models require an additional evaluation of the misfit function at each step (e.g., *Tape et al.*, 2007). At $\mathbf{m}_{02}$, for example, CG has used 7 forward simulations, while SS has used only 5.

Figure 4.3 demonstrates that the SS and CG algorithms perform similarly for the case of a three-parameter inversion for location and origin time.

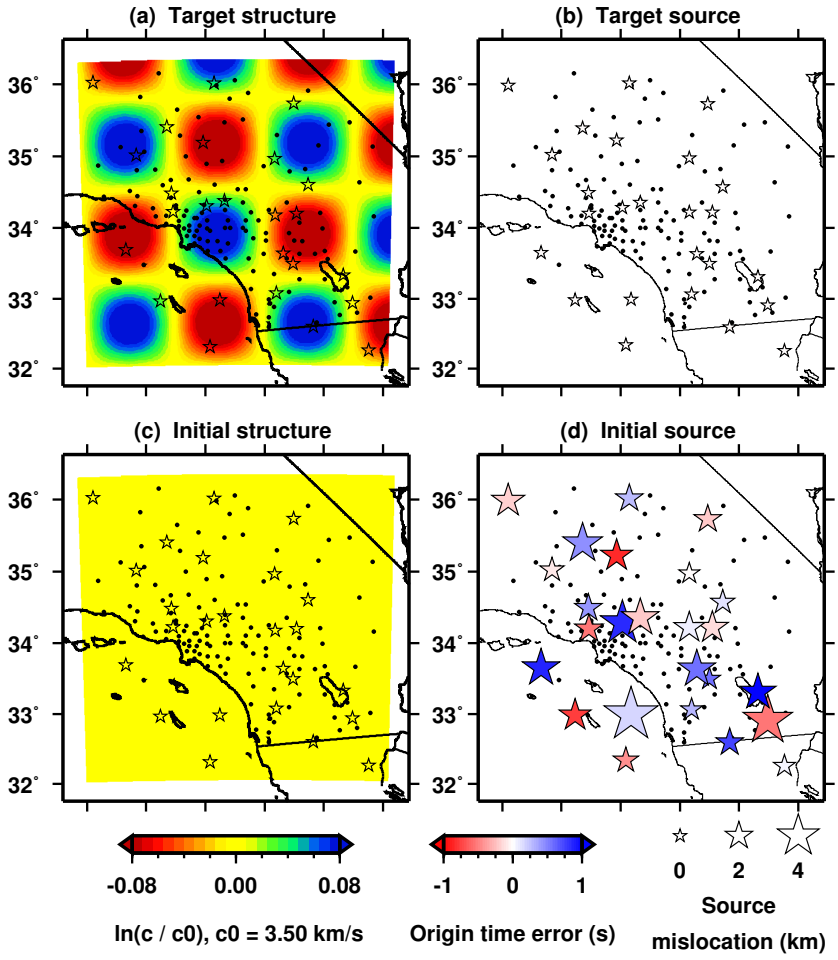


Figure 4.1: Initial and target structure and sources for subspace experiments. Target synthetic seismograms are generated using the target structure or target sources. Initial synthetics are generated using the initial structure or initial sources. Through iterative minimization of a misfit function, the initial model moves in the direction of the target model.

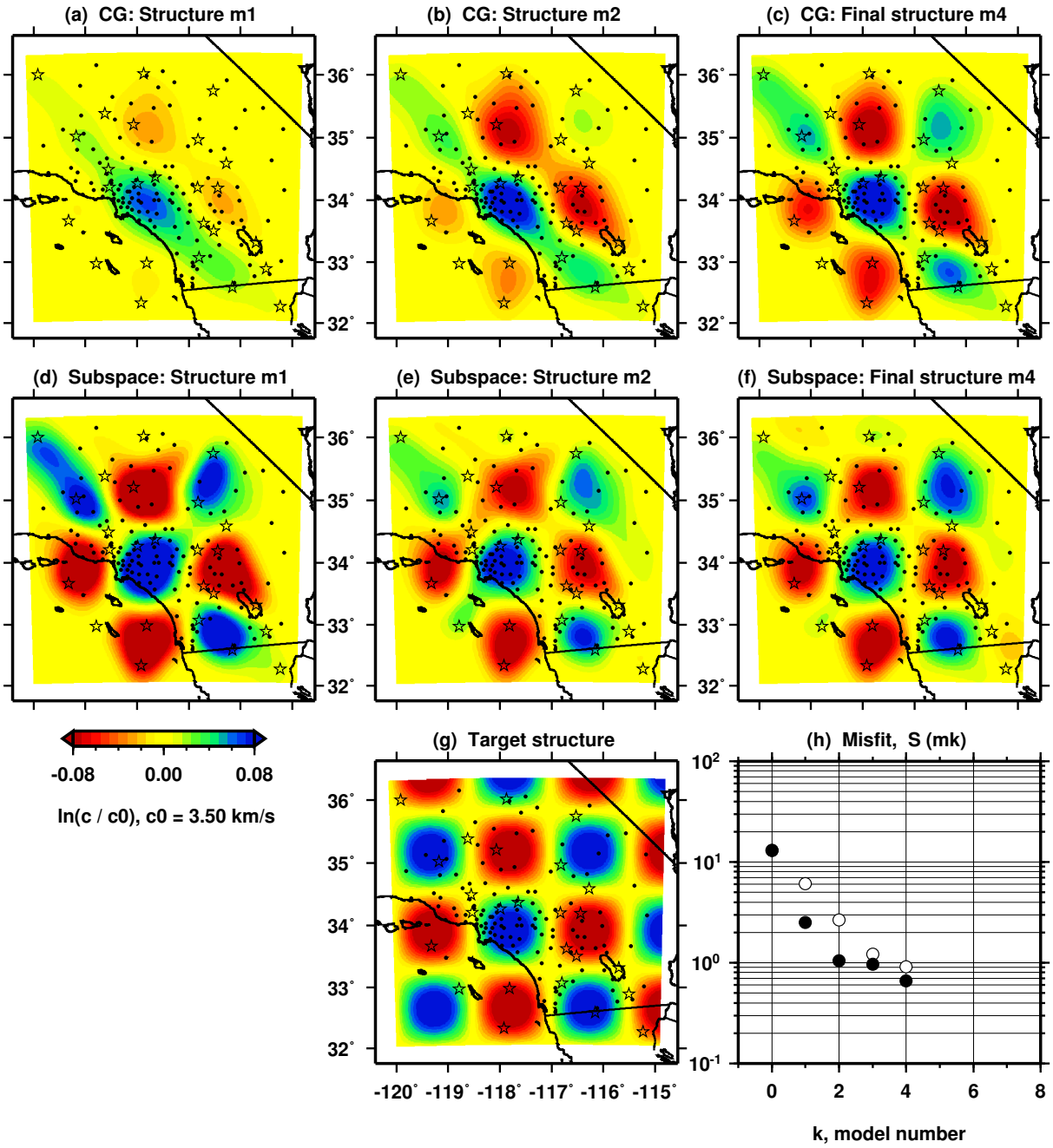


Figure 4.2: Comparison between conjugate-gradient (CG) and source-subspace synthetic inversions for structure parameters. (a)–(c) Conjugate-gradient models $\mathbf{m}_{01}$, $\mathbf{m}_{02}$, and $\mathbf{m}_{04}$. (d)–(f) Source-subspace (SS) models $\mathbf{m}_{01}$, $\mathbf{m}_{02}$, and $\mathbf{m}_{04}$. (g) Target structure. (h) Misfit function evaluations for each model. White circles are for CG models; black circles are for SS models.

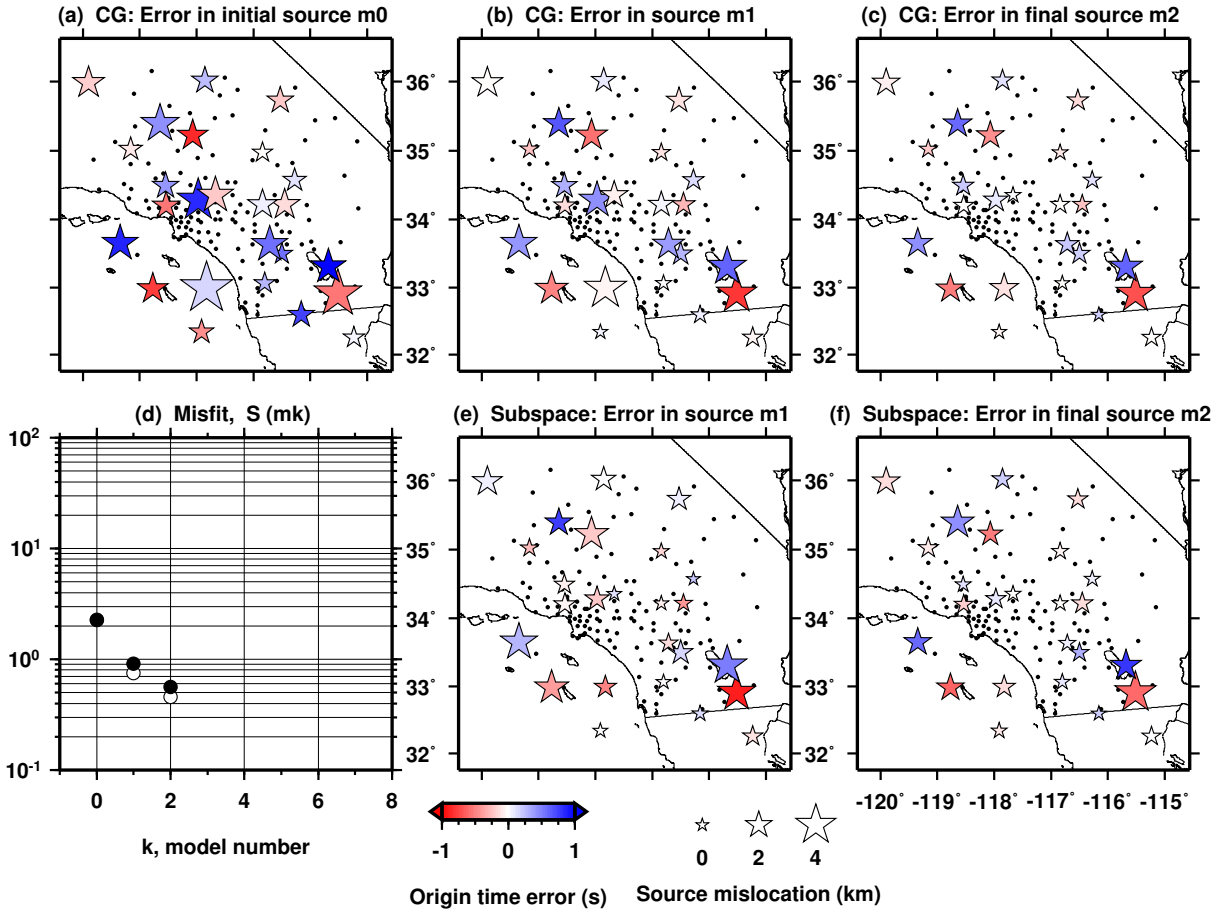


Figure 4.3: Comparison between conjugate-gradient and source-subspace synthetic inversions for source parameters. The three source parameters in the inversion experiment are (x_s, y_s) location and origin time (t_s). (a) Initial source errors for both the conjugate-gradient (CG) and source-subspace (SS) inversions. (b)–(c) CG source errors for models $\mathbf{m}_{01}$ and $\mathbf{m}_{02}$. (d) Reduction in misfit for CG (white circles) and SS (black circles). The performance of CG and SS is essentially the same for the source inversion.

Chapter 5

An automated time-window selection algorithm for seismic tomography

Note

This chapter contains excerpts from a published paper entitled "An automated time-window selection algorithm for seismic tomography," by Alessia Maggi, Carl Tape, Min Chen, Daniel Chao, and Jeroen Tromp. A. Maggi devised the algorithm and wrote the code, and I was involved in testing and refining some parts of the code. A. Maggi, M. Chen, and I tested the code using three different data sets: global earthquakes (A. Maggi), regional earthquakes from the Japan subduction zone (M. Chen), and crustal earthquakes in southern California (C. Tape). D. Chao (Caltech) did a Summer Undergraduate Research Experience project using and refining the algorithm. His work provided a good starting point for determining a set of user parameters, required by the algorithm, for the southern California data set.

The open-source algorithm FLEXWIN is available for download from the webpage of the Computational Infrastructure for Geodynamics (CIG): `www.geodynamics.org`.

Summary

We present FLEXWIN, an open-source algorithm for the automated selection of time windows on pairs of observed and synthetic seismograms. The algorithm was designed specifically to accommodate synthetic seismograms produced from 3D wavefield simulations, which capture complex phases that do not necessarily exist in 1D simulations or traditional traveltime curves. Relying on signal processing tools and several user-tuned parameters, the algorithm is able to include these new phases and to maximize the number of measurements made on each seismic record, while avoiding seismic noise. Our motivation is to use the algorithm for iterative tomographic inversions, in which the synthetic seismograms change from one iteration to the next. Hence, automation is needed to handle the volume of measurements and to allow for an increasing number of windows at each model iteration. The algorithm is sufficiently flexible to be adapted to many tomographic applications and seismological scenarios, including those based on synthetics generated from 1D models. We illustrate the algorithm using datasets from three distinct regions: the entire globe, the Japan subduction zone, and southern California.

5.1 The selection algorithm

Our open-source algorithm, called FLEXWIN to reflect its FLEXibility in picking time WINdows for measurement, operates on pairs of observed and synthetic single component seismograms. The window selection process has five stages, each of which is discussed in *Maggi et al.* (2009): *Stage A:* preprocessing; *Stage B:* definition of preliminary measurement windows; *Stage C:* rejection of preliminary windows based on the content of the synthetic seismogram alone; *Stage D:* rejection of preliminary windows based on the differences between observed and synthetic seismograms; *Stage E:* resolution of preliminary window overlaps. The parameters that permit tuning of the window selection toward a specific tomographic scenario are all contained in a simple parameter file (see Table 5.1). More complexity and finer tuning can be obtained by making some of these parameters time-dependent via user-defined functions that can depend on the source parameters (e.g., event location or depth).

An example of a synthetic seismogram and its corresponding envelope and STA:LTA timeseries $E(t)$ is shown in Figure 5.1. The $E(t)$ timeseries starts at its value for a constant

signal, then rises gradually due to the tapered low level numerical noise on the synthetic. At each seismic arrival, $E(t)$ rises to a local maximum. We can see from Figure 5.1 that these local maxima correspond both in position and in width to the seismic phases in the synthetic, and that the local minima in $E(t)$ correspond to the transitions between one phase and the next. In the following sections we shall explain how we use these correspondences to define time windows.

Figure 5.2 shows the reduction of candidate windows for the seismogram in Figure 5.1.

5.2 Windowing Examples

We present a set of examples showing the results of the FLEXWIN algorithm applied to real data. These examples illustrate the robustness and flexibility of the algorithm. We have applied the algorithm to three tomographic scenarios, with very different geographical extents and distinct period ranges: long-period global tomography (50–150 s), regional tomography of the Japan subduction zone, down to 700 km (6–120 s), and regional tomography of southern California, down to 60 km (2–30 s). For each of these scenarios, we compare observed seismograms to spectral-element synthetics, using our algorithm to select time windows on the pairs of timeseries.

The windowing algorithm itself has little prior knowledge of seismology, other than in the most general terms: it considers a seismogram to be a succession of seismic phases indicated by changes in amplitude and frequency of the signal with time; it is based upon the idea that the short-term to long-term average ratio STA:LTA is a good indicator of the arrival of such phases; it has a notion of the characteristics of an optimal set of data windows. All other prior information — the frequency range to be considered, the portions of the seismogram to be excluded, the acceptable signal-to-noise ratios, the tolerance of dissimilarity between the observed and synthetic seismogram — varies greatly between any two seismological studies. In order to ensure maximum flexibility of our windowing algorithm, all such scenario-dependent information is encapsulated in the tuning parameters of Table 5.1.

We tuned the windowing algorithm separately for each of the three scenarios we present here, and we present examples based on the events listed in Table 5.3. Tuning parameter values for each scenario can be found in Table 5.2, while the functional forms of the time-

dependent parameters can be found in Section 5.3.1. Once tuned for a given scenario, the algorithm is applied to all its events without further modification.

Local tomography in Southern California

Our last scenario is a local tomographic study of southern California. We apply the windowing algorithm to a set of 140 events within southern California, for which we have computed synthetic seismograms using the spectral-element method and a regional 3D crustal and upper mantle model (*Komatitsch et al.*, 2004). This model contains three discontinuities: the surface topography (included in the mesh), the basement layer that separates the sedimentary basins from the bedrock, and the Moho, separating the lower crust from the upper mantle. The model includes several sedimentary basins, such as the Ventura basin, the Los Angeles basin, and the Salton trough (*Komatitsch et al.*, 2004; *Lovely et al.*, 2006). The smooth 3D background velocity model used in *Komatitsch et al.* (2004) was determined by *Hauksson* (2000); we use an updated version provided by *Lin et al.* (2007b). The physical domain of the model is approximately 600 km by 500 km at the surface, and extends to a depth of 60 km. Our simulations of seismic waves are numerically accurate down to a period of 2 s.

The 140 events have M_{w} magnitudes between 3.5 and 5.5 and were recorded between 1999 and 2007. The locations and origin times are primarily from *Lin et al.* (2007a), and the focal mechanisms are from *Clinton et al.* (2006), *Hardebeck and Shearer* (2003), or *Tan* (2006).

We test the windowing code using three period ranges: 6–30 s, 3–30 s, and 2–30 s. The parameters we use for the windowing code are listed in Table 5.2. Figures 5.3 and 5.4 show examples of the output from the windowing algorithm for event 9818433 listed in Table 5.3 recorded at two different stations, while Figure 5.5 shows a summary plot for event 9983429 in the 6–30 s period range.

The windowing algorithm tends to identify five windows on each set of three-component 6–30 s seismograms (Figures 5.3 and 5.5): on the vertical and radial components the first window corresponds to the body-wave arrival and the second to the Rayleigh wave, while windows on the transverse component capture the Love wave. The 2–30 s synthetic seismograms do not agree well with the observed seismograms, especially in the later part of the signal, leading to fewer picked windows. In Figure 5.3c, only three windows are selected by

the algorithm: the P arrival recorded on the radial component, the S arrival on the transverse component, and the Love-wave arrival on the transverse component. The P arrival (PmP or Pn) in fact appears on all three components on both data and synthetics. On the vertical component it is rejected because the cross-correlation value within the time window did not exceed the specified minimum value of 0.85 (Table 5.2). On the transverse component it does not have a large enough signal-to-noise ratio to be picked, but it is evident as a small peak at 36 s in the STA:LTA curve, and more conspicuous when zooming into the synthetics and data. The presence of the P arrival on the transverse component highlights the possibility of measuring subtle phases that may be present in 3D synthetics.

Figure 5.4 shows results for the same event as Figure 5.3, but for a different station, FMP, situated 52 km from the event and within the Los Angeles basin. Comparison of the two figures highlights the characteristic resonance caused by the thick sediments within the basin. This resonance is beautifully captured by the transverse component synthetics (Figure 5.4b, record T), thanks to the inclusion of the basin in the model (*Komatitsch et al.*, 2004). In order to pick such long time windows with substantial frequency-dependent measurement differences, we are forced to lower the minimum cross-correlation value CC_0 for the entire dataset (0.71 in Table 5.2) and increase c_{4b} to capture the slow decay in the STA:LTA curves (Figure 5.4b, record T). It is striking that although these arrivals look nothing like the energy packets typical for the global case, the windowing algorithm is still able to determine the proper start and end times for the windows. In Figure 5.4c the windowing algorithm selects three short-period body-wave time windows with superb agreement between data and synthetics.

5.3 Appendix A: Tuning considerations

FLEXWIN is not a black-box application, and as such cannot be applied blindly to any given dataset or tomographic scenario. The data windowing required by any given problem will differ depending on the inversion method, the scale of the problem (local, regional, global), the quality of the data set and that of the model and method used to calculate the synthetic seismograms. The user must configure and tune the algorithm for the given problem. In this appendix we shall discuss some general considerations the user should bear in mind during the tuning process. For more detailed information on tuning, and for further

examples of tuning parameter sets, we refer the reader to the user manual that accompanies the source code.

The order in which the parameters in Table 5.1 are discussed in the main text of this paper follows the order in which they are used by the algorithm, but is not necessarily the best order in which to consider them for tuning purposes. We suggest the following as a practical starting sequence (the process may need to be repeated and refined several times before converging on the optimal set of parameters for a given problem and dataset).

$T_{0,1}$: In setting the corner periods of the bandpass filter, the user is deciding on the frequency content of the information to be used in the tomographic problem. Values of these corner periods should reflect the information content of the data, the quality of the Earth model and the accuracy of the simulation used to generate the synthetic seismogram. The frequency content in the data depends on the spectral characteristics of the source, on the instrument responses, and on the attenuation characteristics of the medium. As $T_{0,1}$ depend on the source and station characteristics, which may be heterogeneous in any given dataset, these filter periods can be modified dynamically by constructing an appropriate user function (e.g.,*if station is in list of stations with instrument X then reset T0 and T1 to new values*).

$r_{P,A}$: In setting the signal-to-noise ratios for the entire seismogram the user is applying a simple quality control on the data. Note that these criteria are applied after filtering. No windows will be defined on data that fail this quality control.

$w_E(t)$: The short-term average long-term average ratio $E(t)$ of a constant signal converges to a constant value when the length of the time-series is greater than the effective averaging length of the long-term average. We suggest the user start with a constant level for $w_E(t)$ equal to this convergence value. The time dependence of $w_E(t)$ should then be adjusted to exclude those portions of the waveform the user is not interested in, by raising $w_E(t)$ (e.g., to exclude the fundamental mode surface-wave: *if* $t >$ *fundamental mode surface-wave arrival time then set* $w_E(t) = 1$). We suggest finer adjustments to $w_E(t)$ be made after $r0(t)$, $CC_0(t)$, $\Delta T_0(t)$ and $\Delta \ln A_0(t)$ have been configured.

$r_0(t)$, $\mathrm{CC}_0(t)$, $\Delta\tau_{\mathrm{ref}}$, $\Delta\tau_0(t)$, $\Delta \ln A_{\mathrm{ref}}$ and $\Delta \ln A_0(t)$: These parameters — window signal-to-noise ratio, normalized cross-correlation value between observed and synthetic seismograms, cross-correlation time lag, and amplitude ratio — control the degree of well-behavedness of the data within accepted windows (Stage D). The user first sets constant

values for these four parameters, then adds a time dependence if required. Considerations that should be taken into account include the quality of the Earth model used to calculate the synthetic seismograms, the frequency range, the dispersed nature of certain arrivals (e.g.,*for t corresponding to the group velocities of surface-waves, reduce $CC_0(t)$*), and *a priori* preferences for picking certain small-amplitude seismic phases (e.g.,*for t close to the expected arrival for $P_{\rm diff}$, reduce $r_0(t)$*). $\Delta\tau_{\rm ref}$ and $\Delta \ln A_{\rm ref}$ should be set to zero at first, and only reset if the synthetics contain a systematic bias in traveltimes or amplitudes.

c_{0-4} : These parameters are active in Stage C of the algorithm, the stage in which the suite of all possible data windows is pared down using criteria on the shape of the STA:LTA $E(t)$ waveform alone. We suggest the user start by setting these values to those used in our global example (see Table 5.2). Subsequent minimal tuning should be performed by running the algorithm on a subset of the data and closely examining the lists of windows rejected at each stage to make sure the user agrees with the choices made by the algorithm.

$w_{\rm CC}$, $w_{\rm len}$ and $w_{\rm nwin}$: These parameters control the overlap resolution stage of the algorithm (Stage E). Values of $w_{\rm CC} = w_{\rm len} = w_{\rm nwin} = 1$ should be reasonable for most applications.

The objective of the tuning process summarily described here should be to maximize the selection of windows around desirable features in the seismogram, while minimizing the selection of undesirable features, bearing in mind that the desirability or undesirability of a given feature is subjective, and depends on how the user subsequently intends to use the information contained within the data windows.

5.3.1 Examples of user functions for southern California

As concrete examples of how the time dependence of the tuning parameters can be exploited, we present here the functional forms of the time dependencies used for the southern California tomographic scenario described in the text (Section 5.2). We use predicted arrival times derived from 1D Earth models to help modulate certain parameters. Note, however, that the actual selection of individual windows is based on the details of the waveforms, and not on information from 1D Earth models.

In the following, t_P and t_S denote the start of the time windows for the crustal P wave and the crustal S wave, computed from a 1D layered model appropriate to Southern California (*Wald et al.*, 1995). The start and end times for the surface-wave time window,

t_{R0} and t_{R1}, as well as the criteria for the time shifts $\Delta\tau_0(t)$, are derived from formulas in *Komatitsch et al.* (2004). The source-receiver distance (in km) is denoted by Δ.

For the 6–30 s and 3–30 s data, we use constant values of $r_0(t) = r_0$, $\mathrm{CC}_0(t) = \mathrm{CC}_0$, $\Delta\tau_0(t) = \Delta\tau_0$, and $\Delta \ln A_0(t) = \Delta \ln A_0$. We exclude any arrivals before the P wave and after the Rayleigh wave. This is achieved by the box-car function for $w_E(t)$:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \le t \le t_{R1}, \\ 10w_E & t > t_{R1}. \end{cases} \tag{5.1}$$

For the 2–30 s data, we avoid selecting surface-wave arrivals as the 3D model used to calculate the synthetics cannot produce the required complexity. The water-level criteria then becomes:

$$w_E(t) = \begin{cases} 10w_E & t < t_P, \\ w_E & t_P \le t \le t_S, \\ 10w_E & t > t_S. \end{cases} \tag{5.2}$$

Table 5.1: Overview of standard tuning parameters, and of fine tuning parameters. Values are defined in a parameter file, and the time dependence of those that depend on time is described by user-defined functions.

Standard tuning parameters:	
$T_{0,1}$	bandpass filter corner periods
$r_{P,A}$	signal-to-noise ratios for whole waveform
$r_0(t)$	signal-to-noise ratios single windows
$w_E(t)$	water-level on short-term:long-term ratio
$\mathrm{CC}_0(t)$	acceptance level for normalized cross-correlation
$\Delta\tau_0(t)$	acceptance level for time lag
$\Delta \ln A_0(t)$	acceptance level for amplitude ratio
$\Delta\tau_{\mathrm{ref}}$	reference time lag
$\Delta \ln A_{\mathrm{ref}}$	reference amplitude ratio
Fine tuning parameters:	
c_0	for rejection of internal minima
c_1	for rejection of short windows
c_2	for rejection of unprominent windows
$c_{3a,b}$	for rejection of multiple distinct arrivals
$c_{4a,b}$	for curtailing of windows with emergent starts and/or codas
w_{CC} w_{len} w_{nwin}	for selection of best non-overlapping window combination

Table 5.2: Values of standard and fine-tuning parameters for the three seismological scenarios discussed in *Maggi et al.* (2009).

	Global	Japan		S. California		
$T_{0,1}$	50, 150	24, 120	6, 30	6, 30	3, 30	2, 30
$r_{P,A}$	3.5, 3.0	3.5, 3.0	3.5, 3.0	3.0, 2.5	2.5, 3.5	2.5, 3.5
r_0	2.5	1.5	3.0	3.0	4.0	4.0
w_E	0.08	0.10	0.12	0.18	0.11	0.07
CC_0	0.85	0.70	0.73	0.71	0.80	0.85
$\Delta\tau_0$	15	12.0	3.0	8.0	4.0	3.0
$\Delta \ln A_0$	1.0	1.0	1.5	1.5	1.0	1.0
$\Delta\tau_{\mathrm{ref}}$	0.0	0.0	0.0	4.0	2.0	1.0
$\Delta \ln A_{\mathrm{ref}}$	0.0	0.0	0.0	0.0	0.0	0.0
c_0	0.7	0.7	0.7	0.7	1.3	1.0
c_1	4.0	3.0	3.0	2.0	4.0	5.0
c_2	0.3	0.0	0.6	0.0	0.0	0.0
$c_{3a,b}$	1.0, 2.0	1.0, 2.0	1.0, 2.0	3.0, 2.0	4.0, 2.5	4.0, 2.5
$c_{4a,b}$	3.0, 10.0	3.0, 25.0	3.0, 12.0	2.5, 12.0	2.0, 6.0	2.0, 6.0
$w_{\mathrm{CC}}, w_{\mathrm{len}}, w_{\mathrm{nwin}}$	1, 1, 1	1, 1, 1	1, 1, 1	0.5,1.0,0.7	0.70,0.25,0.05	1,1,1

Table 5.3: Example events used in *Maggi et al.* (2009). The identifier refers to the CMT catalog for global events and Japan events, and refers to the Southern California Earthquake Data Center catalog for southern California events.

Identifier	Latitude	Longitude	Depth, km	Moment, N m	M_{w}	Location
			Global			
101895B	28.06	130.18	18.5	5.68e19	7.1	Ryukyu Islands
200808270646A	-10.49	41.44	24.0	4.68e17	5.7	Comoros Region
050295B	-3.77	-77.07	112.8	1.27e19	6.7	Northern Peru
060994A	-13.82	-67.25	647.1	2.63e21	8.2	Northern Bolivia
			Japan			
051502B	24.66	121.66	22.4	1.91e18	6.1	Taiwan
200511211536A	30.97	130.31	155.0	2.13e18	6.2	Kyushu, Japan
091502B	44.77	130.04	589.4	4.24e18	6.4	Northeastern China
			Southern California			
9983429	35.01	-119.14	13.5	9.19e15	4.6	Wheeler Ridge
9818433	33.91	-117.78	9.4	3.89e15	4.3	Yorba Linda

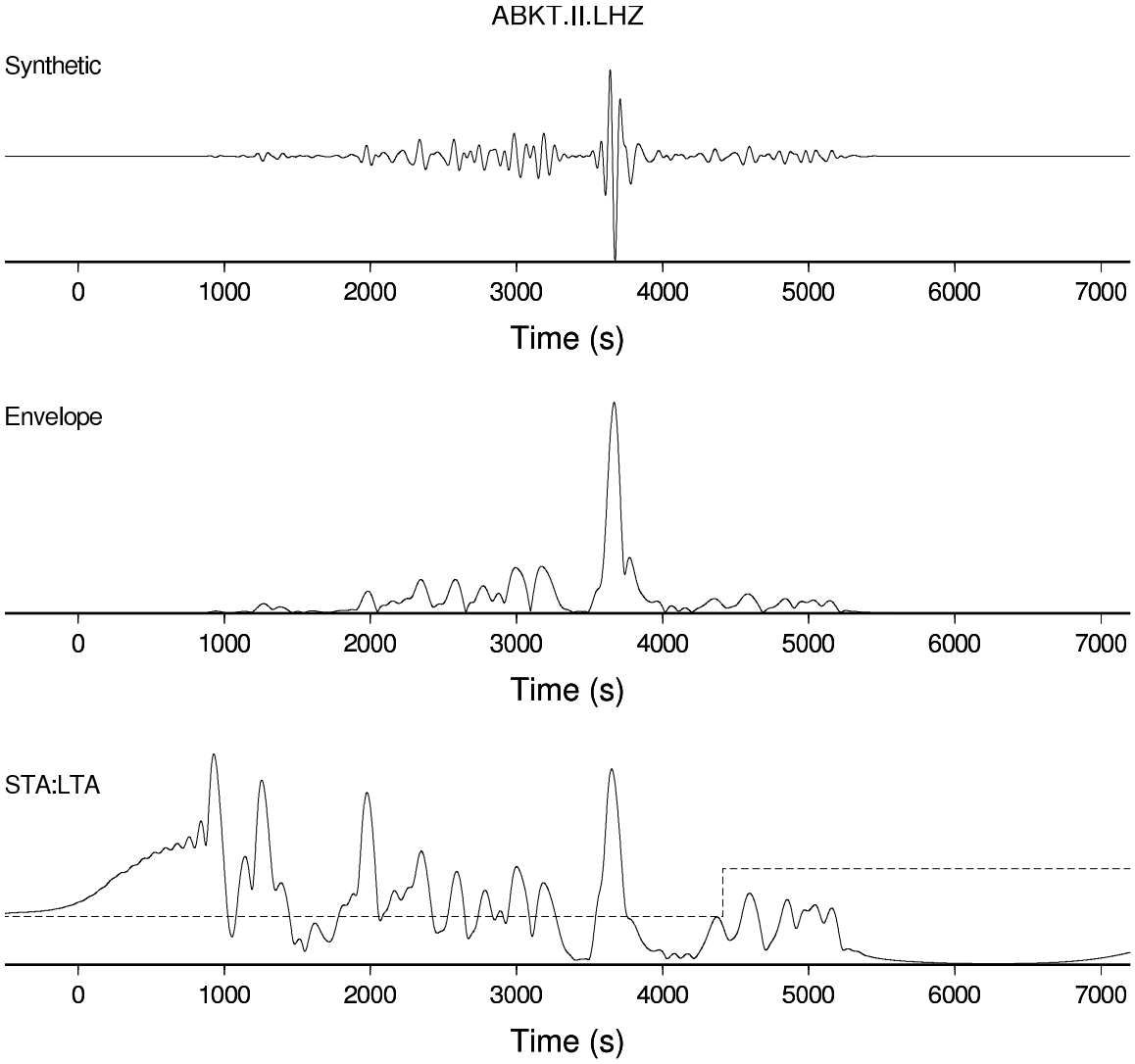


Figure 5.1: Synthetic seismogram and its corresponding envelope and STA:LTA timeseries (*Maggi et al.*, 2009, Figure 1). The seismogram was calculated using SPECFEM3D and the Earth model S20RTS (*Ritsema et al.*, 1999) for the CMT catalog event 050295B, whose details can be found in Table 5.3. The station, ABKT, is at an epicentral distance of 14100 km and at an azimuth of 44 degrees from the event. The top panel shows the vertical component synthetic seismogram, filtered between periods of 50 and 150 seconds. The center panel shows its envelope, and the bottom panel shows the corresponding STA:LTA waveform. The dashed line overlaid on the STA:LTA waveform is the water-level $w_E(t)$.

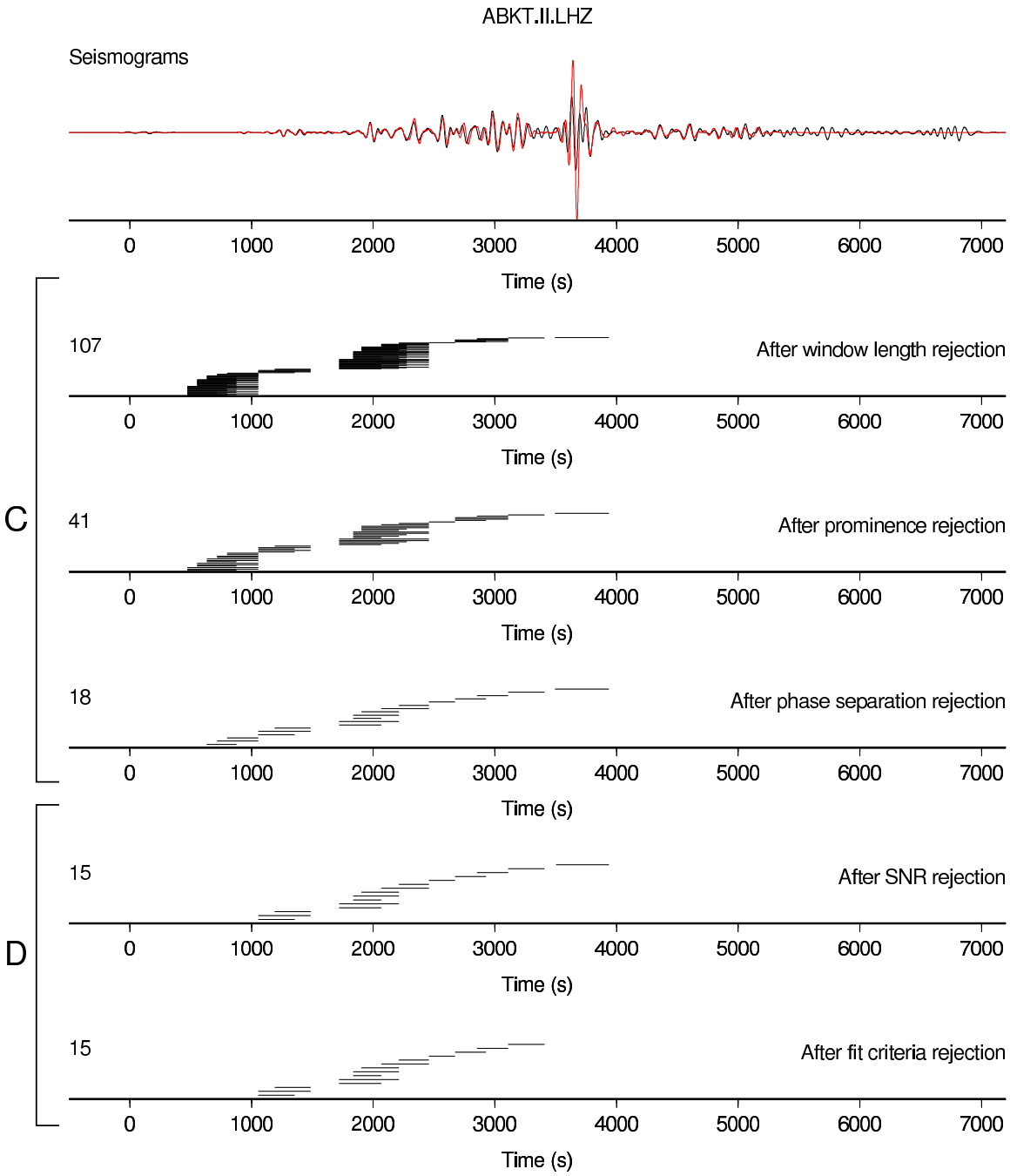


Figure 5.2: Window rejection applied to real data (*Maggi et al.*, 2009, Figure 4). Top panel: observed (black) and synthetic (red) seismograms for the 050295B event recorded at ABKT (see Figure 5.1). Subsequent panels: candidate windows at different stages, separated into Stage C (shape based rejection) and Stage D (fit based rejection). Each candidate window is indicated by a black segment. The number of windows at each stage is shown to the left of the panel.

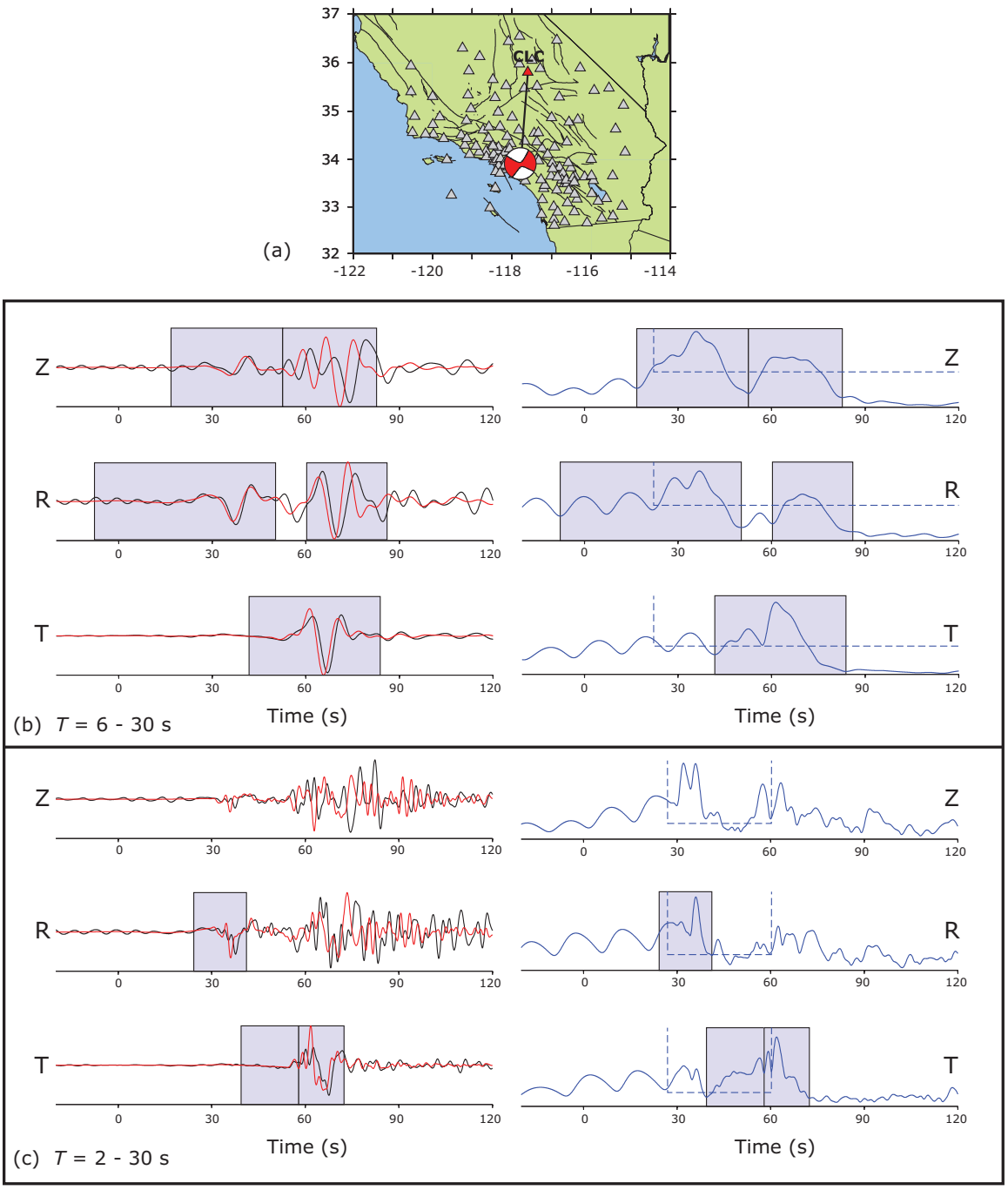


Figure 5.3: Window selection results for event 9818433 from Table 5.3 recorded at station CLC (Δ = 211.7 km) (*Maggi et al.*, 2009, Figure 15). (a) Map showing all stations with at least one measurement window for the period range 6–30 s for this event. Red triangle denotes station CLC. (b) Results for station CLC for the period range 6–30 s. Vertical (Z), radial (R), and transverse (T) records of data (black, left column) and synthetics (red, left column), as well as the STA:LTA records (right column) used to produce the window picks. (c) Results for station CLC for the period range 2–30 s. Note that corresponding lower-passed filtered versions are shown in (b).

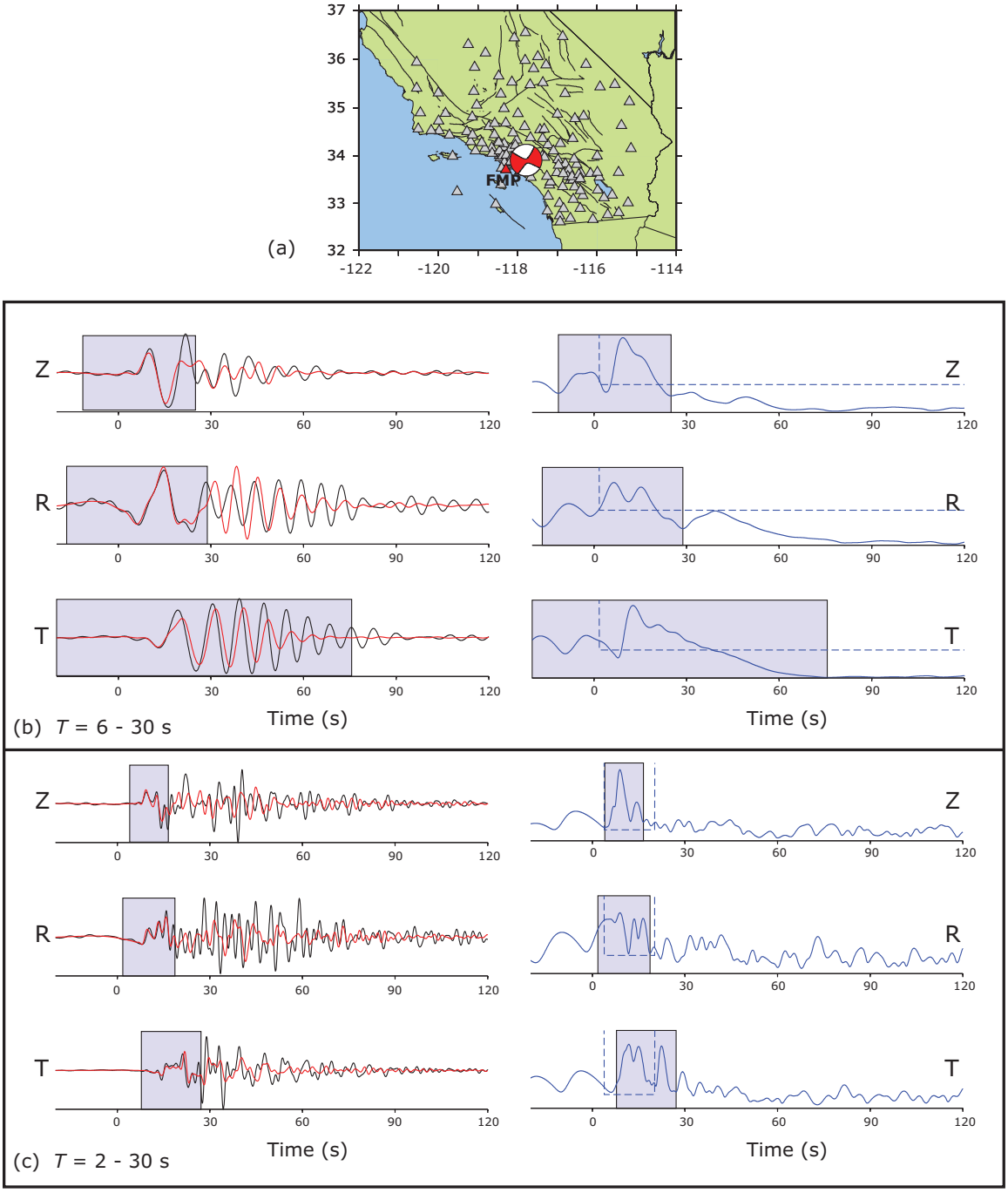


Figure 5.4: Window selection results for event 9818433 from Table 5.3 recorded at station FMP ($\Delta = 52.2$ km) (*Maggi et al.*, 2009, Figure 16). Same caption as Figure 5.3, but for a different station.

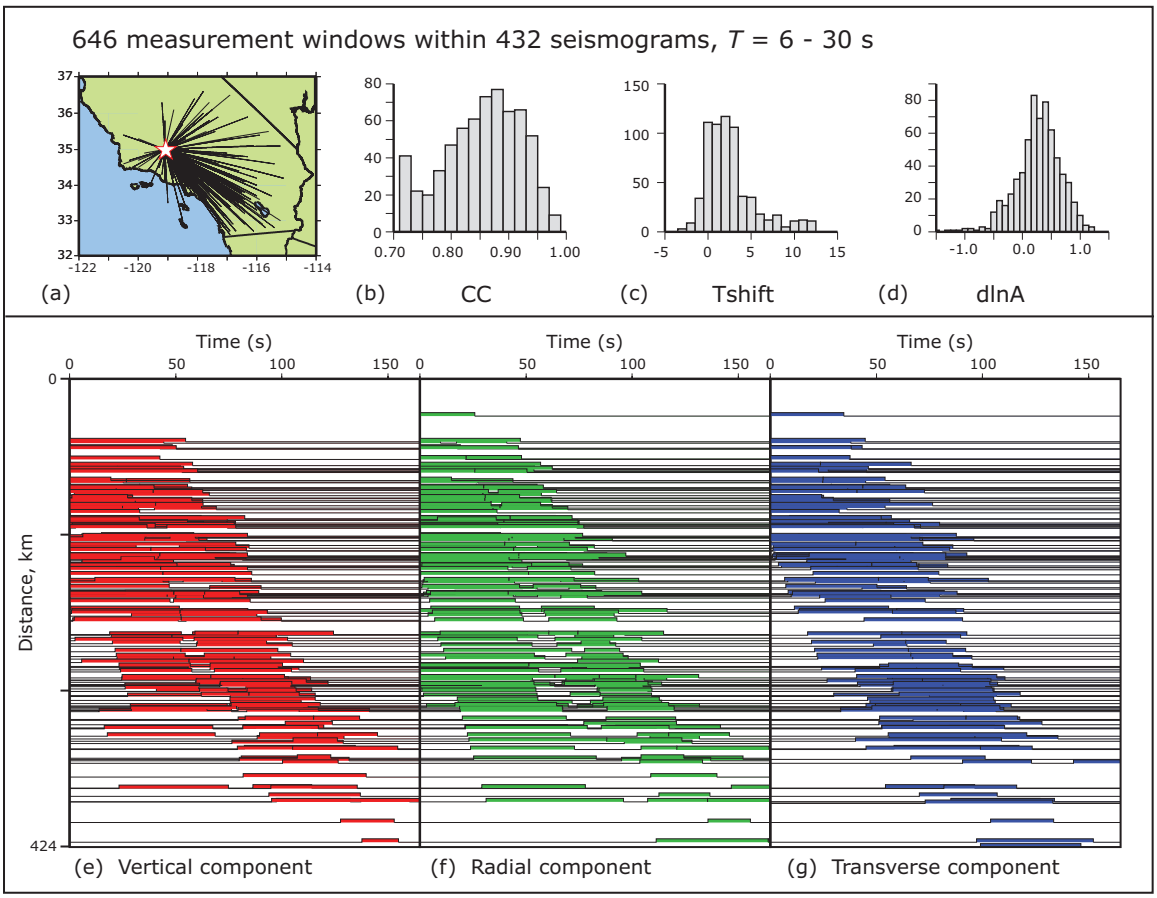


Figure 5.5: Summary plots of windowing results for event 9983429 in Table 5.3, for the period range 6–30 s (*Maggi et al.*, 2009, Figure 17). (a) Map showing paths to each station with at least one measurement window. (b)-(d) Histograms of number of windows as a function of normalized cross-correlation CC, time lag τ and amplitude ratio $\Delta \ln A$. (e)-(g) Record sections of selected windows for the vertical, radial and transverse components. The two branches observed on the vertical and radial components correspond to the body-wave arrivals and the Rayleigh wave arrivals.

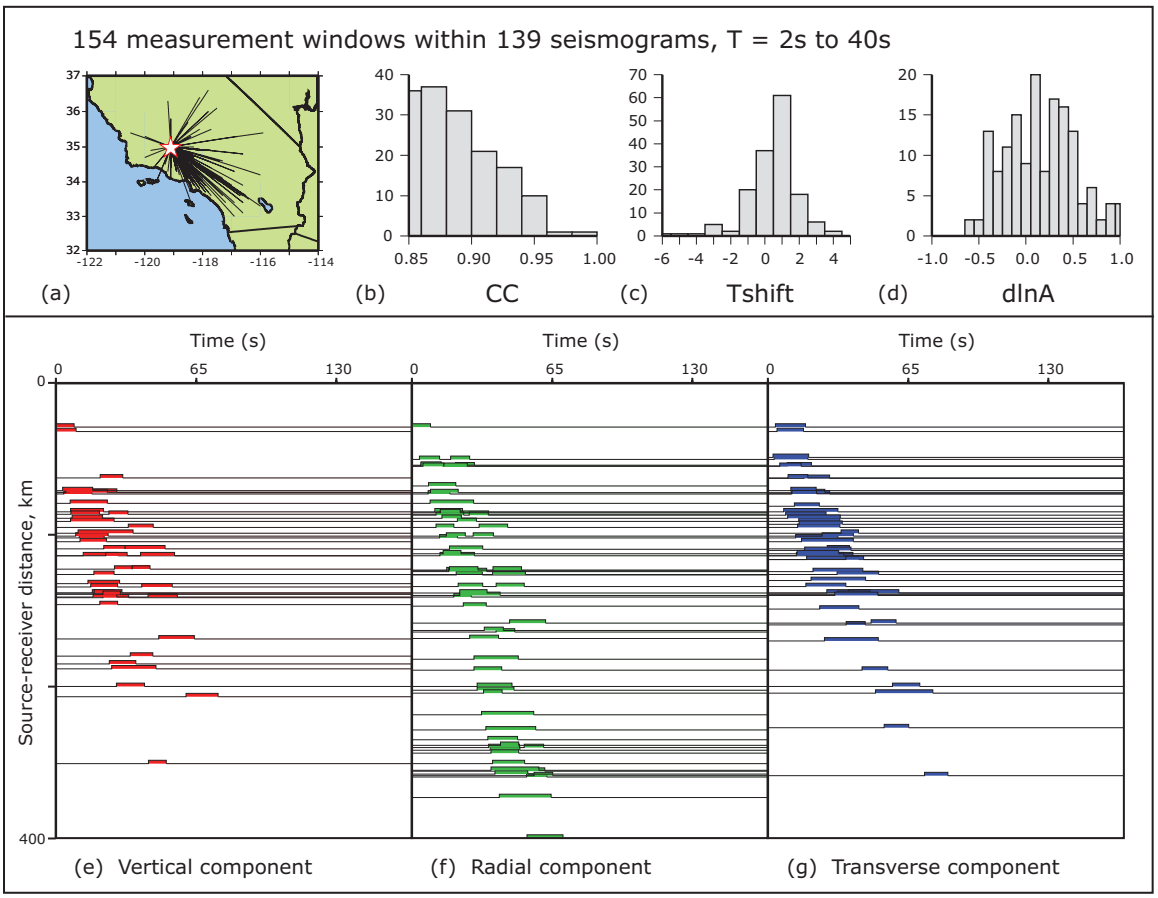


Figure 5.6: Summary plots of windowing results for event 9983429 in Table 5.3, for the period range 2–30 s. Same as Figure 5.5, but the windowing code has been run using a different set of parameters (Table 5.2), so that primarily only the body-wave arrivals are selected.

Chapter 6

Adjoint tomography of the southern California crust

Note

This chapter is a draft of a paper to be submitted to *Geophysical Journal International*. Authors are Carl Tape, Qinya Liu, Alessia Maggi, and Jeroen Tromp. This research was the primary focus of my thesis from January 2007 to April 2009. Synthetics for the first updated model ($\mathbf{m}_{01}$) were computed in May 2008, and synthetics for the final model ($\mathbf{m}_{16}$) were computed February 2009. Key preceding studies by each coauthor are *Tape et al.* (2007), *Liu and Tromp* (2006), *Maggi et al.* (2009), and *Tromp et al.* (2005).

Summary

We iteratively improve a three-dimensional seismological model of the southern California crust using an inversion strategy based upon adjoint methods. The resulting model involved 16 tomographic iterations, which required 6800 wavefield simulations and a total of 0.8 million CPU-hours. The new crustal model reveals strong heterogeneity, including local changes of $\pm 30\%$ with respect to the initial 3D model provided by the Southern California Earthquake Center. The improved crustal model illuminates shallow features with a close correspondence to surface geology, such as sedimentary basins. It also reveals crustal features at depth that aid in the tectonic reconstruction of southern California, such as subduction-captured oceanic crustal fragments. The new model enables more realistic and accurate assessments of seismic hazard.

6.1 Introduction

The objective of seismic tomography is to produce detailed three-dimensional (3D) images of Earth's interior by minimizing the differences between simulated (or "synthetic") seismograms and recorded (or "observed") seismograms. Seismic tomography has been successful in producing images of Earth's interior, such as large-scale variations in the mantle (*Woodhouse and Dziewonski*, 1984; *Romanowicz*, 2003), subducting slabs (*Grand et al.*, 1997), and mantle plumes (*Montelli et al.*, 2004). These tomographic studies adopt a simple one-dimensional (layered) reference model (*Dziewonski and Anderson*, 1981), which allows for computationally inexpensive procedures within the minimization problem. Highly accurate numerical methods, such as the spectral-element method (SEM), may now be used to compute synthetic seismograms at regional and global scales, allowing tomographers to start the minimization procedure with more realistic 3D initial models and simulations (*Komatitsch et al.*, 2002; *Akçelik et al.*, 2003; *Chen et al.*, 2007; *Fichtner et al.*, 2009). Furthermore, as demonstrated in this study, these numerical methods may be exploited within the minimization problem by using so-called adjoint methods (*Tarantola*, 1984; *Talagrand and Courtier*, 1987; *Tromp et al.*, 2005).

Southern California provides an excellent motivation and setting for the two-fold objective of fitting seismograms and characterizing the crust. The station coverage (Figure 6.2a), especially in the Los Angeles region, is one of the densest in the world. A detailed 3D seismological model of the southern California crust has been constructed from a variety of seismic datasets (*Komatitsch et al.*, 2004). Several different approaches have been used to determine earthquake source parameters. Finally, an accurate wave propagation code, employing the SEM, has been applied to simulate seismic wave propagation in the region (*Komatitsch et al.*, 2004), with recent modifications to facilitate an inverse problem (*Liu and Tromp*, 2006).

"Adjoint tomography" involves the application of adjoint methods and 3D simulations of seismic wave propagation to seismic tomography (*Tromp et al.*, 2005; *Tape et al.*, 2007). The approach is that of a minimization problem: (1) specification of an initial model described in terms of a set of earthquake source parameters and 3D variations in density, shear-wave and bulk-sound speeds; (2) specification of a misfit function; (3) computation of the value of the misfit function for the initial model; (4) computation of the gradient (and Hessian,

if feasible) of the misfit function for the initial model; and (5) iterative minimization of the misfit function. This paper is organized following these steps, with emphasis on the new crustal model (Section 6.5).

Here we demonstrate the feasibility of our tomographic approach by iteratively improving a 3D crustal model of southern California. Within the iterative procedure we use traveltime measurements of body and surface waves from 52,000 three-component seismograms of 143 crustal earthquakes. After 16 iterations, the resulting crustal model generates seismic waveforms with substantially improved fits to observed waveforms and it captures features in the data that are not produced by the initial 3D model. The quality of the new crustal model allows us to simulate the details of earthquake ground motion at periods of two seconds and longer for hundreds of different paths in southern California. Our new crustal model contains strong vertical and lateral heterogeneity. Many new tomographic features are revealed, the most dramatic of which are the Coast Ranges and their numerous sedimentary basins, the southern San Joaquin basin, the mid-crust of the Mojave Desert region, and the mid-crust of the western Transverse Ranges.

6.2 Initial model

We compute synthetic seismograms using the spectral-element method (*Komatitsch et al.*, 2004). Due to the accuracy of the SEM, the goodness of fit between observed and synthetic seismograms depends only on the quality of the Earth structure model and the quality of the earthquake source model. Here we describe the structure and source parameters, followed by a description of the model vector $\mathbf{m}$ used in the tomographic inversion.

6.2.1 Initial 3D crustal model

We wish to use an initial seismological Earth model that produces the maximum number of measurements for a given set of earthquakes. Hence we begin with a 3D model (*Komatitsch et al.*, 2004) rather than a standard 1D layered model for southern California. The initial seismological model is provided by the Southern California Earthquake Center and contains results from several different seismic datasets: seismic reflection and industry well-log data to constrain the geometry and structure of major basins (*Süss and Shaw*, 2003; *Komatitsch et al.*, 2004; *Lovely et al.*, 2006), receiver function data to estimate the

depth to the Moho (*Zhu and Kanamori*, 2000), and local earthquake data to obtain the 3D background wavespeed structure (*Hauksson*, 2000; *Lin et al.*, 2007b). The seismological model is described in terms of shear wavespeed (V_S) and bulk sound speed (V_B), which can be combined to compute compressional wavespeed

$$V_P^2 = \tfrac{4}{3} V_S^2 + V_B^2 \, . \tag{6.1}$$

We extend the simulation region of *Komatitsch et al.* (2004) westward, so as to include the Coast Ranges (Figure 6.1). We also implement a more recent version of the background model (*Lin et al.*, 2007b), and we obtain density (ρ) by empirically scaling V_P (*Brocher*, 2005).

6.2.2 Earthquake sources

Each earthquake source is described by ten parameters: origin time (one), hypocenter (three), and moment tensor (six). Most of the epicenters and origin times were previously determined based upon the relocation technique of *Lin et al.* (2007a), and these remain unchanged during the iterative improvement of the seismological model. We perform numerous tests to determine the best focal mechanism for each earthquake, and we invert for the focal mechanisms (*Liu et al.*, 2004) once in the initial 3D model ($\mathbf{m}_{00}$) and again at the twelfth iteration ($\mathbf{m}_{12}$). The earthquake and station coverage for our study is shown in Figure 6.1.

Four criteria, in order of importance, influenced our selection of earthquakes for the tomographic inversion:

1. availability of quality seismic waveforms for the period range of interest (2–30 s) (must have at least 10 good stations);
2. availability of a relocated hypocenter (with origin time);
3. occurrence in a region with few other earthquakes;
4. availability of a "reasonable" initial focal mechanism.

These criteria led us to select earthquakes with $M_w \geq 3.4$, with the smallest ones in regions of sparse coverage. We represent the earthquakes as point sources in our simulations. Within

our period range of interest (2–30 s), this leads us to exclude earthquakes with $M_w > 5.5$. Because the computational cost of our technique is independent of the number of stations, we prefer to use larger earthquakes ($M_w \geq 4$). Larger earthquakes produce higher signal-to-noise ratios at more stations, thereby leading to more measurement windows.

We use the variable σ_0 to indicate the "water-level" minimum uncertainty associated with a traveltime measurement. For our tomographic inversion, we choose

$$\sigma_0 = 1.0 \text{ s} \tag{6.2}$$

based on the estimated uncertainties of earthquake source parameters. The tomographic inversion concentrates on reducing time shifts between synthetic and recorded waveforms that exceeded σ_0.

Epicenters

Earthquake epicenters are primarily from *Lin et al.* (2007a), where available. We supplement these with results from other local studies that used local and temporary stations. These local studies include *Lohman and McGuire* (2007) for a swarm of earthquakes in the Salton trough, *Thurber et al.* (2006) for earthquakes in the Parkfield region, and *McLaren et al.* (2008) for earthquakes in the San Simeon region. These studies all employed the double-difference method of *Waldhauser and Ellsworth* (2000), and generally provide the hypocenters and origin times of earthquakes, as well as a 3D tomographic model.

Epicenters—the latitudes and longitudes of the earthquakes—are the best determined source parameters. The seismicity studies listed above used a much larger dataset of traveltime picks to locate the earthquakes than we do in the tomographic inversion. The majority of the epicenters are from *Lin et al.* (2007a,b), who used P and S arrival times from 783 stations, including strong motion stations and temporary arrays. The dense coverage of stations in the vicinity of the earthquakes (e.g., the Parkfield array used in *Thurber et al.*, 2006) is important for epicenter estimation, as well as for depth and origin time. *Lin et al.* (2007a,b) also used information from controlled sources (quarry blasts and shots) to estimate uncertainties of absolute locations and absolute origin times (*Lin et al.*, 2006). The changes in wavespeed produced by our tomographic inversion are large (±30% locally) but

they impart only minor traveltime shifts[1] compared to $\sigma_0 = 1.0$ s. Therefore, we do not change the epicenters during the iterative tomographic inversion.

Depths and origin times

Our initial depths and origin times are from the relocated catalogs in Section 6.2.2, where available. If relocated source parameters are not available, then we use the source parameters listed in the Southern California Earthquake Data Center catalog. At model iteration $\mathbf{m}_{12}$, we performed source inversions that allowed the depths to change (Section 6.2.2), and we also applied an empirical correction to the origin times (discussed next). These adjustments to the depths and origin times induced time shifts much smaller than $\sigma_0 = 1.0$ s.

We noticed a minor, magnitude-dependent time-shift pattern, based on analysis of near-source seismograms for model $\mathbf{m}_{12}$ for the period range 2–30 s. The pattern indicated that larger events had systematically positive time shifts, even for stations in the vicinity of the earthquake source. We therefore modified the origin times of the sources with an empirical relationship given by $t_s = t'_s + 0.5h$, where t'_s is the listed origin time, h is the half-duration of the source (determined directly from M_{w}), and t_s is the new origin time. (t_s and t'_s are in "absolute" seconds, that is, with a particular reference zero-time; h is in seconds.) The maximum correction factor $0.5h$ for all 234 earthquakes was 0.65 s, the minimum was 0.06 s, the mean was 0.14 s, and the median was 0.11 s. Thus, all correction factors were less than $\sigma_0 = 1.0$ s. This minor adjustment in origin time is due to the fact that the origin time is derived from P wave picks on the unfiltered seismograms (thus, highest frequency) (*Allen*, 1978), while the origin time in our simulation is taken to be the *center* of a Gaussian source-time function with half duration h.

Focal mechanisms

Our initial focal mechanisms are selected from published catalogs of *Tan* (2006), *Hardebeck and Shearer* (2003), and the Southern California Earthquake Data Center (*Clinton et al.*, 2006). *Tan* (2006) implemented the "cut-and-paste" method (*Zhao and Helmberger*, 1994; *Zhu and Helmberger*, 1996), which uses both surface waves and body waves. The method of

[1] We illustrate this with an extreme example, considering an earthquake at $z = 10$ km depth, overlain by a region of $v_1 = 3$ km/s in the initial model, and we then apply a (very large) $n = \ln(v_1/v_2) = -0.3$ change in wavespeed. For a station immediately above the earthquake, the change in traveltime due to the change in structure will be $\Delta t = (z/v_1)[\exp(-n) - 1] \approx 1.2$ s. Thus, this extreme scenario produces a time shift slightly larger than $\sigma_0 = 1.0$ s.

Hardebeck and Shearer (2003) uses amplitude ratios between P and S waves. *Clinton et al.* (2006) implemented the method of *Pasyanos et al.* (1996) and *Dreger et al.* (1998), which uses relatively long-period surface waves (10–50 s). In cases where there were significant discrepancies among the mechanisms reported by different studies, we performed 3D simulations and compared the synthetics directly with the data to determine the best starting mechanism for the SEM-based inversion, discussed next.

Earthquake source inversions using 3D models

We performed inversions for earthquake focal mechanisms (*Liu et al.*, 2004) twice using the 3D crustal models: once at $\mathbf{m}_{00}$ and again at $\mathbf{m}_{12}$. For the inversion at $\mathbf{m}_{00}$, we inverted for the focal mechanism only, while keeping the hypocenter and origin time fixed. For some earthquakes, such as those near the Salton trough, the Ventura basin, and offshore Continental Borderlands, the quality of the inverted mechanism was affected by the poor quality of the initial source.

For the inversion at $\mathbf{m}_{12}$, we performed an SEM inversion first in the routine manner (*Liu et al.*, 2004), which starts with an initial focal mechanism. We also performed a (new) global grid search inversion that does not use the initial focal mechanism at all. Equipped with a much improved 3D crustal model ($\mathbf{m}_{12}$), we allowed both the focal mechanism and depth to change, while keeping the epicenter and origin time fixed. The mean and median depth changes for all earthquakes were both less than 0.5 km. We did not adjust the origin times based on changes to the depths, primarily because the changes in depth induced a traveltime shift[2] that was generally much less than $\sigma_0 = 1.0$ s.

6.2.3 Model variables, model parameterization, and model vector

We are therefore faced with constructing a model vector $\mathbf{m}$ for the tomographic inversion. The elements of $\mathbf{m}$ must describe the two continuous scalar fields $V_{\mathrm{S}}(\mathbf{x})$ and $V_{\mathrm{B}}(\mathbf{x})$, where $\mathbf{x}$ is a point in the volume. The continuous fields are represented using basis functions, which we choose to be the same ones used in solving the forward problem numerically (*Tape et al.*, 2007). Thus, each (local) gridpoint in the numerical mesh, $\mathbf{x}_i$, has a corresponding value

[2]To illustrate this point, consider the extreme example of an earthquake at depth z with a station directly above it. The change in traveltime due to a change in earthquake depth is $\Delta t = \Delta z / v$. In order to match $\sigma_0 = 1.0$ s, at which point we would consider modifying the origin time, we would need a 1 km depth change (larger than what we generally applied) beneath a layer with a wavespeed of 1 km/s, which is much slower than the (2–30 s) wavespeed structure in most earthquake source regions.

of V_S and V_B that appear as elements of the model vector $\mathbf{m}$. The model vector has $2G$ elements, with G the number of (local) gridpoints in the mesh, and 2 the number of inversion variables (V_S and V_B). We use a subscript to denote the model iterations, such that $\mathbf{m}_{00}$ is out initial model, and $\mathbf{m}_{16}$ is our final model.

6.3 Misfit function

6.3.1 Selection of bandpasses

The quality of fit between observed and synthetic seismograms is strongly dependent on the frequency content of the seismic waves, because the overall quality of the model generally diminishes with shortening scalelength. We therefore examine multiple period ranges: 6–30 s, 3–30 s, and 2–30 s. Our choice emphasizes fitting seismic waveforms in the period range 6–30 s, which, for crustal earthquakes in southern California, is dominated by surface waves. Table 6.4 summarizes measurements for the final tomographic model.

6.3.2 Selection of time windows

Because our synthetic seismograms are computed using a complex 3D seismological model, we require a measurement tool that can capture the complex effects of wave propagation. We use an automated algorithm (*Maggi et al.*, 2009) to select time windows for the entire seismic dataset. A given time window, or "measurement window," is selected if there is a user-specified, quantifiable level of agreement between the observed and simulated seismograms.

We use an automated algorithm, FLEXWIN, for picking seismogram time windows that contain a quantifiable level of agreement between synthetics and data (*Maggi et al.*, 2009). The algorithm requires several user parameters that need to be adjusted for a given dataset (Table 6.2). The only differences between our values and those listed in *Maggi et al.* (2009, Table 3) are for the quantities that specify the maximum allowable time shifts between data and synthetics: $\Delta\tau_0$ and $\Delta\tau_{\text{ref}}$. The values in *Maggi et al.* (2009) are based on the $\mathbf{m}_{00}$ synthetics, which contain time shifts in excess of 10 s, comparable to what is shown in Figure 6.3b. However, using $\mathbf{m}_{16}$, there are no identifiable time shifts larger than 5 s, and the standard deviation of the time shifts is less than 1 s. Thus, with $\mathbf{m}_{16}$ we specify FLEXWIN parameters to only allow windows to be considered in which the absolute value

of the time shift is less than 4 s (6–30 s records), 3 s (3–30 s records), and 2 s (2–30 s records) (Table 6.2).

As our crustal model improved, we began to notice agreement between data and synthetics for time intervals *after* the expected surface-wave arrival times. We therefore modified the user functions stated in *Maggi et al.* (2009, Section A1.3), by *not* raising the water-level at the end-time of the expected surface wave.

Although the window-picking algorithm is automated, it is important to examine every single time window for the period ranges 6–30 s and 3–30 s. By carefully examining all the window picks at each iteration, we were able to lower the window acceptance criteria, thereby allowing more windows to be selected. This led to the automated selection of additional windows that needed to be manually excluded. If computation is unlimited, then one could instead raise the acceptance criteria to the point where very little hand-checking of the windows is needed, although fewer windows are then used in the in the tomographic inversion.

6.3.3 Misfit measures

We consider two measures of misfit: a traveltime difference (F_{t}) and a waveform difference (F_{w}). We use the traveltime misfit measure within the tomographic inversion, such as the generic equations in Equations (6.5) and (6.6). We use the waveform misfit measure to assess the misfit reduction, because in many cases there is a waveform in the $\mathbf{m}_{16}$ synthetics to align with the data, but there is no corresponding waveform in the $\mathbf{m}_{00}$ synthetics (e.g., Figure 6.3b, 6.4b).

For a single time window on a single seismogram, the traveltime and waveform misfit measures are given by

$$F_{\mathrm{t}}(\mathbf{m}) = \int_{-\infty}^{\infty} \frac{h(\omega)}{H} \left[\frac{\Delta T(\omega, \mathbf{m})}{\sigma_{\mathrm{t}}}\right]^2 \mathrm{d}\omega \,, \tag{6.3}$$

$$F_{\mathrm{w}}(\mathbf{m}) = \frac{\displaystyle\int_{-\infty}^{\infty} w(t) \left[d(t) - s(t, \mathbf{m})\right]^2 \mathrm{d}t}{\left(\displaystyle\int_{-\infty}^{\infty} w(t) \left[d(t)\right]^2 \mathrm{d}t \int_{-\infty}^{\infty} w(t) \left[s(t)\right]^2 \mathrm{d}t\right)^{1/2}} \,, \tag{6.4}$$

where $d(t)$ denotes the recorded time series, $s(t, \mathbf{m})$ the simulated time series, $\sigma_{\mathrm{t}} \geq \sigma_0$ the estimated uncertainty associated with the traveltime measurement, $w(t)$ a time-domain

window, and $h(\omega)$ a frequency-domain window with associated normalization constant $H = \int_{-\infty}^{\infty} h(\omega)\mathrm{d}\omega$. The frequency-dependent traveltime measurement, $\Delta T(\omega, \mathbf{m})$, is made using a multitaper method (e.g., *Laske and Masters*, 1996; *Zhou et al.*, 2004). In the case of a frequency independent measurement, $\Delta T(\omega, \mathbf{m})$ reduces to a cross-correlation traveltime measurement. The expression for F_{w} (Eq. 6.4) contains the same normalization as the standard cross-correlation formula and has been used for source inversions (*Zhu and Helmberger*, 1996, Eq. 3).

6.3.4 Misfit function

In the tomographic inversion, within each measurement window we choose to measure the frequency-dependent traveltime difference between observed and simulated seismic arrivals. Measurements are made by cross-correlation or by a frequency-dependent multitaper technique. Our measurement misfit function for a single earthquake is defined by

$$F_e(\mathbf{m}) = \frac{1}{N_e} \sum_{i=1}^{N_e} \left[\Delta T_i(\mathbf{m})\right]^2 / \sigma_i^2 \ , \tag{6.5}$$

where $\mathbf{m}$ is a model vector, N_e denotes the total number of measurement windows for earthquake e, $\Delta T_i(\mathbf{m}) = T_i^{\mathrm{obs}} - T_i^{\mathrm{syn}}(\mathbf{m})$ is the traveltime difference between observed and synthetic waveforms associated with the ith window, and $\sigma_i \geq \sigma_0$ is the associated standard deviation. Our overall misfit function F is simply

$$F(\mathbf{m}) = \frac{1}{E} \sum_{e=1}^{E} F_e(\mathbf{m}) \ , \tag{6.6}$$

where E is the number of earthquakes. Implicit in Equation (6.6) are the choices of the L2-norm and an associated diagonal data covariance matrix $\mathbf{C}_{\mathrm{D}}$ containing terms of E, N_e, and σ_i.

6.4 Misfit gradient and iterative inversion procedure

A distinguishing feature of adjoint tomography is that the gradient of F_e (Eq. 6.5) is computed from an interaction between two wavefields: the "regular" forward simulation emanating from the earthquake source, and the "adjoint" simulation emanating from stations

(*Tarantola*, 1984; *Tromp et al.*, 2005; *Tape et al.*, 2007). The method we use for iterating from the initial model ($\mathbf{m}_{00}$) to the final model ($\mathbf{m}_{16}$) is adapted from the approach illustrated in *Tape et al.* (2007). The *Tape et al.* (2007) study involved a 2D tomographic problem using only synthetic seismograms, whereas the current study is 3D and uses real data. In the 3D problem, our model vector $\mathbf{m}$ contains two variables that describe the structure, V_S and V_B. Finally, our misfit measure is a frequency-dependent multitaper traveltime difference, made using three different period ranges (6–30 s, 3–30 s, 2–30 s).

Our main computational cost involves computing the gradient of F_e (one for each earthquake), which we call an "event kernel" (*Tape et al.*, 2007). In *Tape et al.* (2007) we combined the event kernels by summing them, and then used a conjugate-gradient algorithm to obtain a model update. In this study, we use a subspace projection technique to compute the model update. Instead of using a subspace of model parameter classes (*Kennett et al.*, 1988; *Sambridge et al.*, 1991), we use a subspace of earthquake sources. The basic idea is to determine a linear combination of event kernels that exploits the features they have in common. This procedure provides a preconditioner for the gradient algorithm that increases convergence of the minimization problem. Not all earthquakes were used in each iteration (Table 6.5). This is because for certain iterations, the event kernels were much stronger in specific regions of the model, indicating that the majority of the observed misfit was originating in those regions. The model updates for these iterations did not change appreciably with the inclusion of all event kernels (although the computation time increased considerably).

6.4.1 Computational demands

The computational demands of adjoint tomography are formidable, due to the number of simulations needed to evaluate the misfit function and its gradient at every iteration. Our simulation region is 639 km × 503 km at the surface and extends to 60 km depth (Figure 6.1). For each earthquake we calculated four minutes of seismograms that are accurate down to a period of two seconds. Each simulation took approximately 43 minutes (of wall-clock time) on 168 cores of a parallel computer. For each earthquake we performed three simulations, one to evaluate the misfit function (Eq. 6.5) and two to compute its gradient. Each model iteration thus required $3E_k$ simulations, where E_k is the number of earthquakes used for the kth iteration. The total number of 168-core simulations used in producing model $\mathbf{m}_{16}$

was 6794, totaling 0.8 million CPU-hours. Tabulations of these simulations are shown in Table 6.5 and discussed next.

Table 6.5 lists the number of forward simulations used in constructing the final tomographic model. The total number includes: (1) both sets of earthquake source inversions (at models $\mathbf{m}_{00}$ and $\mathbf{m}_{12}$), (2) forward simulations that were used to construct kernels used in the inversion, (3) forward simulations for 91 earthquakes that were not used in the tomographic inversion, and (4) forward simulations that were used in constructing some kernels that were not used in computing the model update. For each computed kernel, we list "3" as the number of forward simulations performed; it would be possible, in theory, to reduce this number to "2" if we were to hold the final snapshot of the forward wavefield in memory on the parallel computer while computing the misfit function and adjoint sources, prior to creating the adjoint wavefield (*Liu and Tromp*, 2006).

The total number of simulations per earthquake is listed in the fourth column of Table 6.5: a total of 6794 168-core simulations. The number of CPU-hours per earthquake is then the product of the following parameters: number of cores per simulation, number of simulations, duration of simulation (seismogram hours), seismogram-to-wall-clock factor. For our simulations we use 168 cores, and for the desired accuracy of periods of 2 s, the simulation-to-wall-clock-factor is about 13. For example, 300 s of seismograms requires 65 minutes (13×300 s) of wall-clock time. The total for all 6794 simulations is then 0.80 million CPU-hours.

6.5 New 3D crustal model

We present our new crustal model on both relative and absolute scales. First, the update to the seismological model (the relative scale) reveals the changes to the initial SCEC model that are required by the data. We compute the update as $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$. Second, the seismological model itself reveals the "absolute" model parameters (e.g., wavespeed in units of km/s) and is more relevant for geologic and geodynamic interpretations. All cross sections discussed below are of shear wavespeed (V_S) models ($\mathbf{m}_{00}$ and $\mathbf{m}_{16}$). The bulk sound speed model is discussed in Section 6.5.2.

Figure 6.3 shows the iterative improvement of a single three-component seismogram for an earthquake source beneath the southern San Joaquin basin recorded in the eastern

Mojave. The evolution of the particle motion from $\mathbf{m}_{00}$ to $\mathbf{m}_{16}$ is illustrated by the red synthetic seismograms. The varying time windows are a reminder that the measurement windows may change at each iteration as the fit improves (*Maggi et al.*, 2009). The initial traveltime difference (or "time shift") for the Rayleigh wave in the standard 1D southern California model is 9.60 s on the vertical component (Z) and 9.85 s on the radial component (R). In the initial SCEC model ($\mathbf{m}_{00}$; Figure 6.3b), the Rayleigh wave time shift is 7.00 s (Z) and 7.30 s (R). After 16 iterations, the time shifts are 0.95 s (Z) and 1.00 s (Z). The evolution of the transverse component is more dramatic, as there is virtually no energy between 110 s and 140 s in the SCEC model, and thus there is no time shift to identify. After 16 iterations, the time shift is 0.05 s, and the synthetic transverse-component seismogram captures the main shape of the waveform up to 130 s.

Having demonstrated reasonable fits on all three components, we examine the net change between the initial and final models that resulted in the misfit reduction (Figure 6.3a). The tomographic update, $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$, contains three principal features. First, there is the addition of the southern San Joaquin basin (*Goodman and Malin*, 1992), marked as a -35% (slow) anomaly immediately above the earthquake source. The basin resonates, influencing the Love wave observed after 110 s. Second, the shear wavespeed of the western Mojave is increased in the upper 3 km, decreased in the depth range 5 km to 15 km, and increased in the depth range 18 km to 26 km. Third, east of the Camp Rock fault, there is a -10% change in wavespeed, probably associated with Quaternary volcanism (*Luffi et al.*, 2009). Only through multiple iterations is it possible to isolate the locations and amplitudes of these changes.

In Figure 6.4, we highlight the improvement in fits for two additional earthquakes. We first consider a Parkfield earthquake on the San Andreas fault recorded at station WGR, north of the Ventura basin (Figure 6.4b). The synthetics for the final model exhibit improved fit of amplitude and phase for both the Rayleigh wave arrival (Z and R: 75–100 s) and the Love wave arrival (T: 60–90 s). In comparison to the 1D and $\mathbf{m}_{00}$ synthetics, the $\mathbf{m}_{16}$ synthetics constitute a dramatic improvement.

We also observe improved fits within the region containing the higher-resolution basin models (*Komatitsch et al.*, 2004). Figure 6.4d shows the improvement in fits for an earthquake that was not used in the tomographic inversion. The seismic wavefield interacts with the Los Angeles and Ventura basins before reaching station STC. The SCEC model captures

the resonance, duration, and approximate amplitude of the observed seismogram, but the final 3D model is markedly better. In particular, we note that the fits for the amplitudes are improved, even though amplitude differences are not built into the misfit function. This demonstrates that the 3D structural changes to the initial model induce additional focusing and amplification of the seismic wavefield.

We emphasize five key points about the model updates (Figure 6.5):

1. The net changes in the model are large, locally in excess of $\pm 30\%$, but the changes during any one iteration are small, locally less than $\pm 10\%$.

2. The areas in the initial model that require changes are highly variable and generally unknown. For example, it takes more than ten iterations to isolate the -35% anomaly related to the southern San Joaquin basin.

3. Although only traveltime differences are used in the misfit function (Eq. 6.5), amplitude differences also decrease in the final model, due to 3D effects of focusing (Figures 6.4d and 6.8).

4. Frequency-dependent surface waves are necessary to resolve crustal structure (Figure 6.3).

5. We are able to use more seismograms for measurements as we improve the seismological model.

We now illustrate model $\mathbf{m}_{16}$ with four additional paths (Figure 6.7). Figure 6.7a shows a striking three-component seismogram of an earthquake that originated near the base of the Salton trough recorded at LA basin station LAF. This path is important for seismic hazard assessments, since large earthquakes have occurred in the past along the southern San Andreas (*Olsen et al.*, 2006). The synthetic seismograms capture the phase and amplitude of most of the first 180 s of the observed records.

We also capture wave propagation effects that occur far from the direct path between the earthquake and station. The three windows in Figure 6.7b highlight three different Rayleigh-wave paths from the earthquake (near Hollywood) to station RVR. The latter two waveforms are not apparent in the synthetics for the initial 3D model. Inclusion of such waveforms in the tomographic inversion shows we can increase the coverage by exploiting additional information already present in the seismograms.

Waveform fits for the shortest period range, 2–30 s, are shown in Figure 6.7c–d. Most of the body-wave pulses in Figure 6.7c fit the observed pulses to within 1 s (our target value). The downward pulse at 33 s on the transverse component of the $\mathbf{m}_{16}$ synthetic seismogram is not apparent in the corresponding $\mathbf{m}_{00}$ synthetic seismogram. Seismograms for a path crossing the entire Mojave (*Fuis et al.*, 2003) are shown in Figure 6.7d. These seismograms match the Love wave at 68 s (T), the P wave at 30 s (Z and R), the Rayleigh wave at 75 s (Z and R), and some additional complexity, particularly on the radial component.

6.5.1 Connections with geology and tectonics

Within a single vertical cross section for a path crossing the Mojave region (Figure 6.3a), we identify reductions in wavespeed due to both compositional and thermal features: above the earthquake source is the southern San Joaquin sedimentary basin (*Goodman and Malin*, 1992), and east of the Camp Rock fault (and north of 34.3°N) there is higher heat flow (*Bonner et al.*, 2003) likely related to Quaternary volcanism (e.g., *Luffi et al.*, 2009, Figure 1). West of the Camp Rock fault, slow wavespeeds and high heat flow are not observed at the surface, nor is volcanic activity.

The middle panels of the horizontal cross sections (Figure 6.5) reveal lateral variations in the new crustal model. At 2 km depth, large-scale slow regions (<2.8 km/s) reveal several known Neogene basins (Figure 6.1a). The fastest regions (>3.5 km/s) occur in the Peninsular Ranges west of the Elsinore fault, and in the Sierra Nevada west of the Kern Canyon fault (*Shapiro et al.*, 2005). The eastern front of the Sierra Nevada is marked by an eastward step in wavespeed from about 3.5 km/s to 2.8 km/s, with the Coso geothermal region and the sedimentary fill in Owen's Valley and Indian Wells Valley accounting for the slower wavespeeds. We attribute the slow wavespeeds (2.9 km/s) in the eastern Mojave to Quaternary volcanism.

At 10 km depth in our crustal model (Figure 6.5b), some of the basins are no longer visible (Los Angeles, Salton trough), and a striking pattern of wavespeeds west of the San Andreas fault is evident. The Peninsular Ranges and a magmatic layer beneath the Salton trough (*Fuis and Kohler*, 1984) form a fast region (3.8 km/s) that is separated by the San Andreas fault from slower regions (3.4 km/s) to the northeast. The 50-km-scale variations in wavespeeds along the longitudinal line 119°N illuminate, from north to south, the western Sierra Nevada (fast), the southern San Joaquin basin (slow), the San Gabriel Mountains

(fast), the Ventura basin (slow), the Santa Monica Mountains (fast), and the Santa Monica basin (slow). Wavespeeds in the Coast Ranges are slowest (3.1 km/s) east of the San Andreas (*Bleibinhaus et al.*, 2007) and along the coast, and are somewhat faster (3.4 km/s) in between, where Mesozoic granitic and sedimentary rocks are exposed at the surface. The northwestern Mojave is slow (3.3 km/s) compared to faster material (3.6 km/s) in the southern Sierra Nevada, across the Garlock fault.

At 20 km depth (Figure 6.5c) the most striking feature is the fast wavespeed region (4.1 km/s) beneath the Ventura–Santa Barbara basin and the Santa Monica Mountains, also observed in Figure 6.4c. This region coincides with the surface expression of the western Transverse Ranges block (WTRB *Luyendyk et al.*, 1980) (Figure 6.1a), bound to the north by the Santa Ynez fault and to the south by the Malibu Coast fault. We interpret this feature as subduction-captured Farallon oceanic crust, on top of which the WTRB rotated clockwise by more than 90° from a position near the Peninsular Ranges (3.8 km/s) (*Nicholson et al.*, 1994). It is possible that the WTRB crustal anomaly is related to upper mantle anomalies observed below this region (*Humphreys and Clayton*, 1990).

6.5.2 Bulk sound speed model

Horizontal cross sections of the bulk sound speed (V_B) models are shown in Figure 6.6. At 20 km and 10 km depths, the larger spatial extent of the masked region indicates that our sensitivity to V_B is not as good as at shallow depths.

At 2 km depth the initial 3D model, V_B $\mathbf{m}_{00}$, contains considerable spatial variations that are also present in the final model, because our perturbation contains only longer scalelength variations. Our perturbation, $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$, is almost uniformly negative, indicating that the bulk-sound speed—on the whole—is too fast in the initial model.

Outside the Los Angeles basin, the slowest feature in the initial model is near Indian Wells Valley, just south of the Coso geothermal region (e.g., *Hauksson and Unruh*, 2007). Our model update applies a -15% change in V_B to this anomaly. Interestingly, in the central Mojave region at 2 km depth, the change to V_B is about -10% (Figure 6.6a), whereas the change to V_S is about $+5\%$ (Figure 6.5a). These changes of opposite sign will lead to more pronounced changes in quantities such as Poisson's ratio and the $V_\mathrm{P}/V_\mathrm{S}$ ratio (e.g., *Christensen*, 1996).

6.5.3 Implications for seismic hazard assessment

Our results demonstrate that moderate (M_w = 3.5–5.5), well-recorded earthquakes (Figure 6.1) can be used to make large, necessary changes to the crustal model of southern California. Waveforms from these earthquakes can be extremely complicated, even at relatively long periods (6–30 s). One example is the path from the Salton trough to the Los Angeles basin (Figure 6.7a), the same region covered by the TeraShake simulations (*Olsen et al.*, 2006).

If it is not possible to fit waveforms from moderate, point-source earthquakes, then it will not be possible to fit waveforms from large earthquakes with complex ruptures. By beginning to fit complex propagation paths for moderate earthquakes, we provide hope for accurately simulating larger earthquakes. For example, the transverse-component ground motion from a M_w 4.5 earthquake on the White Wolf fault (Figure 6.3) is not apparent in the simulations based upon the initial 3D model, but is present in those based upon the final model. The new model therefore provides a better starting point for simulating larger earthquakes on the White Wolf fault, such as the 1952 M_w 7.2 Kern Country earthquake. Similarly, the improvement in fits for the M_w 4.6 Parkfield earthquake (Figure 6.3) suggests that the new crustal model is more appropriate for simulations of larger earthquakes on this segment of the San Andreas fault, such as the 1857 M_w 7.9 Fort Tejon earthquake.

In southern California, there are no high-quality seismic waveforms available for major "scenario" earthquakes, that is, earthquakes that have occurred in the past and that are likely to reoccur in the future (*Olsen et al.*, 2006). However, there are several strong (M_w = 6–7) earthquakes that have been recorded well enough to determine rupture models (*Custódio et al.*, 2005). These earthquakes present the formidable challenge of fitting near-source and regional waveforms that capture the complexities of both the rupture process and the heterogeneous structure. Our more accurate 3D crustal model will benefit the development of rupture models for strong earthquakes.

An improved crustal model will allow for the systematic search for "exotic" seismic waveforms that result from wave propagation in complex 3D structure that may contain interfaces at all possible orientations. This search should be undertaken in the regions with the strongest heterogeneity, including surface topography variations (*Ma et al.*, 2007; *Lee et al.*, 2008), and emphasis should be placed on waveforms with anomalous amplitude,

indicative of damaging energy. It is likely that quantities such as the maximum amplitude of ground motion at a particular location will likely be extremely sensitive to changes in source–station geometry. In the same manner in which we can illuminate the sensitivity region of a Pn head wave (*Liu and Tromp*, 2006), we should be able to also illuminate the region of more exotic waves.

6.6 Misfit analysis

In Figure 6.3 we illustrated the misfit reduction for one particular path between an earthquake source and a station. Our overall assessment of the misfit reduction from the initial SCEC model to the final model is based on 12,583 different paths (Table 6.4). This assessment cannot be based simply on a traveltime misfit function, because there are many seismic waveforms in the final model that do not have a measurable traveltime difference in the initial model (e.g., Figure 6.3b). Thus, in order to facilitate a direct comparison between the two models, we compute a simple waveform difference using the time windows that were selected for the final model for the 143 earthquakes used in the tomographic inversion (Figure 6.8a).

We also consider a separate set of 91 earthquakes that was not used in the tomographic inversion. An earthquake not used in the tomographic inversion—or any future earthquake, for that matter—may be used to independently assess the misfit reduction from $\mathbf{m}_{00}$ to $\mathbf{m}_{16}$. Remarkably, the reduction in waveform difference misfit for the extra earthquakes is almost the same as it is for the earthquakes used in the inversion (Figure 6.8). This result provides validation for the tomographic model and suggests that future earthquakes will see the same misfit reduction.

Many of the 91 earthquakes occur in similar regions to the 143 earthquakes (Figure 6.2), and thus one might argue that the misfit reduction would be similar, since the paths are similar to those used in the tomographic inversion. We would agree to some extent, but would counter with two points. First, even for near-identical paths, the occurrence of similar misfit confirms the quality of the initial uninverted source parameters (focal mechanism, origin time, hypocenter). This is a critical aspect of the tomographic inversion, that we are (for the most part) not mapping seismogram misfit due to source errors into structure changes; these consequences are illustrated in (*Tape et al.*, 2007, Figure 19b). Second,

"similar" paths—in the sense of a "nearby" earthquake recorded at the same station—may not be very similar in regions of strong heterogeneity. A change of source location (or station location, for that matter) of 2 km can have a profound impact on seismic waveforms, especially at higher frequencies.

6.6.1 Waveform misfit, $F_{\mathrm{w}}(\mathbf{m})$

We use a direct waveform difference, $F_{\mathrm{w}}(\mathbf{m})$ (Eq. 6.4), as the primary measure of misfit. In tabulating the histograms in Figure 6.8, we exclude all windows whose time shifts in *both* $\mathbf{m}_{00}$ and $\mathbf{m}_{16}$ are ≤ 1 s, our target measurement value in the tomographic problem. This leaves behind only those windows that have changed appreciably (for better or for worse). The waveform misfit measure is applied to either the portions of records within the time windows (Figure 6.8a, c) or to the entire seismogram containing at least one (non-excluded) time window (Figure 6.8b, d), including time before the expected P-wave arrival and after the surface wave arrivals. For a given set of windows, the number of seismograms containing windows will be less than (or equal to) the number of windows. The number of seismograms listed includes (up to three) different bandpassed versions.

The waveform misfit analysis is shown in Figure 6.8. There are several comparisons to make among the subplots.

1. There is a strong similarity between the earthquakes used in the inversion ("tomo") and the earthquakes not used in the inversion ("extra"). (In some cases, the "extra" earthquakes actually display a better misfit reduction than the "tomo" earthquakes.)

2. The waveform misfit of the full seismograms (Figure 6.8b, d) is reduced.

3. The waveform misfit of the measurement windows is better than that computed for the full seismograms, as expected.

4. For the measurement windows, neither the overall misfit nor the misfit reduction show a dependence on period range. This is not a one-to-one comparison, since the comparison is for different sets of windows, but it suggests that for many windows, such as those common to all three period ranges, the synthetic waveforms are capturing the dominant features of the wavefield. For the full seismograms, however, both the misfit and misfit reduction get progressively better from 2–30 s to 3–30 s to 6–30 s.

This is because measurement windows selected on the 6–30 s records cover more of the full seismogram than those selected on the 3–30 s and 2–30 s records, and thus we expect more of a misfit reduction.

6.6.2 Traveltime misfit, $F_t(\mathbf{m})$

We use a multitaper traveltime difference, $F_t(\mathbf{m})$ (Eq. 6.3), within the tomographic inversion. The traveltime differences in the final model have a standard deviation of less than 1 s for the entire dataset (Figure 6.9). In other words, given an adequate location, origin time, and focal mechanism for any $M_w = 3.5$–5.5 earthquake in southern California, we expect most traveltime differences computed using our crustal model to be ≤ 1 s for seismic records in the period range 2–30 s. For the three period ranges, the patterns do not change appreciably, and all listed standard deviations of the time shifts are <0.8 s (Figure 6.9).

6.7 Resolution considerations

We compute the composite volumetric sensitivity of all measurements (e.g., *Chen et al.*, 2007). This is achieved by using the same procedure that was used to compute each event kernel (*Tape et al.*, 2007), and omitting the traveltime measurement weight for each adjoint source (*Tromp et al.*, 2005). Because we have two inversion variables, V_S and V_B, we also have two corresponding volumetric sensitivity kernels, which we refer to as "coverage kernels" for brevity.

Coverage kernels for V_S and V_B are shown in Figures 6.10 and 6.11. The left column shows the field without the mask, and the right column includes the mask. The threshold for the mask is given by a subjective value of $K_0 = 10^{-16}$ m^{-3}. In regions where the coverage kernel is less than this value, the tomographic model is masked out, as shown in Figure 6.10. The coverage kernel decreases with depth, and has maximum sensitivity near the surface, which is sampled by the shorter-period surface waves in the dataset. The masks shown in Figure 6.10 are applied to the V_S cross sections shown in Figure 6.5. The coverage kernel for V_B contains lower amplitudes, leading to larger masks in Figure 6.11. These masks are applied to the V_B cross sections shown in Figure 6.6.

A seismic tomographic study will typically include a resolution analysis that shows how well a model perturbation (e.g., a delta function or a checkerboard pattern) is expected

to be resolved by the inversion procedure. These representations, however, are limited by the forward model embedded within the inverse problem. If the forward model is a simple computation (such as with ray theory), then it is possible to perform resolution tests with limited additional computation.

In our case, the forward model is computationally expensive, and a resolution test would require a comparable number of simulations as the real problem. Instead of a formal resolution analysis, we qualitatively examine the model update $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$, which provides estimates of the resolvable scalelengths in our problem. In particular, the minimum scalelength is about 2 km in depth (visible in Figure 6.3a) and about 6 km laterally. These values correspond to the regions of densest coverage, and lower resolution is expected in regions of poor coverage, such as the Great Valley or near the boundaries of the simulation region. We note that shorter scalelength features are present in the unsmoothed event kernels, but our choices of regularization prevent them from appearing in the model updates. In other words, we adopt a conservative approach that will introduce the finer details into the model only if seismograms from many different earthquakes require them.

We advocate monitoring the uncertainty of model parameters rather than conducting formal resolution analyses, because the former can be achieved without repeating the full inverse problem. Future work will address the uncertainties of model parameters. From a Bayesian perspective, our final model ($\mathbf{m}_{16}$) represents the mean model of a posterior distribution of possible models ($\mathbf{C}_{\mathrm{M}}$) (*Tarantola*, 2005). It is important that we construct a distribution of possible models that provides a guide for future studies that inevitably will include more stations and higher-frequency seismograms. The inclusion of additional data will then help reduce the distribution of all possible models. One promising approach to considering uncertainties of model parameters is the square-root variable metric method. Using this method it is possible to obtain *samples* of the posterior distribution $\mathbf{C}_{\mathrm{M}}$ without computing the prohibitively large $M \times M$ $\mathbf{C}_{\mathrm{M}}$ itself.

6.8 Summary

In this study we have exploited the accuracy of the SEM (*Komatitsch and Vilotte*, 1998; *Komatitsch et al.*, 2004) within an inverse problem. After 16 iterations, we have obtained a model with local perturbations of $\pm 30\%$ from the initial 3D model. We have fit thousands

of three-component waveforms that capture both the phase and amplitude of the particle motion at the stations. Direct waveform difference measurements of full-length seismograms that were never used in the inversion (Figure 6.8b, d) provide compelling support for the quality of the new crustal model. The ability to capture "exotic" seismic waveforms, such as from basin resonance (Figure 6.4b) and from laterally reflected surface waves (Figure 6.7b; also *Stich et al.* (2009)), suggests the prospect of enhancing tomographic images by delving deeper into the seismic records.

Seismic studies indicate the possible presence of interfaces at all scales and all possible orientations (*Ni et al.*, 2002; *Fuis et al.*, 2003; *Bleibinhaus et al.*, 2007). Interfaces in our tomography model that were sharp to begin with—the topography surface, the basement surface, and the Moho—remain sharp in the final model, since we only allow relatively smooth changes to the model. However, it is clear that the model update illuminates several features that appear to be sharp, including the southern San Joaquin basin and the eastern margin of the Sierra Nevada. And it is also clear from previous studies that interfaces such as the Moho are not uniformly sharp (*Mori and Helmberger*, 1996). The next stage of improving the model will be to numerically implement the most robust interfaces, and to check for the improvement of fits to seismograms, in particular for reflected phases from the interfaces.

Our new tomographic model for southern California is described in terms of shear-wave and bulk-sound speeds. The topography of primary interfaces (Moho and basement surface) remains fixed, anisotropy is not permitted, and attenuation does not change. These are all parameters that are specified in the forward model but are not used in the inverse problem, and they constitute future possible extensions of this research. Sensitivity kernels for anisotropic parameters (*Sieminski et al.*, 2007) and boundary surfaces (*Tromp et al.*, 2005; *Liu and Tromp*, 2008) are already available for the inverse problem, but there will be challenges in implementing multiple classes of model parameters.

Table 6.1: Standard 1D reference model for southern California (*Kanamori and Hadley*, 1975; *Dreger and Helmberger*, 1990; *Wald et al.*, 1995). V_B is listed for comparison and is computed via $V_\mathrm{B}^2 = V_\mathrm{P}^2 - \frac{4}{3}V_\mathrm{S}^2$. The Moho depth of 32 km is from *Wald et al.* (1995). The listed 60 km base level is based on our numerical simulation region.

layer	top m	bottom m	thickness m	V_P m s^{-1}	V_B m s^{-1}	V_S m s^{-1}	ρ kg m^{-3}
1	0	5500	5500	5500	4095	3180	2400
2	5500	16000	10500	6300	4693	3640	2670
3	16000	32000	16000	6700	4992	3870	2800
4	32000	60000	28000	7800	5817	4500	3000

Table 6.2: Values of parameters used for the window-picking code FLEXWIN (*Maggi et al.*, 2009), based on the synthetics for the final model $\mathbf{m}_{16}$.

$T_{0,1}$	6, 30	3, 30	2, 30
$r_{P,A}$	3.0, 2.5	2.5, 3.5	2.5, 3.5
r_0	3.0	4.0	4.0
w_E	0.18	0.11	0.07
CC_0	0.71	0.80	0.85
$\Delta\tau_0$	4.0	3.0	2.0
$\Delta \ln A_0$	1.5	1.0	1.0
$\Delta\tau_\mathrm{ref}$	0.0	0.0	0.0
$\Delta \ln A_\mathrm{ref}$	0.0	0.0	0.0
c_0	0.7	1.3	1.0
c_1	2.0	4.0	5.0
c_2	0.0	0.0	0.0
$c_{3a,b}$	3.0, 2.0	4.0, 2.5	4.0, 2.5
$c_{4a,b}$	2.5, 12.0	2.0, 6.0	2.0, 6.0
$w_\mathrm{CC}, w_\mathrm{len}, w_\mathrm{nwin}$	0.5,1.0,0.7	0.70,0.25,0.05	1,1,1

Table 6.3: Summary of the tomographic inversion, based only on seismograms with measurements for the final model $\mathbf{m}_{16}$. "TOMO" corresponds to the 143 earthquakes used in the tomographic inversion (Table 1). "EXTRA" corresponds to the 91 extra earthquakes not used in the tomographic inversion. "COMBINED" corresponds to the TOMO+EXTRA set of 243 earthquakes. The number of unique seismograms is indicated next to "seismograms (unique)". The number of total seismograms—including the three 6–30 s, 3–30 s, and 2–30 s—is indicated next to "seismograms (total)". The same is true for "windows (total)". A "path" is a single source-station pair that has at least one measurement.

	TOMO	EXTRA	COMBINED
components (Z,R,T)	3	3	3
networks	8	7	8
earthquakes	**143**	**91**	**234**
stations	203	200	210
paths	12583	4305	16888
seismograms (unique)	27007	8013	35020
seismograms (total)	52138	14803	66941
windows (total)	61673	16758	78431

Table 6.4: Summary of seismogram measurement time windows for final model $\mathbf{m}_{16}$. Each entry corresponds to the number of measurement windows for a particular period range of data (6–30 s, 3–30 s, 2–30 s) for a particular component (Z, R, T).

	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

Table 6.5: Tabulation of simulations for each earthquake and each model iteration. The description of each column is listed below, with the values in the subsequent table, with the totals for each column listed in the final row. The label "relocation available" indicates whether a relocated hypocenter (with origin time) was available for the earthquake (e.g., *Lin et al.*, 2007a; *Thurber et al.*, 2006).

column	label	description
1		index
2	`eid`	earthquake ID (Southern California Earthquake Data Center)
3	`dur`	duration of computed seismograms in seconds (120, 200, 300)
4	`Ne`	total number of forward simulations for earthquake
5	`m00`	number of forward simulations performed at $\mathbf{m}_{00}$: `1` = synthetics only; `3` = kernel (not used); `*3` = kernel (used)
6	`S00`	number of forward simulations performed for source inversion at $\mathbf{m}_{00}$: `6` = SEM inversion; `0` = no SEM inversion
7	`m01`	number of forward simulations performed at $\mathbf{m}_{01}$
⋮	⋮	⋮
18	`m12`	number of forward simulations performed at $\mathbf{m}_{12}$
19	`S12`	number of forward simulations performed for source inversion at $\mathbf{m}_{12}$: `6` = SEM inversion; `0` = no SEM inversion
20	`m13`	number of forward simulations performed at $\mathbf{m}_{13}$
⋮	⋮	⋮
23	`m16`	number of forward simulations performed at $\mathbf{m}_{16}$
24		relocation available (`Y` or `N`)
25		earthquake used in at least one model update (`TOMO`) or kept as an extra earthquake for misfit analysis(`EXTRA`)

	eid	dur	Ne	m00	S00	m01	m02	m03	m04	m05	m06	m07	m08	m09	m10	m11	m12	S12	m13	m14	m15	m16		
1	9967025	200	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
2	9967137	200	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
3	9967249	200	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
4	9967901	300	49	*3	6	*3	*3	*3	*3	*3	*3	1	0	0	0	0	3	6	*3	*3	*3	3	Y	TOMO
5	9968977	300	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	*3	*3	3	Y	TOMO
6	14096736	300	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	*3	*3	3	Y	TOMO
7	14189556	200	25	1	6	0	0	0	0	1	0	1	0	0	0	0	1	6	0	*3	*3	3	Y	TOMO
8	14263252	120	4	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	N	EXTRA
9	14095540	300	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	*3	*3	3	Y	TOMO
10	14096196	300	49	1	6	0	0	0	0	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	*3	*3	3	Y	TOMO
11	10063349	300	39	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	1	6	0	*3	*3	3	Y	TOMO
12	10100053	300	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	0	*3	6	*3	*3	*3	3	Y	TOMO
13	9171679	120	39	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	1	3	Y	TOMO
14	9983429	300	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	3	*3	*3	3	Y	TOMO
15	14138080	300	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	3	*3	*3	3	Y	TOMO
16	10097009	200	36	*3	6	*3	1	0	0	1	0	1	0	0	0	0	*3	6	3	*3	*3	3	Y	TOMO
17	14186612	200	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	3	*3	*3	3	Y	TOMO
18	14186928	120	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
19	9094270	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
20	9151000	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
21	9875657	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
22	9875665	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO

23	9882325	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	*3	*3	3	Y	TOMO
24	9882329	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	*3	*3	3	Y	TOMO
25	14095628	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	*3	*3	3	Y	TOMO
26	14187364	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
27	9095528	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
28	9151609	200	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	*3	3	Y	TOMO
29	9644345	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	*3	*3	3	Y	TOMO
30	9653293	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
31	9653349	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
32	9653493	200	63	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	3	*3	*3	3	Y	TOMO
33	12887732	120	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	*3	*3	3	Y	TOMO
34	9915909	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
35	13986104	200	38	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
36	9994573	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
37	14169456	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	*3	*3	3	Y	TOMO
38	9044494	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
39	3298170	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
40	9044650	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
41	9045109	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
42	9045697	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
43	9116921	200	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	*3	3	Y	TOMO
44	7179710	200	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	*3	3	Y	TOMO
45	9141142	120	30	*3	6	*3	*3	1	1	1	0	1	0	0	0	0	1	6	0	1	0	3	N	TOMO
46	7180136	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
47	9163702	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
48	9642941	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
49	9646589	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
50	10964587	200	48	*3	6	*3	*3	*3	*3	*3	*3	*3	3	0	0	0	3	6	0	3	0	3	Y	TOMO
51	9673577	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
52	9674049	300	40	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
53	9674097	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
54	9674205	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
55	9674213	300	40	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
56	9674653	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
57	10992159	300	48	*3	6	*3	*3	*3	*3	*3	*3	*3	3	0	0	0	3	6	0	3	0	3	Y	TOMO
58	11671240	200	38	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
59	9686565	200	48	*3	6	*3	*3	*3	*3	*3	*3	*3	3	0	0	0	3	6	0	3	0	3	Y	TOMO
60	9688025	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
61	9688709	200	48	*3	6	*3	*3	*3	*3	*3	*3	*3	3	0	0	0	3	6	0	3	0	3	Y	TOMO
62	9828889	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	*3	*3	3	Y	TOMO
63	9829213	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	*3	3	Y	TOMO
64	10023841	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
65	9152038	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
66	9165019	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
67	9171064	120	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
68	9631385	120	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
69	14007388	120	38	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	0	3	Y	TOMO
70	14204000	120	34	*3	6	*3	*3	*3	1	1	0	1	0	0	0	0	1	6	0	*3	0	3	Y	TOMO
71	14219360	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	3	6	0	*3	*3	3	Y	TOMO
72	14418600	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	N	EXTRA
73	3320736	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
74	9109131	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
75	9109254	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
76	9109287	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
77	9109442	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
78	9109496	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
79	9109636	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
80	9110685	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
81	9111353	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
82	9112735	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
83	9113909	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO

84	3321590	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
85	3320884	120	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
86	9114042	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
87	9114612	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
88	3324595	120	4	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
89	9114763	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
90	9114775	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
91	9114812	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
92	9114858	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
93	3320951	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
94	3320940	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
95	3320954	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
96	9117942	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
97	3321426	120	37	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	1	1	0	3	Y	TOMO
98	9119414	120	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	N	EXTRA
99	9120741	120	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
100	9122706	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
101	9130422	120	39	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	1	0	3	Y	TOMO
102	7177729	120	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
103	9147453	120	34	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	1	6	0	1	0	3	Y	TOMO
104	9155518	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
105	9775765	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
106	9805021	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
107	9854597	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
108	13945908	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
109	9930549	200	46	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	*3	*3	3	Y	TOMO
110	14408052	300	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	N	EXTRA
111	12659440	200	38	*3	6	1	0	0	0	*3	*3	*3	*3	3	0	0	1	6	0	3	0	3	Y	TOMO
112	10006857	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	3	3	0	3	Y	TOMO
113	14139108	120	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
114	14139160	200	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
115	14165408	200	35	1	6	0	0	0	0	*3	*3	*3	*3	3	0	0	1	6	0	3	0	3	Y	TOMO
116	7210945	200	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	N	EXTRA
117	9695397	200	38	1	6	0	0	0	0	*3	*3	*3	*3	*3	3	0	1	6	0	3	0	3	Y	TOMO
118	9695549	300	18	1	6	0	0	0	0	1	0	1	0	0	0	0	1	6	0	1	0	1	N	EXTRA
119	10148829	200	18	1	6	0	0	0	0	1	0	1	0	0	0	0	1	6	0	1	0	1	N	EXTRA
120	9096972	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
121	9165761	120	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
122	9173365	200	58	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	1	3	0	3	Y	TOMO
123	9173374	200	58	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	1	3	0	3	Y	TOMO
124	9753485	200	52	*3	6	*3	*3	1	0	*3	*3	*3	*3	*3	*3	*3	3	6	0	3	0	3	Y	TOMO
125	9753489	200	58	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	1	3	0	3	Y	TOMO
126	9753497	200	57	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	3	0	3	Y	TOMO
127	9753949	200	57	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	3	0	3	Y	TOMO
128	9755013	200	43	1	6	0	0	0	0	*3	*3	*3	*3	*3	*3	*3	3	6	0	3	0	3	Y	TOMO
129	9941081	200	57	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	3	6	0	3	0	3	Y	TOMO
130	14000376	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
131	14077668	300	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	3	3	0	3	Y	TOMO
132	9038699	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
133	9064568	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
134	9093975	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
135	9644101	120	58	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	1	0	3	Y	TOMO
136	9703873	200	61	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	1	3	Y	TOMO
137	9716853	120	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	0	3	Y	TOMO
138	9735129	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
139	9818433	200	61	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	1	3	Y	TOMO
140	10094253	120	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
141	14383980	300	16	1	6	0	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	1	N	EXTRA
142	3298292	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
143	9064093	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
144	7112721	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA

145	9069997	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
146	9070083	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
147	9105672	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
148	9128775	200	41	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	1	3	0	3	Y	TOMO
149	9132433	120	4	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
150	9140050	120	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	0	3	Y	TOMO
151	9169867	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
152	9627721	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
153	9627953	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
154	9652545	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
155	9655209	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
156	9666905	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
157	10972299	120	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	0	3	Y	TOMO
158	9734033	120	37	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	1	1	0	3	Y	TOMO
159	13692644	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
160	13935988	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
161	13936432	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
162	13936596	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
163	13936812	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
164	13938812	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
165	13939856	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
166	14079184	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	0	3	Y	TOMO
167	10059745	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
168	14116920	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
169	14116972	200	60	*3	6	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	*3	6	*3	3	0	3	Y	TOMO
170	14155260	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
171	14158696	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
172	10148369	120	39	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	1	0	3	Y	TOMO
173	10148421	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
174	10187953	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
175	14239184	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
176	10370141	300	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	N	EXTRA
177	9085734	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
178	9086693	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
179	3317364	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
180	9096656	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
181	9627557	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
182	9700049	120	26	*3	0	*3	*3	*3	*3	1	0	0	0	0	0	0	0	6	0	1	0	3	Y	TOMO
183	9718013	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
184	9742277	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
185	9774569	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
186	13813696	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
187	9853417	120	12	1	6	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	1	Y	EXTRA
188	9915709	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
189	14073800	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
190	14118096	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
191	14151344	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
192	10223765	200	41	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	1	0	3	Y	TOMO
193	9826789	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
194	14133048	200	34	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	1	6	0	1	0	3	Y	TOMO
195	14183744	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
196	14236768	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
197	14255632	120	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
198	10230869	200	24	1	6	0	0	0	0	3	0	1	0	0	0	0	3	6	0	3	0	1	Y	EXTRA
199	9148510	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
200	9150059	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
201	9152745	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
202	9154092	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
203	9154179	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
204	9154233	200	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
205	9722529	200	18	1	6	0	0	0	0	1	0	1	0	0	0	0	1	6	0	1	0	1	Y	EXTRA

206	9722633	200	40	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
207	9817605	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
208	13966396	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
209	13966672	200	10	1	0	0	0	0	0	1	0	0	0	0	0	0	0	6	0	1	0	1	Y	EXTRA
210	13970876	200	29	1	6	0	0	0	0	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
211	14178184	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
212	14178188	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
213	14178212	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
214	14178236	300	38	*3	6	*3	1	0	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
215	14178248	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
216	14179288	120	22	1	6	0	0	0	0	3	0	1	0	0	0	0	1	6	0	3	0	1	Y	EXTRA
217	14179292	300	29	1	6	0	0	0	0	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
218	14179736	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
219	9075784	120	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
220	9075803	200	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
221	12456160	120	34	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	1	6	0	1	0	3	Y	TOMO
222	14072464	200	27	1	6	0	0	0	0	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
223	10186185	200	41	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
224	10207681	200	38	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	1	0	3	Y	TOMO
225	10215753	300	40	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
226	14263544	200	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
227	14263712	120	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
228	14263768	200	40	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	3	6	0	3	0	3	Y	TOMO
229	10226877	120	39	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	*3	6	*3	1	0	3	Y	TOMO
230	9146641	300	43	*3	6	*3	*3	*3	*3	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
231	9660449	120	4	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	Y	EXTRA
232	9944301	200	29	1	6	0	0	0	0	3	0	1	0	0	0	0	*3	6	*3	3	0	3	Y	TOMO
233	14137160	120	34	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	1	6	0	1	0	3	Y	TOMO
234	14181056	120	34	*3	6	*3	*3	*3	*3	1	0	1	0	0	0	0	1	6	0	1	0	3	Y	TOMO
			6794	502	852	400	395	389	389	370	123	222	121	109	103	97	390	954	264	479	115	520		

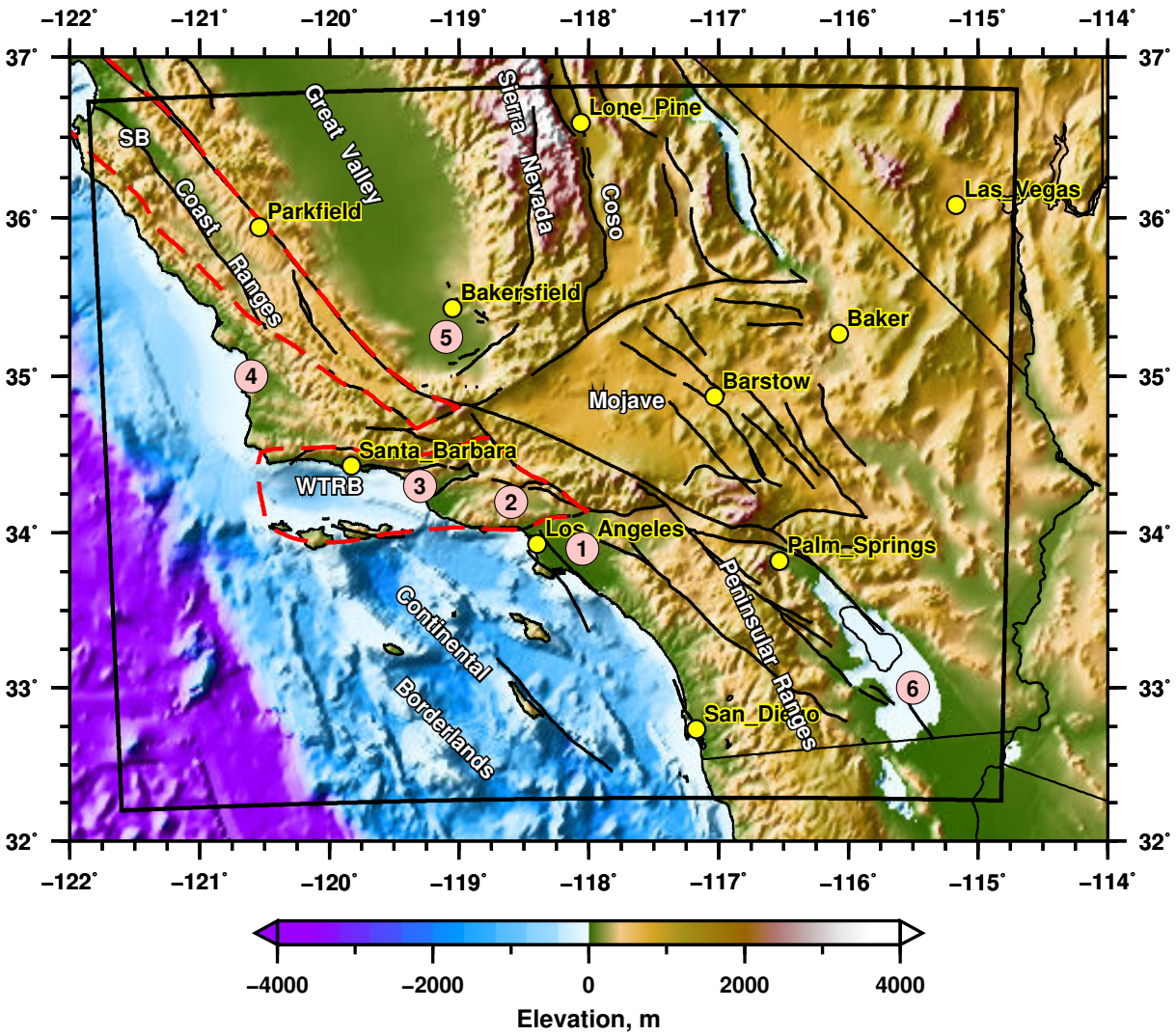


Figure 6.1: Base map for crustal tomography study of southern California. (a) Map shows topography and bathymetry (*Amante and Eakins*, 2008), as well as active faults (*Jennings*, 1994). Labels 1–6 denote the sedimentary basins of (1) Los Angeles, (2) San Fernando, (3) Ventura–Santa Barbara, (4) Santa Maria, (5) southern San Joaquin, and (6) the Salton trough, all of which have been active during the Neogene. Dashed red lines outline blocks that have undergone substantial Neogene motion: the Salinian block (SB) within the Coast Ranges, and the western Transverse Ranges block (WTRB). The black outline denotes the simulation region, which extends to 60 km depth. (b: next page) Selected principal faults in addition to nine segments denoting tomographic cross sections shown in other figures. Each cross section is an extended source-station path, with the earthquake epicenter at the star and the station at the triangle. Active faults are from *Jennings* (1994), plus the Kern Canyon fault (*Nadin and Saleeby*, 2009). Faults labeled for reference in each cross section figure are drawn in bold red and labeled in the boxes: SA, San Andreas, KC, Kern Canyon, SN, Sierra Nevada, G, Garlock, CR, Camp Rock, SG, San Gabriel, SY, Santa Ynez, MC, Malibu Coast, E, Elsinore. The Camp Rock fault, as labeled, includes a connection of faults from north to south: Gravel Hills–Harper fault, Harper Lake fault, Camp Rock fault, and Emerson fault. The Malibu Coast fault is drawn in continuation to the west to mark the southern boundary of the Western Transverse Ranges block (WTRB); the Santa Ynez fault is drawn in continuation to the west to mark the northern boundary of the WTRB (*Luyendyk et al.*, 1980).

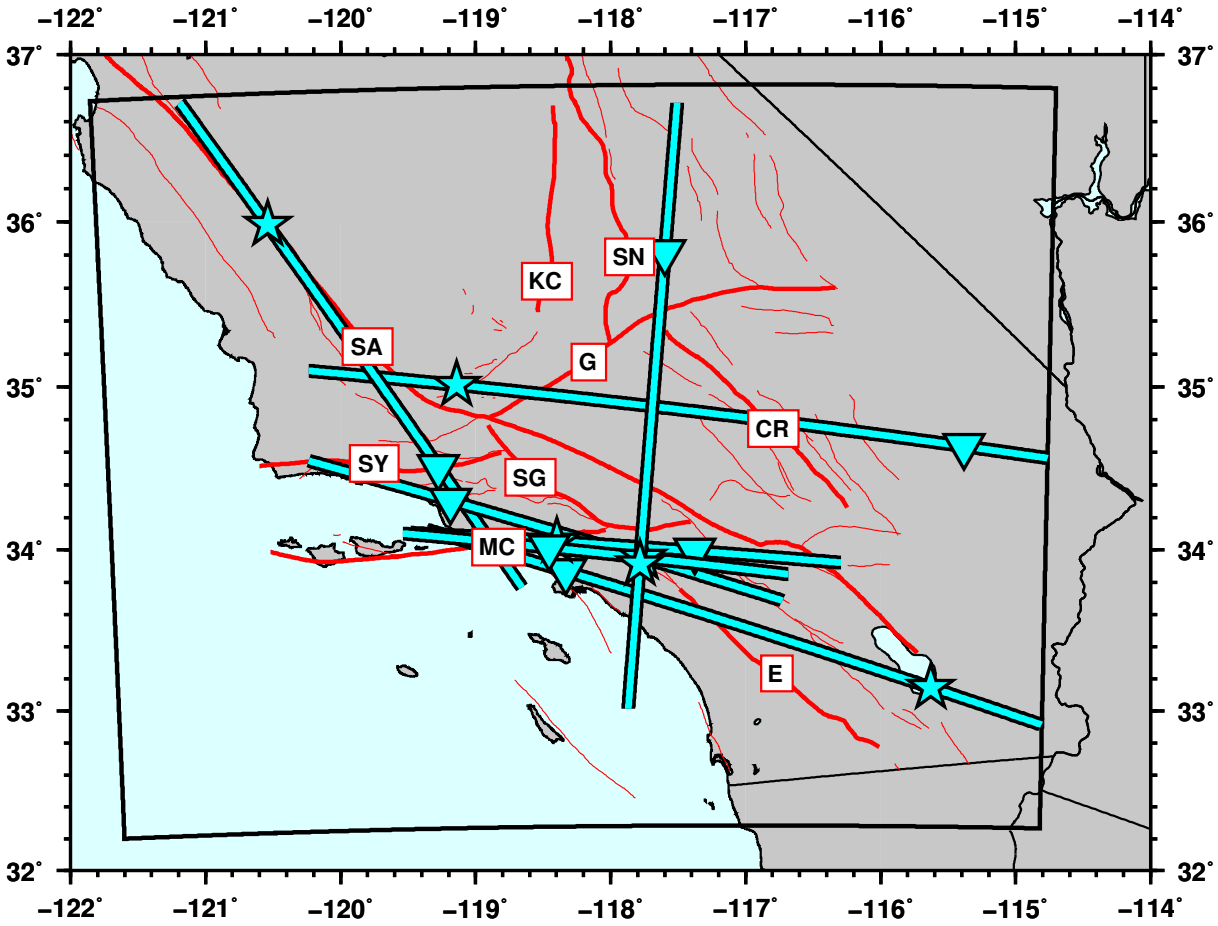

−122˚
−121˚
−120˚
−119˚
−118˚
−117˚
−116˚
−115˚
−114˚
37˚
36˚
35˚
34˚
33˚
32˚
SA
KC
SN
G
CR
SY
SG
MC
E

Figure 6.1b

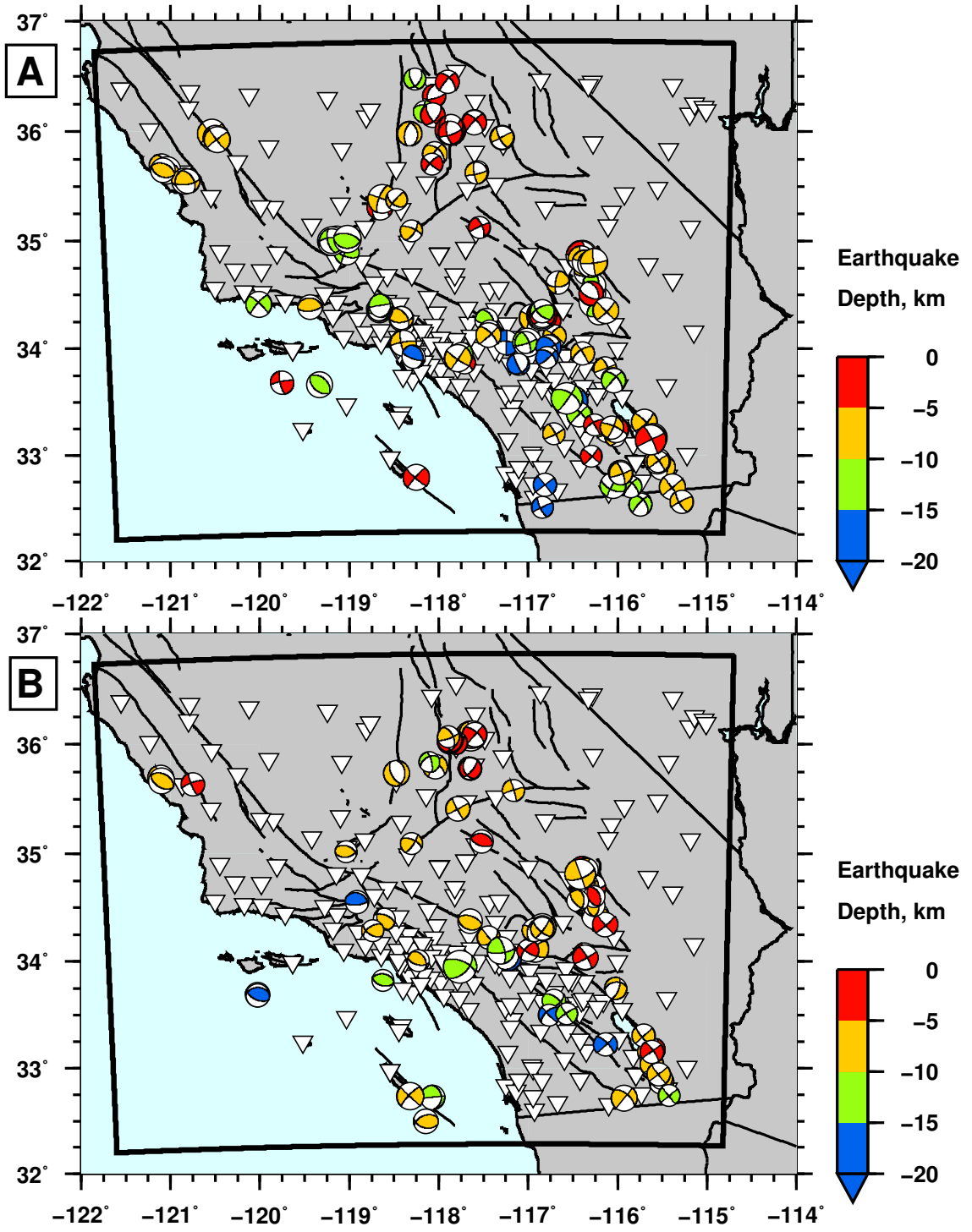


Figure 6.2: (a) Earthquake sources (143) and stations used in the tomographic inversion. Black boundary denotes our simulation region; blue boundary denotes the model of *Lin et al.* (2007b) used in the initial 3D model $\mathbf{m}_{00}$. (b) Extra earthquakes (91) used in validating the final tomographic model, but not used in the tomographic inversion. An earthquake not used in the tomographic inversion—or any future earthquake, for that matter—may be used to independently assess the misfit reduction from $\mathbf{m}_{00}$ to $\mathbf{m}_{16}$.

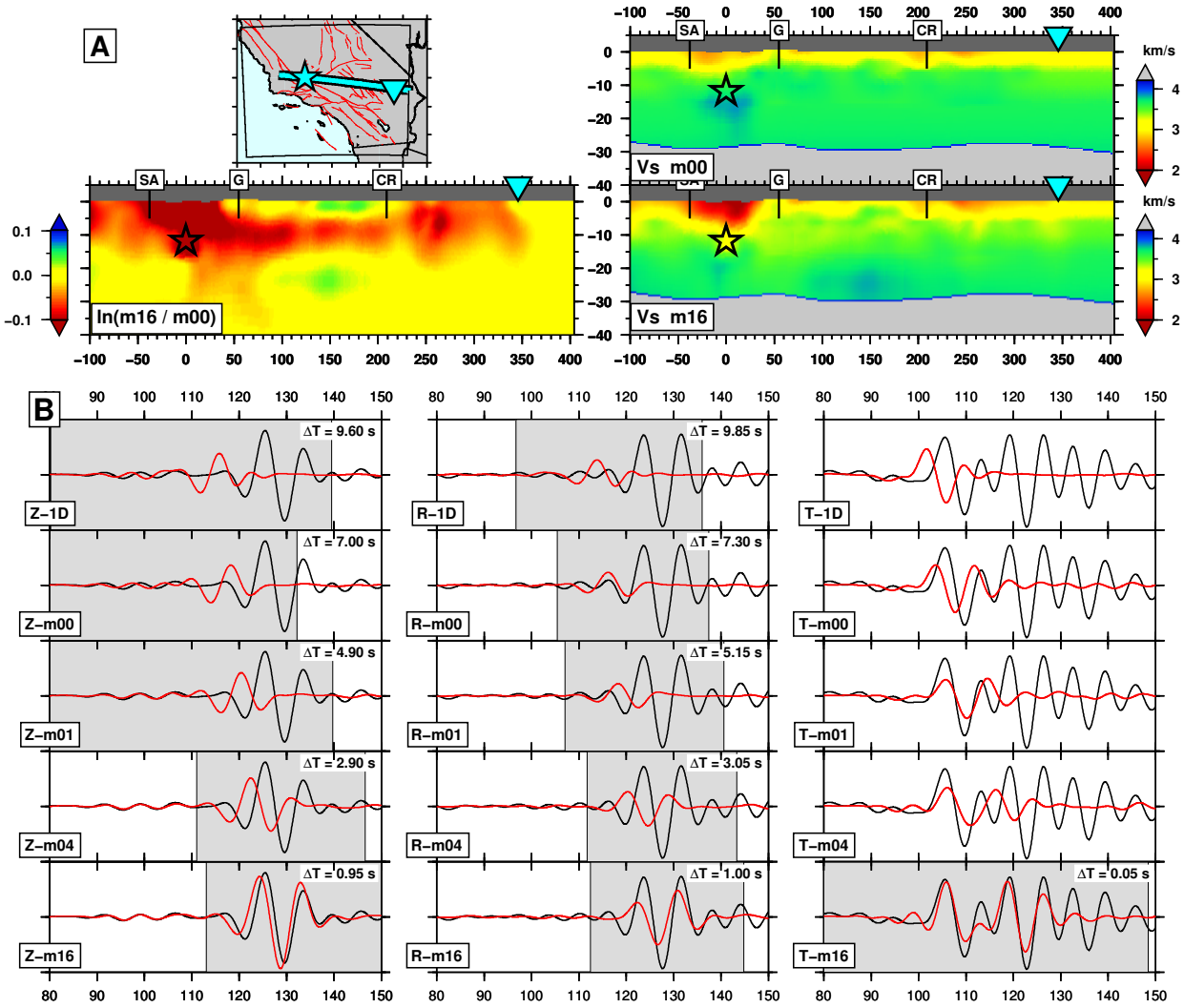


Figure 6.3: (a) Cross section of the V_S tomographic models for a path from a M_w 4.5 earthquake on the White Wolf fault (star label) to station DAN (triangle label) in the eastern Mojave Desert. Upper right is the initial 3D model, $\mathbf{m}_{00}$, lower right is the final 3D model, $\mathbf{m}_{16}$, and lower left is the difference between the two, $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$. Vertical exaggeration in these cross sections, and all cross sections in the paper, is 3.0. Boxed labels with vertical lines denote the position of various faults for reference: San Andreas (SA), Garlock (G), and Camp Rock (CR). (b) Iterative three-component seismogram fits to data for models $\mathbf{m}_{00}$, $\mathbf{m}_{01}$, $\mathbf{m}_{04}$, and $\mathbf{m}_{16}$. Also shown are synthetic seismograms computed for a standard 1D model (Table 6.1). Synthetic seismograms (red) and recorded seismograms (black), filtered over the period range 6–30 s. Left column, vertical component (Z); center column, radial component (R); right column, transverse component (T). Inset "ΔT" label indicates the time shift between the two windowed records *Maggi et al.* (2009) that provides the maximum cross-correlation.

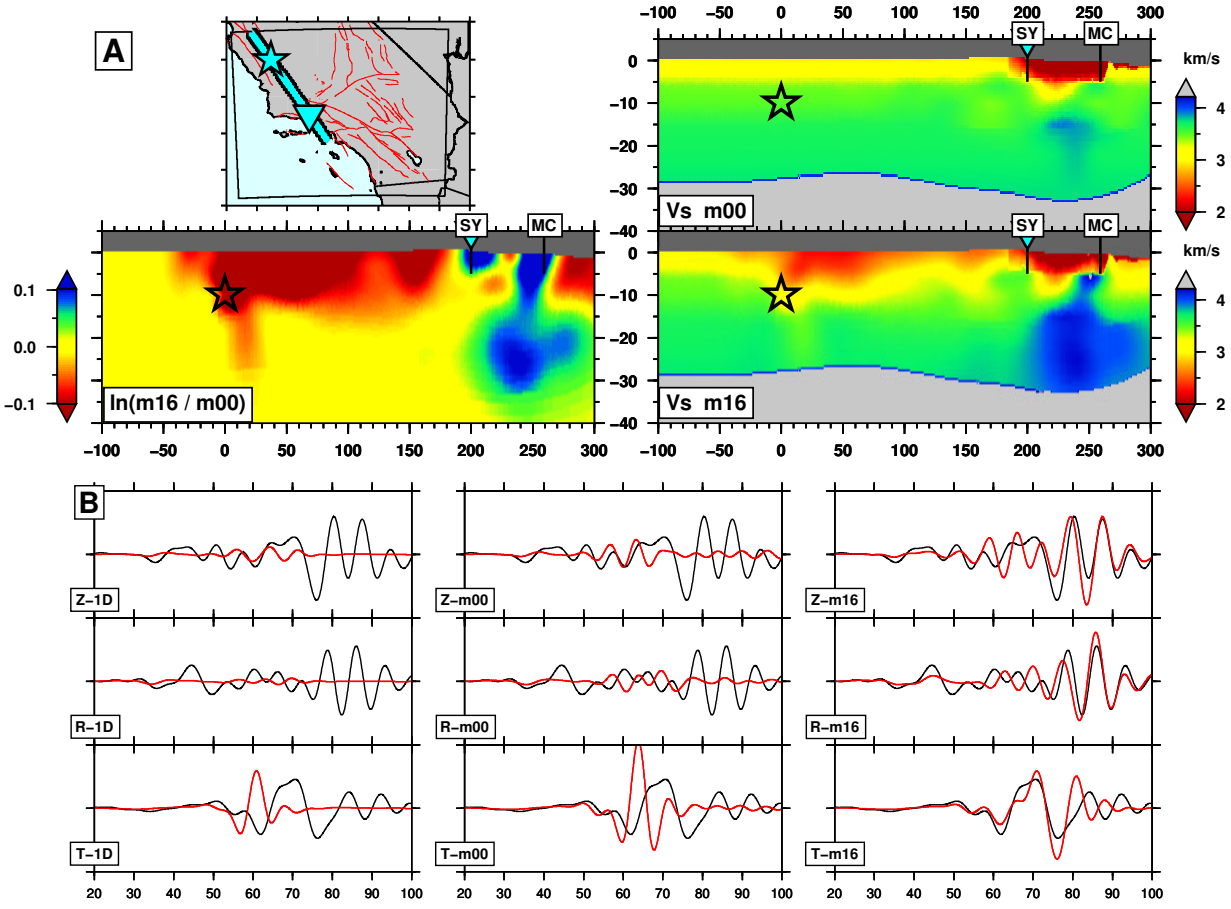


Figure 6.4: Cross sections of the tomographic model with corresponding seismograms. Measurement windows are omitted to emphasize the full waveforms. Boxed labels with vertical lines denote the location of the faults for reference: MC, Malibu Coast, SY, Santa Ynez. (a) Path for a M_w 4.6 earthquake near Parkfield to station WGR, just north of the Ventura basin. (b) Synthetic seismograms (red) and recorded seismograms (black) for the period range 6–30 s. (c: next page) Path for a M_w 5.4 earthquake near Chino Hills to station STC, within the Ventura basin. (d: next page) Corresponding seismograms.

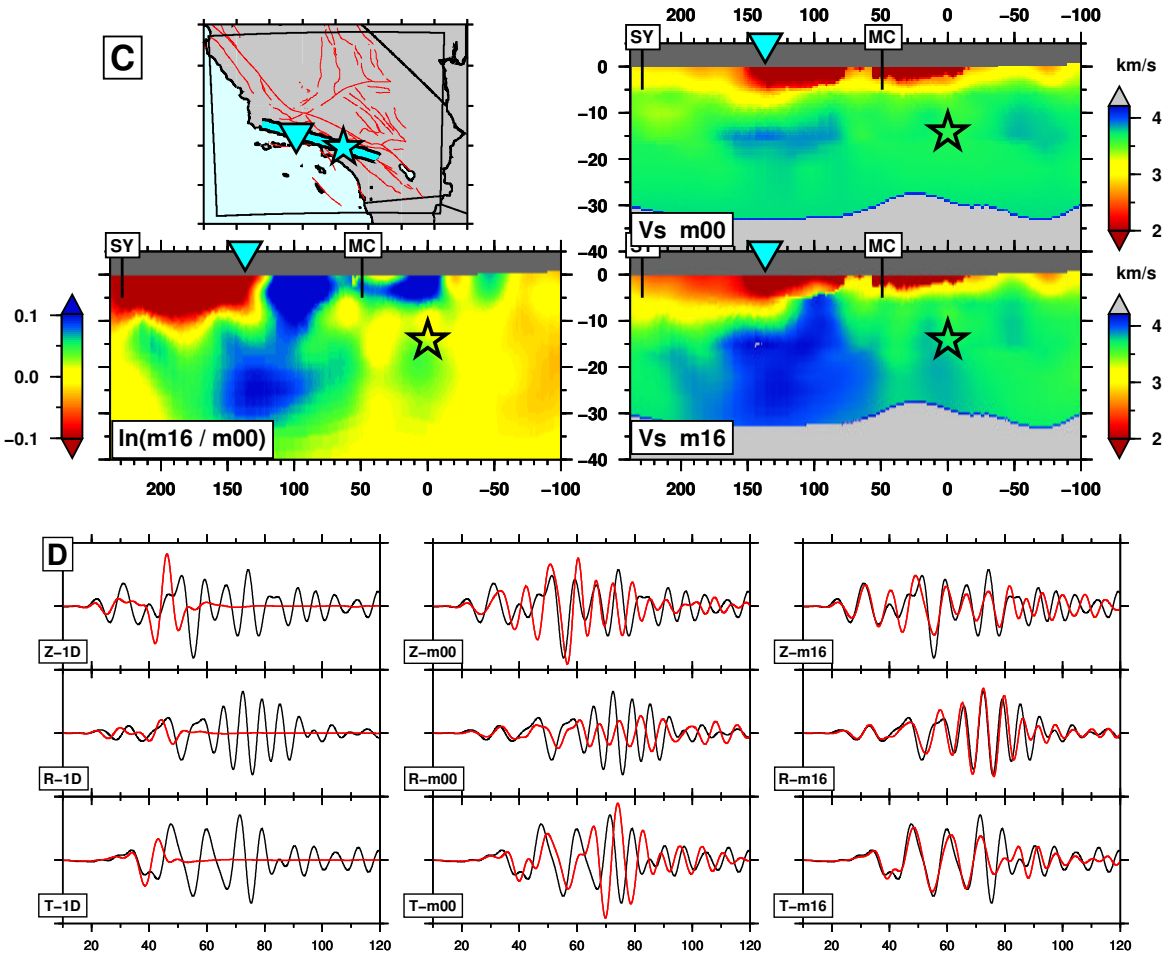

C
SY
MC
Vs m00
Vs m16
ln(m16 / m00)
km/s
0.1
0.0
−0.1
D
Z−1D
R−1D
T−1D
Z−m00
R−m00
T−m00
Z−m16
R−m16
T−m16

Figure 6.4c–d

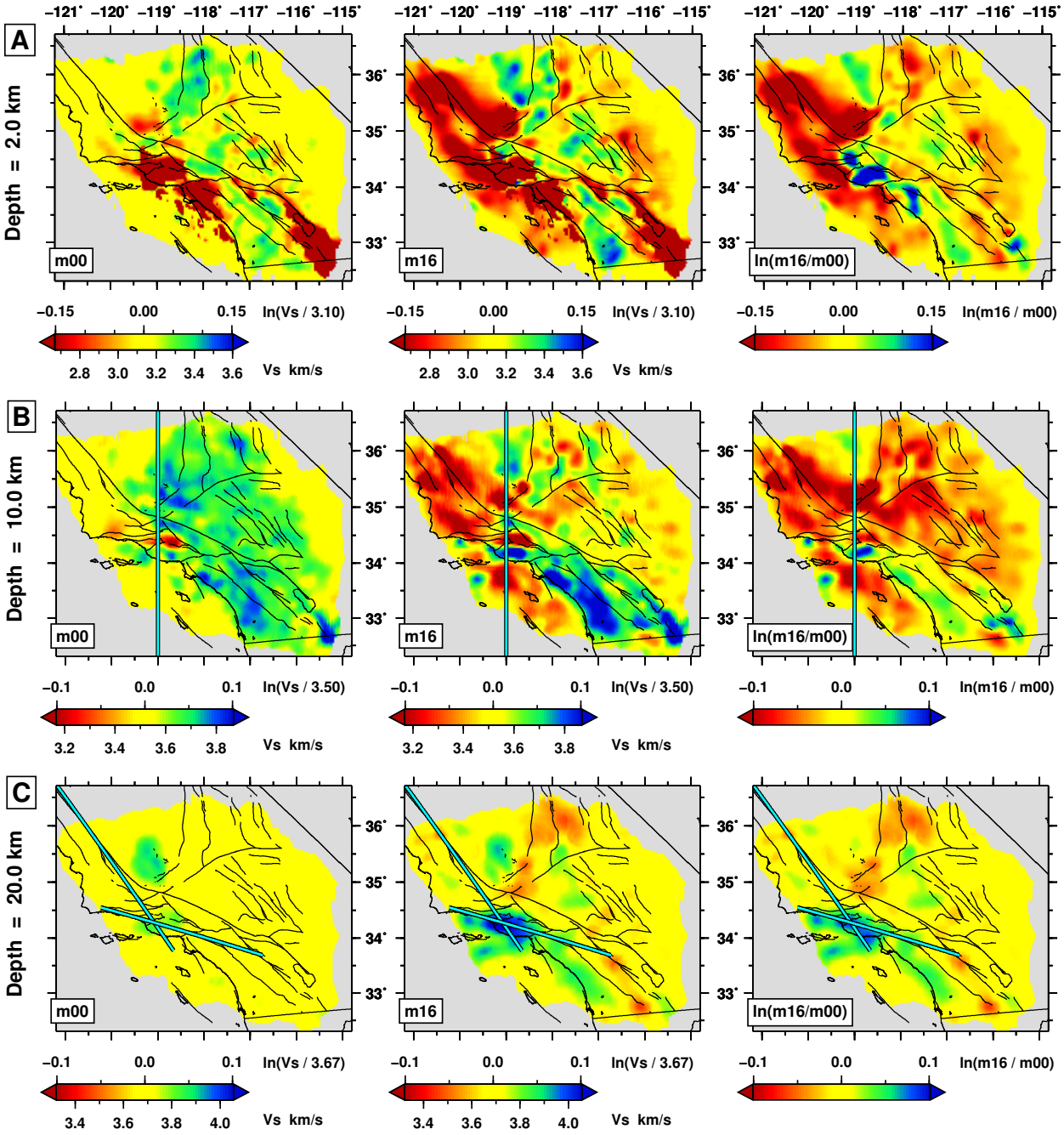


Figure 6.5: Horizontal cross sections of V_S tomographic models, with active faults shown for reference. Refer to Figure 6.1 for locations of principal faults and features. Left column shows the initial model, $\mathbf{m}_{00}$, center column shows the final model, $\mathbf{m}_{16}$, and right column shows the difference between the two models, $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$. The mask covers regions of low sensitivity to changes in V_S. (a) V_S at 2 km depth. (b) V_S at 10 km depth. The variation along longitude 119°W is discussed in the text. (c) V_S at 20 km. The two ray paths crossing the western Transverse Ranges block correspond to the profiles of Figure 6.4a, c.

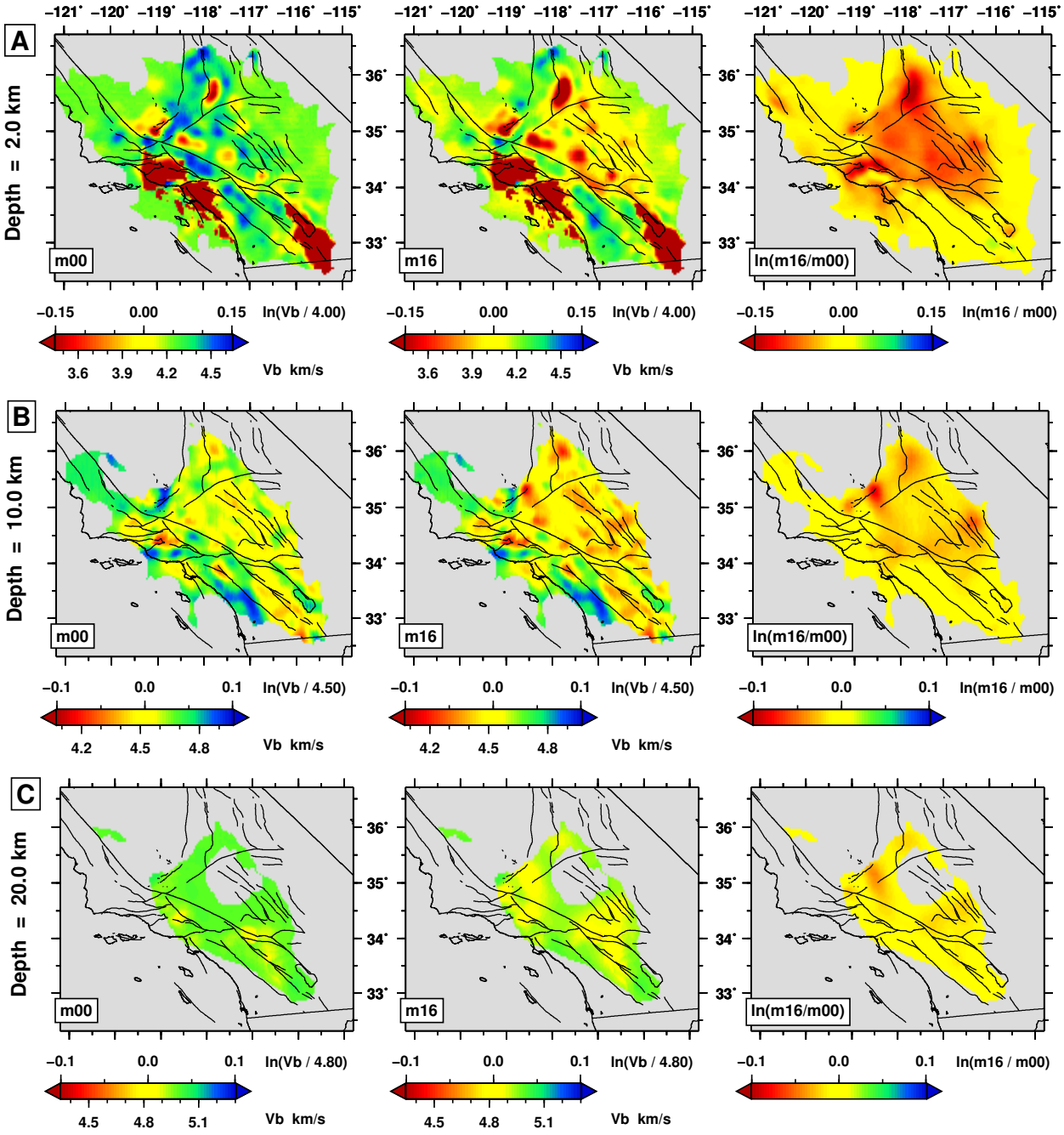


Figure 6.6: Horizontal cross sections of bulk-sound speed V_B tomographic models, with active faults shown for reference. Refer to Figure 6.1 for locations of principal faults and features. Left column shows the initial model, $\mathbf{m}_{00}$, center column shows the final model, $\mathbf{m}_{16}$, and right column shows the difference between the two models, $\ln(\mathbf{m}_{16}/\mathbf{m}_{00})$. The mask covers regions of low sensitivity to changes in V_B. (a) V_B at 2 km depth. (b) V_B at 10 km depth. (c) V_B at 20 km.

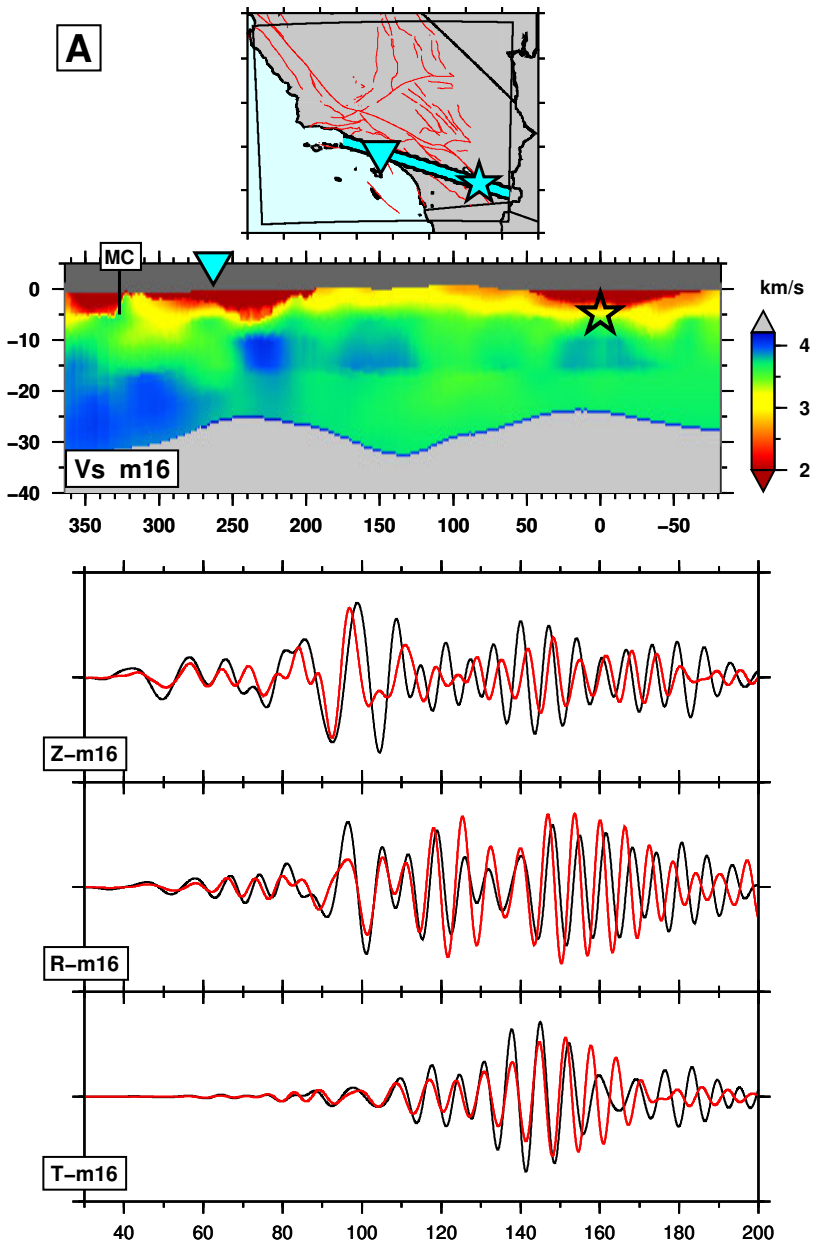


Figure 6.7: Seismogram fits between data (black) and synthetics (red) for selected paths in the final model. Each earthquake (star) is modeled as a point-source focal mechanism. Cross section shows the V_S model $\mathbf{m}_{16}$ between the source and station. Boxed labels with vertical lines denote the position of the faults for reference: SA, San Andreas, G, Garlock, SG, San Gabriel, MC, Malibu Coast. (a) Earthquake 14179736 (M_w 5.0) to station LAF.CI: period range 6–30 s, record length 170 s, source half duration 0.80 s. (b: following pages) Earthquake 9703873 (M_w 4.2) to station RVR.CI: period range 3–30 s, record length 70 s, source half duration 0.33 s. The gray time windows highlight three different Rayleigh wave arrivals whose approximate paths are marked in the adjacent inset map. The later two arrivals have interacted with the Los Angeles basin, south of the source, before reaching the station. (c: following pages) Earthquake 14383980 (M_w 5.4) to SMS.CI: period range 2–30 s, record length 50 s, source half duration 1.29 s. (d: following pages) Earthquake 9818433 (M_w 4.3) to CLC.CI: period range 2–30 s, record length 70 s, source half duration 0.39 s.

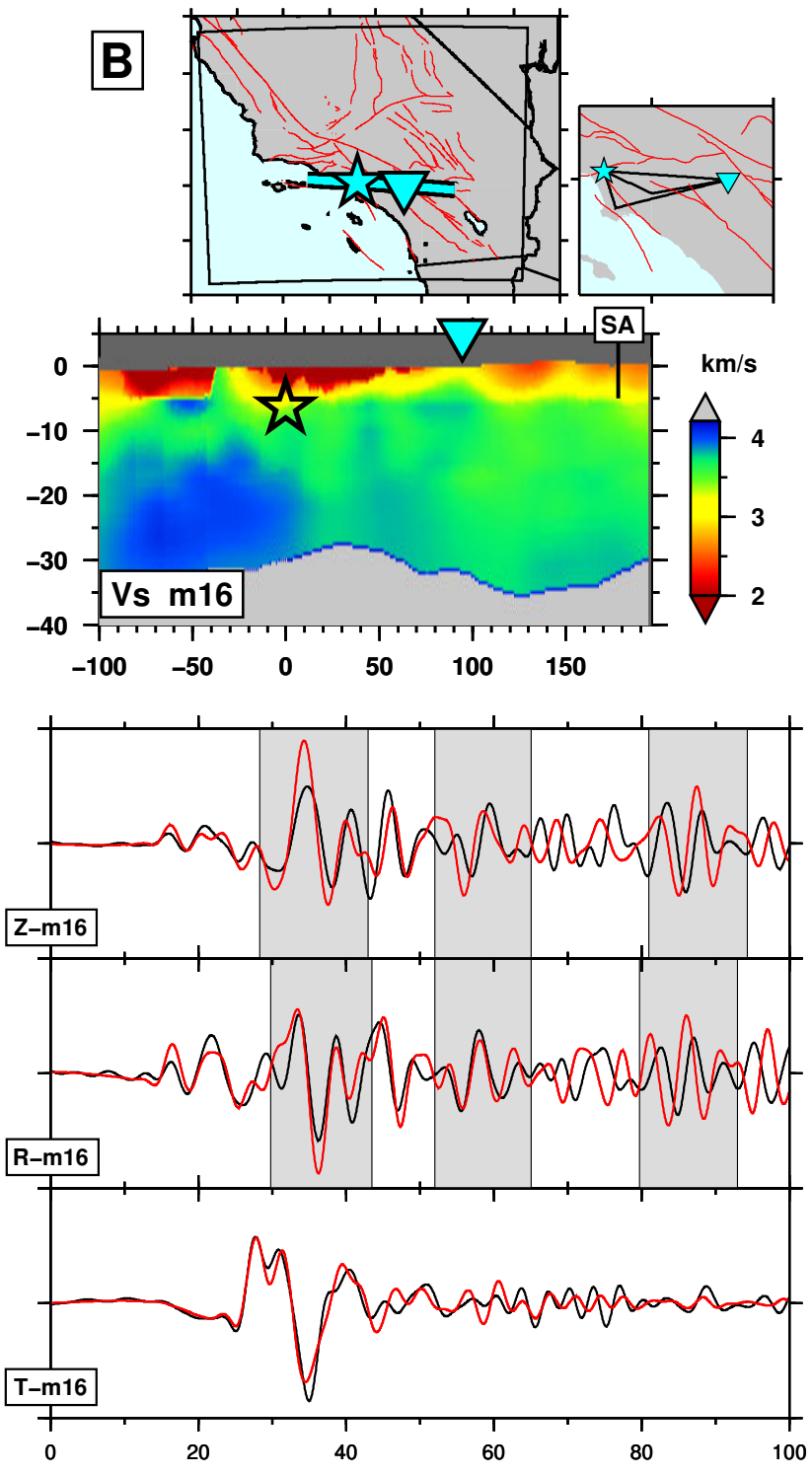

B
SA
km/s
4
3
2
0
−10
−20
−30
−40
Vs m16
−100
−50
0
50
100
150
Z−m16
R−m16
T−m16
0
20
40
60
80
100

Figure 6.7b

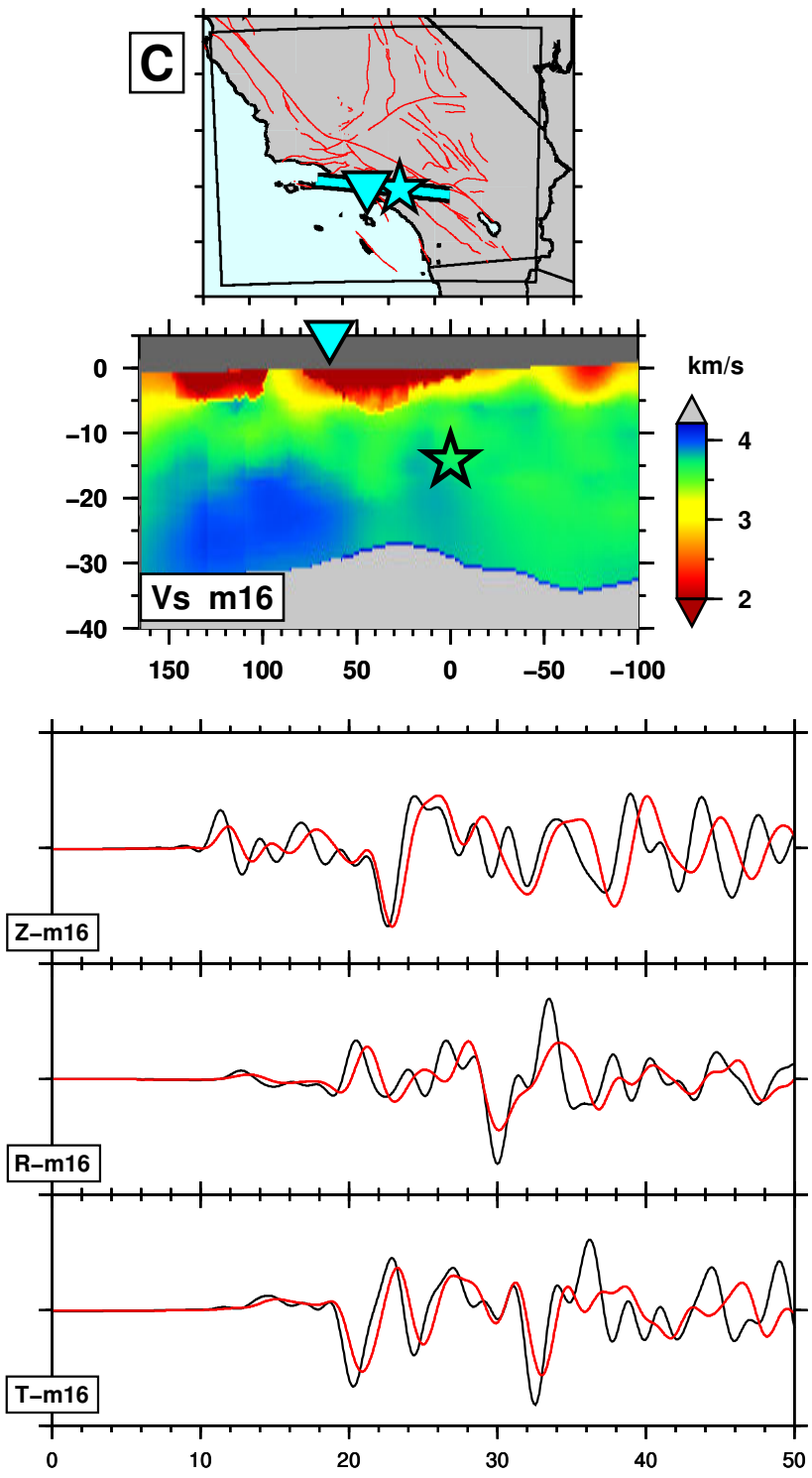

C
0
−10
−20
−30
−40
Vs m16
km/s
4
3
2
150
100
50
0
−50
−100
Z−m16
R−m16
T−m16
0
10
20
30
40
50

Figure 6.7c

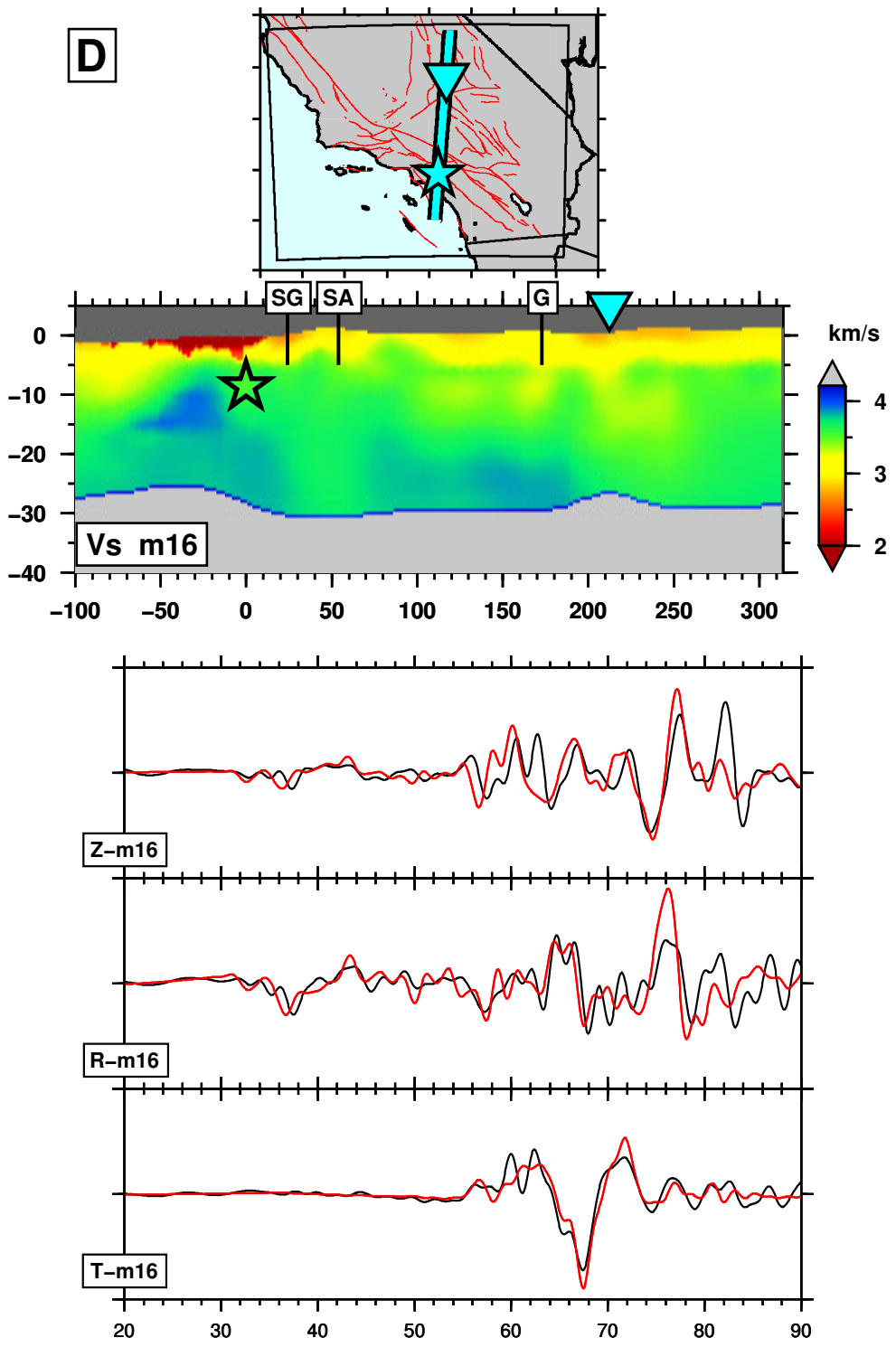

D
SG
SA
G
0
−10
−20
−30
−40
Vs m16
km/s
4
3
2
−100
−50
0
50
100
150
200
250
300
Z−m16
R−m16
T−m16
20
30
40
50
60
70
80
90

Figure 6.7d

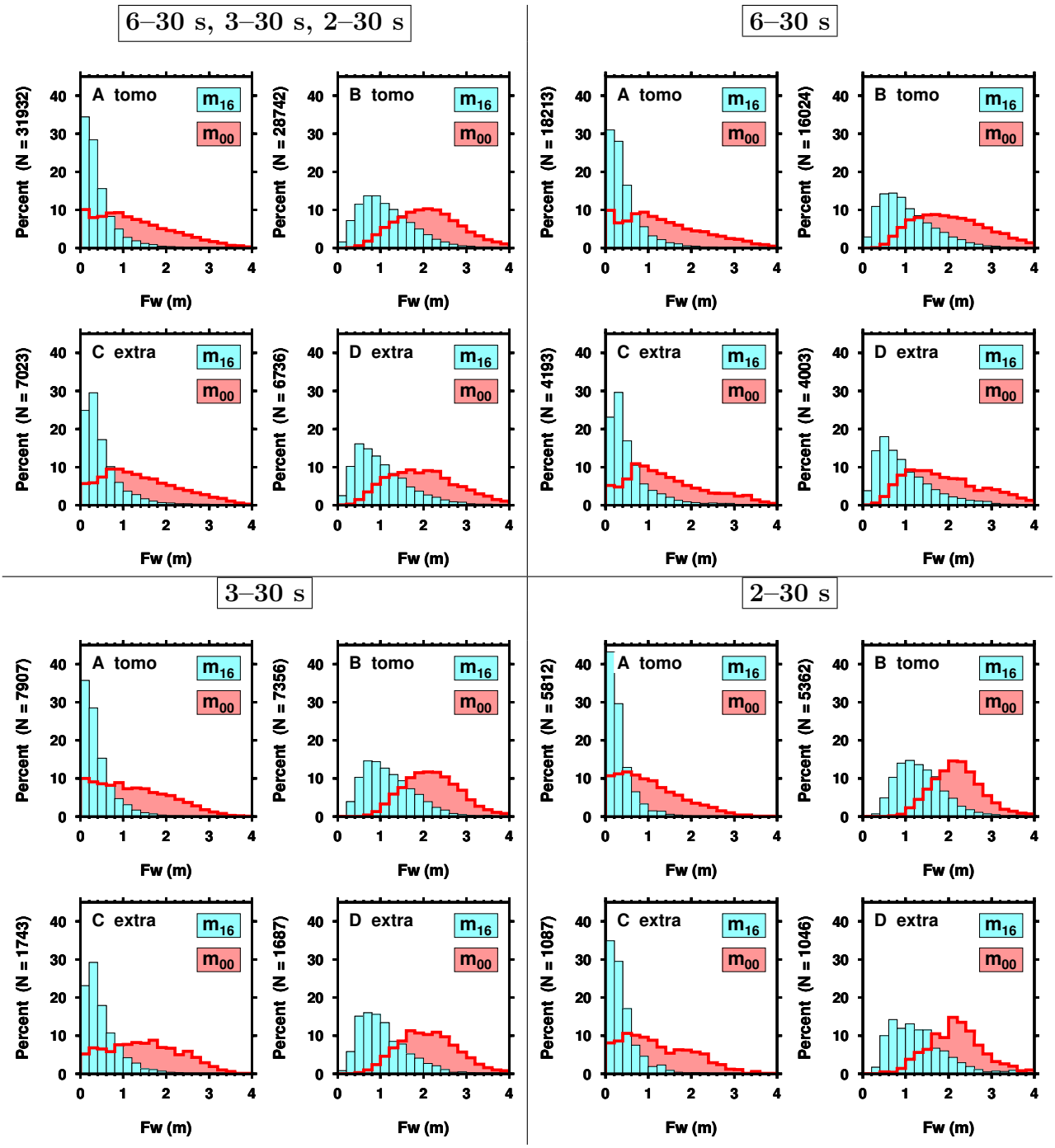


Figure 6.8: Waveform misfit analysis for the initial and final tomographic models, comparing 143 earthquakes used in the inversion (a, b: "tomo") with 91 additional earthquakes not used in the inversion (c, d: "extra"). There are four blocks of four subplots labeled (a)–(d). Each block corresponds to a different period range: all three period ranges (6–30 s, 3–30 s, and 2–30 s) (upper left), 6–30 s only (upper right), 3–30 s only (lower left), and 2–30 s only (lower right). (a) Waveform difference misfit values, $F_{\mathrm{w}}(\mathbf{m})$, for windows used in the inversion. $F_{\mathrm{w}}(\mathbf{m})$ is defined in Section 6.3. (b) Waveform difference misfit values for full seismograms containing a least one measurement window. (c)–(d) Same as (a)–(b), but for the set of extra earthquakes.

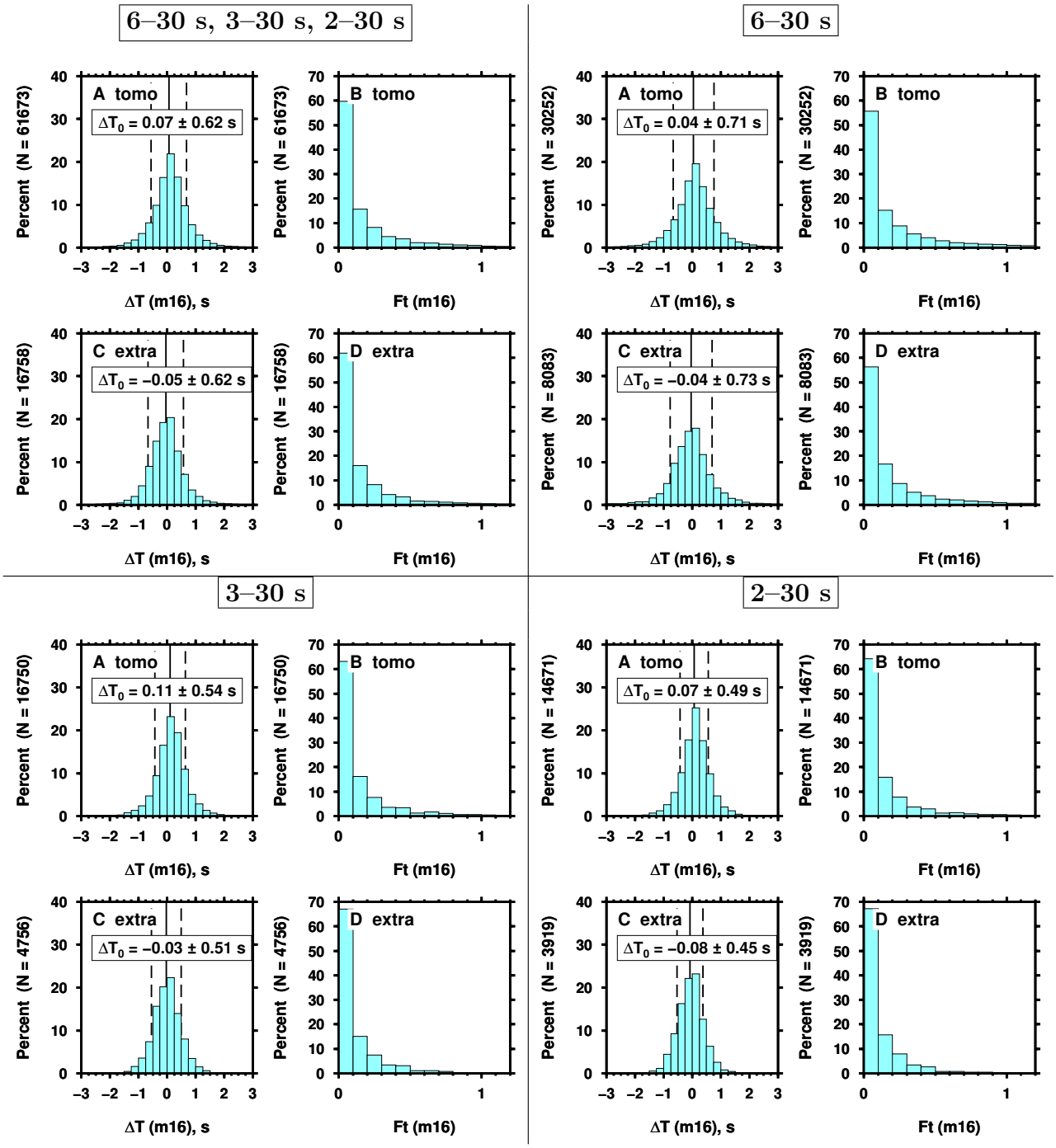


Figure 6.9: Traveltime differences within all seismogram windows used in the final tomographic model ($\mathbf{m}_{16}$) for different period ranges: all three period ranges (6–30 s, 3–30 s, and 2–30 s) (upper left), 6–30 s only (upper right), 3–30 s only (lower left), and 2–30 s only (lower right). In addition to the histogram of traveltime differences, $\Delta T(\mathbf{m}_{16})$, we also show the corresponding misfit measure, $F_{\mathrm{t}}(\mathbf{m}_{16})$.

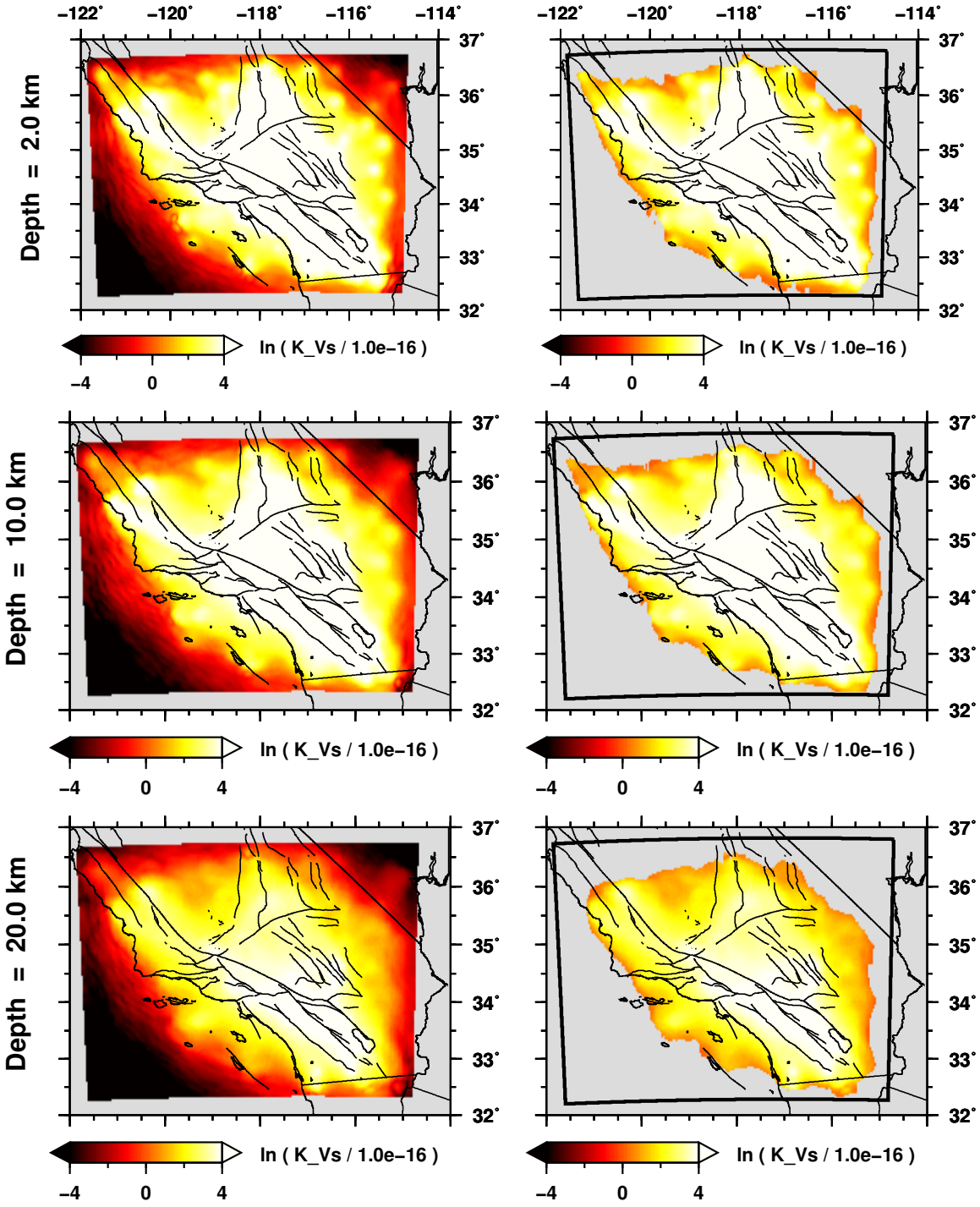


Figure 6.10: Volumetric coverage for the V_S tomographic model, plotted at three depths. The scalar field is computed as $\ln[K_{V_S}(\mathbf{x})/K_0]$, where $K_{V_S}(\mathbf{x})$ is the sum of all V_S kernels, and $K_0 = 10^{-16}$ m^{-3} is the threshold value that determines the mask shown in the right column. Note that the coverage diminishes with depth, because the short-period surface waves, which dominate the sensitivity, are more sensitive near the surface. The black outline denotes the simulation region, which extends to 60 km depth.

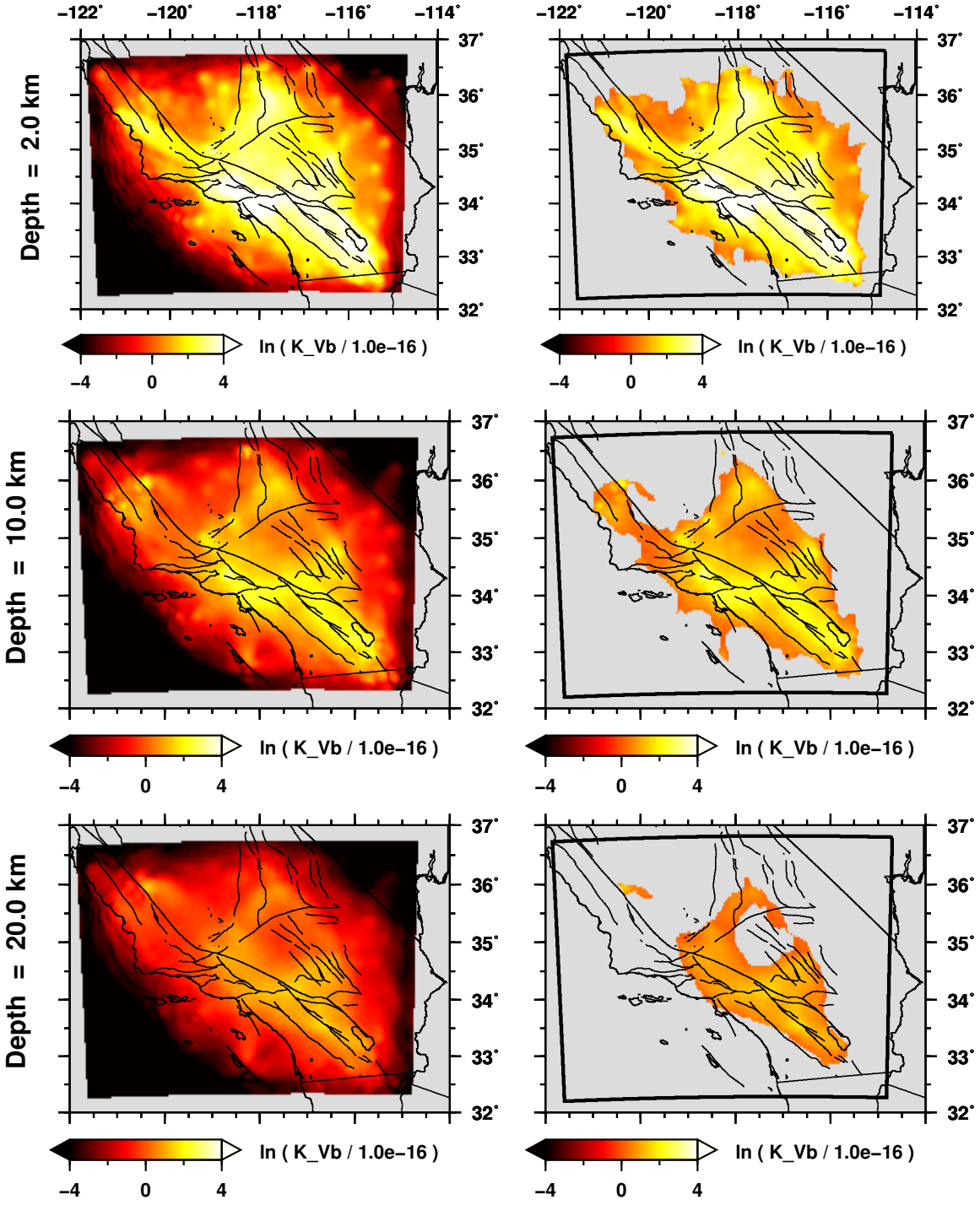


Figure 6.11: Volumetric coverage for the V_B tomographic model, plotted at three depths. The scalar field is computed as $\ln[K_{V_\mathrm{B}}(\mathbf{x})/K_0]$, where $K_{V_\mathrm{B}}(\mathbf{x})$ is the sum of all V_B kernels, and $K_0 = 10^{-16}$ m^{-3} is the threshold value that determines the mask shown in the right column.

Appendix A

Supplemental Material for "Finite-frequency tomography using adjoint methods — Methodology and examples using membrane surface waves" (Chapter 2)

Note

Table A.1 makes a qualitative comparison between "classical" tomography and "adjoint" tomography. Table A.2 highlights all possible source-structure inversion experiments.

A.1 From misfit function to adjoint source: 2D membrane-wave example

Here we derive (2.48), following *Tromp et al.* (2005), which makes use of Green's functions. Alternatively, one could also use the Lagrange multiplier method (e.g., *Liu and Tromp*, 2006; *Fichtner et al.*, 2006).

For ease of notation, we let $\mathbf{x} = (x, y)$ and consider a single event with R recording receivers. The variation in the traveltime misfit function due to a model perturbation $\boldsymbol{\delta}\mathbf{m}$ is given by (2.7):

$$\delta F = -\sum_{r=1}^{R} \Delta T_r \, \delta T_r. \tag{A.1}$$

The cross-correlation traveltime variation δT_r can be written as (*Luo and Schuster*, 1990; *Marquering et al.*, 1999)

$$\delta T_r = \frac{1}{M_{\mathrm{T}r}} \int_0^T w_r(t) \partial_t s(\mathbf{x}_r, t) \delta s(\mathbf{x}_r, t) \, \mathrm{d}t, \tag{A.2}$$

where w_r denotes the cross-correlation window, δs the change in displacement, and $M_{\mathrm{T}r}$ the normalization factor defined as

$$M_{\mathrm{T}r} = \int_0^T w_r(t) s(\mathbf{x}_r, t) \partial_t^2 s(\mathbf{x}_r, t) \, \mathrm{d}t, \tag{A.3}$$

such that $M_{\mathrm{T}r} < 0$ for a pulse with nonzero amplitude.

The equation of motion that is solved by the SEM algorithm is shown in (2.29). Using the standard Green's function approach, we write the wavefield generated by the point source (2.30) as

$$s(\mathbf{x}, t) = \int_0^t \int_\Omega G(\mathbf{x}, \mathbf{x}'; t - t') \, f(\mathbf{x}', t') \, \mathrm{d}^2\mathbf{x}' \, \mathrm{d}t'. \tag{A.4}$$

The change in displacement δs due to a change in the point force δf may be written as

$$\delta s(\mathbf{x}, t) = \int_0^t \int_\Omega G(\mathbf{x}, \mathbf{x}'; t - t') \, \delta f(\mathbf{x}', t') \, \mathrm{d}^2\mathbf{x}' \, \mathrm{d}t'. \tag{A.5}$$

Upon substitution of the perturbation (A.5) into (A.1)–(A.2) we find that the change in

the traveltime misfit function may be expressed as[1]

$$\begin{aligned}\delta F &= \int_0^T \int_\Omega \delta f(\mathbf{x},t') \left[\sum_{r=1}^{R} \Delta T_r \frac{1}{M_{\mathrm{T}r}} \int_0^{T-t'} G(\mathbf{x},\mathbf{x}_r; T-t'-t)\, w_r(T-t) \partial_t s(\mathbf{x}_r, T-t)\, \mathrm{d}t \right] \mathrm{d}^2\mathbf{x}\, \mathrm{d}t' \\ &= \int_0^T \int_\Omega \delta f(\mathbf{x},t)\, s^\dagger(\mathbf{x}, T-t)\, \mathrm{d}^2\mathbf{x}\, \mathrm{d}t, \end{aligned} \tag{A.7}$$

where we have defined the adjoint wavefield by

$$s^\dagger(\mathbf{x},t) \equiv \int_0^t \int_\Omega G(\mathbf{x},\mathbf{x}'; t-t')\, f^\dagger(\mathbf{x}',t')\, \mathrm{d}^2\mathbf{x}'\, \mathrm{d}t' \tag{A.8}$$

and the adjoint source by

$$f^\dagger(\mathbf{x},t) \equiv \sum_{r=1}^{R} \Delta T_r \frac{1}{M_{\mathrm{T}r}}\, w_r(T-t) \partial_t s(\mathbf{x}_r, T-t) \delta(\mathbf{x}-\mathbf{x}_r). \tag{A.9}$$

Note that the spatial integration in (A.8) arises from the delta function in (A.9), and also that the adjoint source includes the time-reversed synthetic velocity recorded at the rth receiver.

A.2 The conjugate gradient algorithm

The gradient is not a vector but rather a tangent plane or set of level lines (*Tarantola*, 2005, p. 205). The metric (tensor) provides a means for selecting the steepest descent vector; using a different metric will lead to a different steepest descent vector. The metric also appears in the conjugate gradient algorithm, and thus the choices of metric will affect the optimization.

[1]The 3D version of Equation (A.7) is given by

$$\delta F = \int_0^T \int_V \delta \mathbf{f}(\mathbf{x},t) \cdot \mathbf{s}^\dagger(\mathbf{x}, T-t)\, \mathrm{d}^3\mathbf{x}\, \mathrm{d}t\ . \tag{A.6}$$

A.2.1 Background and notation

The model covariance matrix $\mathbf{C}$ defines the relationship between the gradient $\hat{\mathbf{g}}$ and the corresponding steepest ascent vector $\mathbf{g}$:

$$\mathbf{g} = \mathbf{C}\,\hat{\mathbf{g}}\,, \tag{A.10}$$

$$\hat{\mathbf{g}} = \mathbf{C}^{-1}\,\mathbf{g}. \tag{A.11}$$

Similarly, for the model vector,

$$\mathbf{m} = \mathbf{C}\,\hat{\mathbf{m}}\,, \tag{A.12}$$

$$\hat{\mathbf{m}} = \mathbf{C}^{-1}\,\mathbf{m}. \tag{A.13}$$

The L2-norm can be defined over either space:

$$\begin{aligned} \|\mathbf{m}\|_2 &= \left(\mathbf{m}^T\,\mathbf{C}^{-1}\,\mathbf{m}\right)^{1/2} = \left((\mathbf{C}\,\hat{\mathbf{m}})^T\;\mathbf{C}^{-1}\;(\mathbf{C}\,\hat{\mathbf{m}})\right)^{1/2} = \left(\hat{\mathbf{m}}^T\,\mathbf{C}^T\,\mathbf{C}^{-1}\,\mathbf{C}\,\hat{\mathbf{m}}\right)^{1/2} \\ &= \left(\hat{\mathbf{m}}^T\,\mathbf{C}\,\hat{\mathbf{m}}\right)^{1/2} = \left(\hat{\mathbf{m}}^T\,\hat{\mathbf{C}}^{-1}\,\hat{\mathbf{m}}\right)^{1/2} = \|\hat{\mathbf{m}}\|_2\,, \end{aligned} \tag{A.14}$$

where

$$\hat{\mathbf{C}} = \mathbf{C}^{-1}.$$

The duality product between the steepest ascent vector and the gradient can be written in several ways:

$$\langle \mathbf{g},\, \hat{\mathbf{g}} \rangle = \mathbf{g}^T\,\hat{\mathbf{g}} = \mathbf{g}^T\mathbf{C}^{-1}\,\mathbf{g} = \|\mathbf{g}\|_2^2\ , \tag{A.15}$$

$$\langle \mathbf{g},\, \hat{\mathbf{g}} \rangle = \mathbf{g}^T\hat{\mathbf{g}} = (\mathbf{C}\,\hat{\mathbf{g}})^T\,\hat{\mathbf{g}} = \hat{\mathbf{g}}^T\mathbf{C}\,\hat{\mathbf{g}} = \|\hat{\mathbf{g}}\|_2^2\ . \tag{A.16}$$

This shows that the norm of the steepest ascent vector is equal to the norm of the gradient, as expected.

A.2.2 Algorithm

The conjugate gradient algorithm we use may be summarized as follows: given an initial model $\mathbf{m}^0$, calculate $F(\mathbf{m}^0)$, $\hat{\mathbf{g}}^0 = \partial F/\partial \mathbf{m}(\mathbf{m}^0)$, and set the initial conjugate gradient search direction equal to minus the initial gradient of the misfit function,

$$\mathbf{p}^0 = -\mathbf{g}^0 = -\mathbf{C}\,\hat{\mathbf{g}}^0. \tag{A.17}$$

If $||\mathbf{p}^0|| < \epsilon$, where ϵ is a suitably small number, then $\mathbf{m}^0$ is the model we seek to determine, otherwise:

1. We denote a model in the direction of the search vector as, and its corresponding gradient, as

$$\mathbf{m}_\nu^k \equiv \mathbf{m}^k + \nu \mathbf{p}^k\,, \tag{A.18}$$
$$\hat{\mathbf{g}}_\nu^k \equiv \frac{\partial F}{\partial \mathbf{m}}\left(\mathbf{m}_\nu^k\right). \tag{A.19}$$

   Perform a line search to obtain the scalar ν^k that minimizes the function $\tilde{F}^k(\nu)$ where

$$\tilde{F}^k(\nu) = F(\mathbf{m}_\nu^k)\,, \tag{A.20}$$
$$\tilde{g}^k(\nu) = \frac{\partial \tilde{F}^k}{\partial \nu} = \left\langle \hat{\mathbf{g}}_\nu^k, \mathbf{p}^k \right\rangle = \left(\hat{\mathbf{g}}_\nu^k\right)^T \mathbf{p}^k\,. \tag{A.21}$$

   - Choose a test parameter $\nu_t^k = -2\tilde{F}^k(0)/\tilde{g}^k(0)$, based on quadratic extrapolation.
   - Calculate the test model $\mathbf{m}_t^k = \mathbf{m}^k + \nu_t^k \mathbf{p}^k$.
   - Calculate $F(\mathbf{m}_t^k)$ and, for cubic interpolation, $\hat{\mathbf{g}}_t^k = \partial F/\partial \mathbf{m}(\mathbf{m}_t^k)$.
   - Interpolate the function $\tilde{F}^k(\nu)$ by a quadratic or cubic polynomial and obtain the ν^k that gives the (analytical) minimum value of this polynomial.

2. Update the model: $\mathbf{m}^{k+1} = \mathbf{m}^k + \nu^k \mathbf{p}^k$, then calculate

$$\mathbf{g}^{k+1} = \mathbf{C}\,\hat{\mathbf{g}}^{k+1} = \mathbf{C}\,\frac{\partial F}{\partial \mathbf{m}}(\mathbf{m}^{k+1}). \tag{A.22}$$

3. Update the conjugate gradient search direction: $\mathbf{p}^{k+1} = -\mathbf{g}^{k+1} + \beta^{k+1}\mathbf{p}^k$, where

$$\beta^{k+1} = \frac{\left\langle \hat{\mathbf{g}}^{k+1} - \hat{\mathbf{g}}^k,\ \mathbf{g}^{k+1} \right\rangle}{\left\langle \hat{\mathbf{g}}^k, \mathbf{g}^k \right\rangle} = \frac{\left(\hat{\mathbf{g}}^{k+1} - \hat{\mathbf{g}}^k\right)^T \mathbf{C}\, \hat{\mathbf{g}}^{k+1}}{\left(\hat{\mathbf{g}}^k\right)^T \mathbf{C}\, \hat{\mathbf{g}}^k} \ . \tag{A.23}$$

4. If $||\mathbf{p}^{k+1}|| < \epsilon$, then $\mathbf{m}^{k+1}$ is the desired model; otherwise replace k with $k+1$ and restart from Step 1.

A.2.3 Inversion details of *Tape et al.* (2007)

Here we show how the description of the CG algorithm in *Tape et al.* (2007) leads to the general expressions in Section A.2. We use the tilde notation (e.g., $\widetilde{\mathbf{m}}$) to distinguish the notation in *Tape et al.* (2007) from the notation previously discussed.

From the CG algorithm (Section A.2.2), the first test model is given by

$$\widetilde{\mathbf{m}}_t^0 = \widetilde{\mathbf{m}}^0 + \widetilde{\nu}_t^0\ \widetilde{\mathbf{p}}^0 = \widetilde{\mathbf{m}}^0 - \widetilde{\nu}_t^0\ \widetilde{\mathbf{C}}\ \widetilde{\hat{\mathbf{g}}}^0 \ , \tag{A.24}$$

where the step length is

$$\widetilde{\nu}_t^0 = -\frac{2\, F\left(\widetilde{\mathbf{m}}^0\right)}{\left(\widetilde{\hat{\mathbf{g}}}^0\right)^T \widetilde{\mathbf{C}}\, \widetilde{\hat{\mathbf{g}}}^0} \ . \tag{A.25}$$

In *Tape et al.* (2007) we expanded the model into orthonormal basis functions and scaled the source parameters in a manner that allowed us to use

$$\widetilde{\mathbf{C}} = \mathbf{I} \tag{A.26}$$

in Equations (A.24) and (A.25).

The model vector and gradient vector are shown within the schematic expression for the (test) model update,

$$\widetilde{\boldsymbol{\delta}\mathbf{m}} \ = \ \widetilde{\mathbf{m}}^{0t} - \widetilde{\mathbf{m}}^0 = -\widetilde{\nu}_t^0\ \widetilde{\hat{\mathbf{g}}}^0 \ , \tag{A.27}$$

which is expanded as

$$\widetilde{\boldsymbol{\delta}\mathbf{m}} = \begin{bmatrix} C_1^{0t}\sqrt{V_1}/J \\ \vdots \\ C_i^{0t}\sqrt{V_i}/J \\ \vdots \\ C_H^{0t}\sqrt{V_H}/J \\ B_1^{0t}\sqrt{V_1}/J \\ \vdots \\ B_i^{0t}\sqrt{V_i}/J \\ \vdots \\ B_H^{0t}\sqrt{V_H}/J \\ (T_s)_1^{0t}/\widetilde{\sigma}_{t_s} \\ (X_s)_1^{0t}/\widetilde{\sigma}_{x_s} \\ (Y_s)_1^{0t}/\widetilde{\sigma}_{y_s} \\ (Z_s)_1^{0t}/\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_s^{0t}/\widetilde{\sigma}_{t_s} \\ (X_s)_s^{0t}/\widetilde{\sigma}_{x_s} \\ (Y_s)_s^{0t}/\widetilde{\sigma}_{y_s} \\ (Z_s)_s^{0t}/\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_S^{0t}/\widetilde{\sigma}_{t_s} \\ (X_s)_S^{0t}/\widetilde{\sigma}_{x_s} \\ (Y_s)_S^{0t}/\widetilde{\sigma}_{y_s} \\ (Z_s)_S^{0t}/\widetilde{\sigma}_{z_s} \end{bmatrix} - \begin{bmatrix} C_1^{0}\sqrt{V_1}/J \\ \vdots \\ C_i^{0}\sqrt{V_i}/J \\ \vdots \\ C_H^{0}\sqrt{V_H}/J \\ B_1^{0}\sqrt{V_1}/J \\ \vdots \\ B_i^{0}\sqrt{V_i}/J \\ \vdots \\ B_H^{0}\sqrt{V_H}/J \\ (T_s)_1^{0}/\widetilde{\sigma}_{t_s} \\ (X_s)_1^{0}/\widetilde{\sigma}_{x_s} \\ (Y_s)_1^{0}/\widetilde{\sigma}_{y_s} \\ (Z_s)_1^{0}/\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_s^{0}/\widetilde{\sigma}_{t_s} \\ (X_s)_s^{0}/\widetilde{\sigma}_{x_s} \\ (Y_s)_s^{0}/\widetilde{\sigma}_{y_s} \\ (Z_s)_s^{0}/\widetilde{\sigma}_{z_s} \\ \vdots \\ (T_s)_S^{0}/\widetilde{\sigma}_{t_s} \\ (X_s)_S^{0}/\widetilde{\sigma}_{x_s} \\ (Y_s)_S^{0}/\widetilde{\sigma}_{y_s} \\ (Z_s)_S^{0}/\widetilde{\sigma}_{z_s} \end{bmatrix} = -\widetilde{\nu}_t^0 \begin{bmatrix} K_1^C\sqrt{V_1}\,J \\ \vdots \\ K_i^C\sqrt{V_i}\,J \\ \vdots \\ K_H^C\sqrt{V_H}\,J \\ K_1^B\sqrt{V_1}\,J \\ \vdots \\ K_i^B\sqrt{V_i}\,J \\ \vdots \\ K_H^B\sqrt{V_H}\,J \\ K_1^{t_s}\widetilde{\sigma}_{t_s} \\ K_1^{x_s}\widetilde{\sigma}_{x_s} \\ K_1^{y_s}\widetilde{\sigma}_{y_s} \\ K_1^{z_s}\widetilde{\sigma}_{z_s} \\ \vdots \\ K_s^{t_s}\widetilde{\sigma}_{t_s} \\ K_s^{x_s}\widetilde{\sigma}_{x_s} \\ K_s^{y_s}\widetilde{\sigma}_{y_s} \\ K_s^{z_s}\widetilde{\sigma}_{z_s} \\ \vdots \\ K_S^{t_s}\widetilde{\sigma}_{t_s} \\ K_S^{x_s}\widetilde{\sigma}_{x_s} \\ K_S^{y_s}\widetilde{\sigma}_{y_s} \\ K_S^{z_s}\widetilde{\sigma}_{z_s} \end{bmatrix}, \tag{A.28}$$

where J is a constant, and the values from *Tape et al.* (2007) are

$$\widetilde{\sigma}_{t_s} \equiv \tau = 20 \text{ s} , \tag{A.29}$$

$$\widetilde{\sigma}_{x_s} \equiv \lambda = 70,000 \text{ m} , \tag{A.30}$$

$$\widetilde{\sigma}_{y_s} \equiv \lambda = 70,000 \text{ m} , \tag{A.31}$$

$$\widetilde{\sigma}_{z_s} \equiv \lambda = 70,000 \text{ m}. \tag{A.32}$$

These terms are analogous to the uncertainties in the prior model parameters. For southern California tomography, reasonable values are

$$\sigma_{t_s} = 0.5 \text{ s} , \tag{A.33}$$

$$\sigma_{x_s} = 2000.0 \text{ m} , \tag{A.34}$$

$$\sigma_{y_s} = 2000.0 \text{ m} , \tag{A.35}$$

$$\sigma_{z_s} = 2000.0 \text{ m} . \tag{A.36}$$

We now define the scaling vector $\mathbf{w}$ as

$$\mathbf{w} \equiv \begin{bmatrix} J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \end{bmatrix}. \tag{A.37}$$

With $\mathbf{W} = \mathrm{diag}(\mathbf{w})$, we multiply Equation (A.28) by $\mathbf{W}$, and the (test) model update is then

$$\widetilde{\mathbf{W}\boldsymbol{\delta}\mathbf{m}} \;=\; \mathbf{W}\widetilde{\mathbf{m}}^{0t} - \mathbf{W}\widetilde{\mathbf{m}}^{0} = -\mathbf{W}\;\widetilde{\nu}_t^0\;\widetilde{\hat{\mathbf{g}}}^0\;,$$

$$\widetilde{\mathbf{W}\boldsymbol{\delta}\mathbf{m}} \;=\; \begin{bmatrix} C_1^{0t} \\ \vdots \\ C_i^{0t} \\ \vdots \\ C_H^{0t} \\ B_1^{0t} \\ \vdots \\ B_i^{0t} \\ \vdots \\ B_H^{0t} \\ (T_s)_1^{0t} \\ (X_s)_1^{0t} \\ (Y_s)_1^{0t} \\ (Z_s)_1^{0t} \\ \vdots \\ (T_s)_s^{0t} \\ (X_s)_s^{0t} \\ (Y_s)_s^{0t} \\ (Z_s)_s^{0t} \\ \vdots \\ (T_s)_S^{0t} \\ (X_s)_S^{0t} \\ (Y_s)_S^{0t} \\ (Z_s)_S^{0t} \end{bmatrix} - \begin{bmatrix} C_1^{0} \\ \vdots \\ C_i^{0} \\ \vdots \\ C_H^{0} \\ B_1^{0} \\ \vdots \\ B_i^{0} \\ \vdots \\ B_H^{0} \\ (T_s)_1^{0} \\ (X_s)_1^{0} \\ (Y_s)_1^{0} \\ (Z_s)_1^{0} \\ \vdots \\ (T_s)_s^{0} \\ (X_s)_s^{0} \\ (Y_s)_s^{0} \\ (Z_s)_s^{0} \\ \vdots \\ (T_s)_S^{0} \\ (X_s)_S^{0} \\ (Y_s)_S^{0} \\ (Z_s)_S^{0} \end{bmatrix} = -\widetilde{\nu}_t^0 \begin{bmatrix} J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ J/\sqrt{V_1} \\ \vdots \\ J/\sqrt{V_i} \\ \vdots \\ J/\sqrt{V_H} \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \\ \vdots \\ \widetilde{\sigma}_{t_s} \\ \widetilde{\sigma}_{x_s} \\ \widetilde{\sigma}_{y_s} \\ \widetilde{\sigma}_{z_s} \end{bmatrix} \begin{bmatrix} K_1^C\,\sqrt{V_1}\,J \\ \vdots \\ K_i^C\,\sqrt{V_i}\,J \\ \vdots \\ K_H^C\,\sqrt{V_H}\,J \\ K_1^B\,\sqrt{V_1}\,J \\ \vdots \\ K_i^B\,\sqrt{V_i}\,J \\ \vdots \\ K_H^B\,\sqrt{V_H}\,J \\ K_1^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_1^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_1^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_1^{z_s}\,\widetilde{\sigma}_{z_s} \\ \vdots \\ K_s^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_s^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_s^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_s^{z_s}\,\widetilde{\sigma}_{z_s} \\ \vdots \\ K_S^{t_s}\,\widetilde{\sigma}_{t_s} \\ K_S^{x_s}\,\widetilde{\sigma}_{x_s} \\ K_S^{y_s}\,\widetilde{\sigma}_{y_s} \\ K_S^{z_s}\,\widetilde{\sigma}_{z_s} \end{bmatrix} = -\widetilde{\nu}_t^0 \begin{bmatrix} J^2\,K_1^C \\ \vdots \\ J^2\,K_i^C \\ \vdots \\ J^2\,K_H^C \\ J^2\,K_1^C \\ \vdots \\ J^2\,K_i^C \\ \vdots \\ J^2\,K_H^C \\ (\widetilde{\sigma}_{t_s})^2\,K_1^{t_s} \\ (\widetilde{\sigma}_{x_s})^2\,K_1^{x_s} \\ (\widetilde{\sigma}_{y_s})^2\,K_1^{y_s} \\ (\widetilde{\sigma}_{z_s})^2\,K_1^{z_s} \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2\,K_s^{t_s} \\ (\widetilde{\sigma}_{x_s})^2\,K_s^{x_s} \\ (\widetilde{\sigma}_{y_s})^2\,K_s^{y_s} \\ (\widetilde{\sigma}_{z_s})^2\,K_s^{z_s} \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2\,K_S^{t_s} \\ (\widetilde{\sigma}_{x_s})^2\,K_S^{x_s} \\ (\widetilde{\sigma}_{y_s})^2\,K_S^{y_s} \\ (\widetilde{\sigma}_{z_s})^2\,K_S^{z_s} \end{bmatrix}. \tag{A.38}$$

This can be rearranged as

$$\mathbf{W}\widetilde{\boldsymbol{\delta}\mathbf{m}} = \mathbf{W}\widetilde{\mathbf{m}}^{0t} - \mathbf{W}\widetilde{\mathbf{m}}^{0} = -\mathbf{W}\,\widetilde{\nu}_t^0\,\widetilde{\hat{\mathbf{g}}}^0\,, \tag{A.39}$$

$$\boldsymbol{\delta}\mathbf{m} = \mathbf{m}^{0t} - \mathbf{m}^0 = -\mathbf{W}\,\widetilde{\nu}_t^0\,\widetilde{\hat{\mathbf{g}}}^0 \tag{A.40}$$

$$\begin{aligned}
\boldsymbol{\delta}\mathbf{m} &= -\widetilde{\nu}_t^0 \begin{bmatrix} J^2/V_1 \\ \vdots \\ J^2/V_i \\ \vdots \\ J^2/V_H \\ J^2/V_1 \\ \vdots \\ J^2/V_i \\ \vdots \\ J^2/V_H \\ (\widetilde{\sigma}_{t_s})^2 \\ (\widetilde{\sigma}_{x_s})^2 \\ (\widetilde{\sigma}_{y_s})^2 \\ (\widetilde{\sigma}_{z_s})^2 \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2 \\ (\widetilde{\sigma}_{x_s})^2 \\ (\widetilde{\sigma}_{y_s})^2 \\ (\widetilde{\sigma}_{z_s})^2 \\ \vdots \\ (\widetilde{\sigma}_{t_s})^2 \\ (\widetilde{\sigma}_{x_s})^2 \\ (\widetilde{\sigma}_{y_s})^2 \\ (\widetilde{\sigma}_{z_s})^2 \end{bmatrix} \begin{bmatrix} K_1^C\,V_1 \\ \vdots \\ K_i^C\,V_i \\ \vdots \\ K_H^C\,V_H \\ K_1^B\,V_1 \\ \vdots \\ K_i^B\,V_i \\ \vdots \\ K_H^B\,V_H \\ K_1^{t_s} \\ K_1^{x_s} \\ K_1^{y_s} \\ K_1^{z_s} \\ \vdots \\ K_s^{t_s} \\ K_s^{x_s} \\ K_s^{y_s} \\ K_s^{z_s} \\ \vdots \\ K_S^{t_s} \\ K_S^{x_s} \\ K_S^{y_s} \\ K_S^{z_s} \end{bmatrix} \\
&= -\widetilde{\nu}_t^0\,\mathbf{C}'\,\hat{\mathbf{g}}^0\,.
\end{aligned} \tag{A.41}$$

Thus, the "old" method is equivalent to the new method, but using a diagonal covariance matrix defined by $\mathbf{C}'$ instead of by Equation (A.42). Furthermore, $\widetilde{\nu}_t^0$ will differ from ν_t^0, since a difference covariance matrix is present.

$$\mathbf{c} = \begin{bmatrix} \sigma_C^2 \, V \,/\, V_1 \\ \vdots \\ \sigma_C^2 \, V \,/\, V_i \\ \vdots \\ \sigma_C^2 \, V \,/\, V_H \\ \sigma_B^2 \, V \,/\, V_1 \\ \vdots \\ \sigma_B^2 \, V \,/\, V_i \\ \vdots \\ \sigma_B^2 \, V \,/\, V_H \\ \sigma_{t_s}^2 \\ \sigma_{x_s}^2 \\ \sigma_{y_s}^2 \\ \sigma_{z_s}^2 \\ \vdots \\ \sigma_{t_s}^2 \\ \sigma_{x_s}^2 \\ \sigma_{y_s}^2 \\ \sigma_{z_s}^2 \\ \vdots \\ \sigma_{t_s}^2 \\ \sigma_{x_s}^2 \\ \sigma_{y_s}^2 \\ \sigma_{z_s}^2 \end{bmatrix} . \tag{A.42}$$

Table A.1: Comparison between two generic tomography approaches, 'classical' and 'adjoint'. The reference model is expanded in terms of basis functions $B_k(\mathbf{x})$. $K_i(\mathbf{x})$ denotes the data-independent kernel for the ith source-receiver pair, while $K(\mathbf{x})$ denotes the data-dependent misfit kernel computed via adjoint methods.

	classical tomography	adjoint tomography
reference model	1D	3D
physical domain	3D	3D
Born approximation	yes	yes
forward modelling technique	e.g., ray theory, modes, or banana-doughnut kernels	fully numerical (e.g., SEM)
gradient method	$\mathbf{g} = -\mathbf{G}^T\mathbf{d}$ $G_{ik} = \int_V K_i B_k \, \mathrm{d}^3\mathbf{x}$	$g_k = \int_V K B_k \, \mathrm{d}^3\mathbf{x}$
Newton method	$\mathbf{G}^T\mathbf{G}\,\delta\mathsf{m} \approx -\mathbf{g}$	(too costly)
number of iterations	1	multiple

Table A.2: Source inversions, structure inversions, and joint inversions. T07 = *Tape et al.* (2007).

	type	pert source	pert structure	invert source	invert structure	comments	T07 figure
1	source	Y	N	Y	N	basic source inversion	16
2	structure	N	Y	N	Y	basic structure inversion	17a–c
3	joint	Y	Y	Y	Y	basic joint inversion	17d–f
4	structure	Y	Y	N	Y	map src error to structure	19a–c
5	structure	Y	N	N	Y	map src error to structure	(none)
6	source	Y	Y	Y	N	map structure error to src	19d–f
7	source	N	Y	Y	N	map structure error to src	(none)
8	joint	Y	N	Y	Y	map src error to structure	(none)
9	joint	N	Y	Y	Y	map structure error to src	(none)

Appendix B

Sensitivity kernels for different model parameterizations

Note

Here we consider three different parameterizations of seismic structure. The formulas are summarized in Table B.1 and shown for a 2D synthetic example in Figure 3.10.

B.1 General formulas

For an *isotropic earth*, the variation of the misfit function F with respect to model parameters κ-μ-ρ or α-β-ρ or c-β-ρ can be expressed in the following forms:

$$\delta F = \int_V \left[K_{\kappa(\mu\rho)} \frac{\delta\kappa}{\kappa} + K_{\mu(\kappa\rho)} \frac{\delta\mu}{\mu} + K_{\rho(\kappa\mu)} \frac{\delta\rho}{\rho} \right] d^3\mathbf{x} \,, \tag{B.1}$$

$$\delta F = \int_V \left[K_{\alpha(\beta\rho)} \frac{\delta\alpha}{\alpha} + K_{\beta(\alpha\rho)} \frac{\delta\beta}{\beta} + K_{\rho(\alpha\beta)} \frac{\delta\rho}{\rho} \right] d^3\mathbf{x} \,, \tag{B.2}$$

$$\delta F = \int_V \left[K_{c(\beta\rho)} \frac{\delta c}{c} + K_{\beta(c\rho)} \frac{\delta\beta}{\beta} + K_{\rho(c\beta)} \frac{\delta\rho}{\rho} \right] d^3\mathbf{x} \,. \tag{B.3}$$

Each kernel, $K(\mathbf{x})$, should be read as follows: $K_{\kappa(\mu\rho)}$ is "the sensitivity kernel for κ with μ and ρ fixed." In Equations (B.1)–(B.3) the kernels are defined according to model parameters defined as fractional perturbations, as used in *Tromp et al.* (2005). Below we derive expressions for the kernels.

B.1.1 Kernels for different model parameterizations

To get from Equation (B.1) to Equation (B.2), we insert Equations (B.12) and (B.13) into the integrand in Equation (B.1):

$$
\begin{aligned}
& \frac{\delta\kappa}{\kappa}\,K_{\kappa(\mu\rho)} + \frac{\delta\mu}{\mu}\,K_{\mu(\kappa\rho)} + \frac{\delta\rho}{\rho}\,K_{\rho(\kappa\mu)} \\
= \;& \left[\frac{\delta\rho}{\rho} \;+\; 2\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}\left(\alpha^2\,\frac{\delta\alpha}{\alpha} - \tfrac{4}{3}\beta^2\,\frac{\delta\beta}{\beta}\right)\right] K_{\kappa(\mu\rho)} \;+\; \left[\frac{\delta\rho}{\rho} \;+\; 2\,\frac{\delta\beta}{\beta}\right] K_{\mu(\kappa\rho)} \;+\; \frac{\delta\rho}{\rho}\,K_{\rho(\kappa\mu)} \\
= \;& \frac{\delta\rho}{\rho}\,K_{\kappa(\mu\rho)} \;+\; 2\,\frac{\delta\alpha}{\alpha}\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}\alpha^2\,K_{\kappa(\mu\rho)} \;-\; 2\,\frac{\delta\beta}{\beta}\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}\tfrac{4}{3}\beta^2\,K_{\kappa(\mu\rho)} \\
& + \frac{\delta\rho}{\rho}\,K_{\mu(\kappa\rho)} \;+\; 2\,\frac{\delta\beta}{\beta}\,K_{\mu(\kappa\rho)} \;+\; \frac{\delta\rho}{\rho}\,K_{\rho(\kappa\mu)} \\
= \;& \frac{\delta\alpha}{\alpha}\left\{2\alpha^2\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}K_{\kappa(\mu\rho)}\right\} \;+\; \frac{\delta\beta}{\beta}\left\{2\,K_{\mu(\kappa\rho)} - 2\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}\tfrac{4}{3}\beta^2\,K_{\kappa(\mu\rho)}\right\} \\
& + \frac{\delta\rho}{\rho}\left\{K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}\right\} \\
= \;& \frac{\delta\alpha}{\alpha}\left\{2\frac{\kappa + \frac{4}{3}\mu}{\rho}\frac{\rho}{\kappa}K_{\kappa(\mu\rho)}\right\} \;+\; \frac{\delta\beta}{\beta}\left\{2\,K_{\mu(\kappa\rho)} - \frac{8}{3}\beta^2\frac{\rho}{\kappa}\,K_{\kappa(\mu\rho)}\right\} \;+\; \frac{\delta\rho}{\rho}\left\{K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}\right\} \\
= \;& \frac{\delta\alpha}{\alpha}\left\{\left(2 + \frac{8}{3}\frac{\mu}{\kappa}\right)K_{\kappa(\mu\rho)}\right\} \;+\; \frac{\delta\beta}{\beta}\left\{2\,K_{\mu(\kappa\rho)} - \frac{8}{3}\frac{\mu}{\kappa}\,K_{\kappa(\mu\rho)}\right\} \;+\; \frac{\delta\rho}{\rho}\left\{K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}\right\} \\
\equiv \;& \frac{\delta\alpha}{\alpha}\,K_{\alpha(\beta\rho)} \;+\; \frac{\delta\beta}{\beta}\,K_{\beta(\alpha\rho)} \;+\; \frac{\delta\rho}{\rho}\,K_{\rho(\alpha\beta)}\;,
\end{aligned}
$$

where the misfit kernels $K_{\alpha(\beta\rho)}$, $K_{\beta(\alpha\rho)}$, and $K_{\rho(\alpha\beta)}$ represent Fréchet derivatives with respect to compressional-wave speed, shear-wave speed, and density, respectively (*Tromp et al.*, 2005, Eq. 20):

$$K_{\alpha(\beta\rho)} \;=\; \left(2 + \frac{8}{3}\frac{\mu}{\kappa}\right)K_{\kappa(\mu\rho)} \;=\; 2\,K_{\kappa(\mu\rho)} + A\,K_{\kappa(\mu\rho)}\;, \tag{B.4}$$

$$K_{\beta(\alpha\rho)} \;=\; 2\,K_{\mu(\kappa\rho)} - \frac{8}{3}\frac{\mu}{\kappa}\,K_{\kappa(\mu\rho)} \;=\; 2\,K_{\mu(\kappa\rho)} - A\,K_{\kappa(\mu\rho)}\;, \tag{B.5}$$

$$K_{\rho(\alpha\beta)} \;=\; K_{\rho(\kappa\mu)} + K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)}\;, \tag{B.6}$$

where $A \equiv 8\mu/(3\kappa)$. For the (constant) κ and μ values listed in Section 2.3.2, $A = 1.36$. Note that in explicit notation $A(\mathbf{x}) = [8\,\mu(\mathbf{x})]\,/\,[3\,\kappa(\mathbf{x})]$. Note that in deriving Equations (B.4)–(B.6), we could have stopped earlier, leaving the kernels in terms of α-β-ρ parameters rather than κ-μ-ρ parameters.

To get from Equation (B.1) to Equation (B.3), we substitute the expressions in Sec-

tion B.2 into Equation (B.1):

$$\begin{aligned}
& \frac{\delta\kappa}{\kappa} K_{\kappa(\mu\rho)} + \frac{\delta\mu}{\mu} K_{\mu(\kappa\rho)} + \frac{\delta\rho}{\rho} K_{\rho(\kappa\mu)} \\
= & \left[2\frac{\delta c}{c} + \frac{\delta\rho}{\rho}\right] K_{\kappa(\mu\rho)} \; + \; \left[2\frac{\delta\beta}{\beta} + \frac{\delta\rho}{\rho}\right] K_{\mu(\kappa\rho)} \; + \; \frac{\delta\rho}{\rho} K_{\rho(\kappa\mu)} \\
= & \frac{\delta c}{c}\left\{2K_{\kappa(\mu\rho)}\right\} \; + \; \frac{\delta\beta}{\beta}\left\{2K_{\mu(\kappa\rho)}\right\} \; + \; \frac{\delta\rho}{\rho}\left\{K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)} + K_{\rho(\kappa\mu)}\right\} \\
= & \frac{\delta c}{c}\left\{2K_{\kappa(\mu\rho)}\right\} \; + \; \frac{\delta\beta}{\beta}\left\{2K_{\mu(\kappa\rho)}\right\} \; + \; \frac{\delta\rho}{\rho}\left\{K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)} + K_{\rho(\kappa\mu)}\right\},
\end{aligned}$$

where we have defined the c-β-ρ kernels in terms of the κ-μ-ρ kernels:

$$K_{c(\beta\rho)} = 2K_{\kappa(\mu\rho)} \,, \tag{B.7}$$

$$K_{\beta(c\rho)} = 2K_{\mu(\kappa\rho)} \,, \tag{B.8}$$

$$K_{\rho(c\beta)} = K_{\kappa(\mu\rho)} + K_{\mu(\kappa\rho)} + K_{\rho(\kappa\mu)}. \tag{B.9}$$

Note that these expressions do not contained factors of μ or κ, as in Equations (B.4)–(B.6).

B.2 Model parameterizations: α-β-ρ or κ-μ-ρ or c-β-ρ

There are several different variables that can be used to describe an isotropic elastic structure for wave propagation. Here we consider three sets of model variables: α-β-ρ, κ-μ-ρ, and c-β-ρ. Going back and forth among them, or their perturbation formulas, is a matter of algebra.

B.2.1 Perturbations

We can derive the perturbations $\delta\kappa$ and $\delta\mu$ in terms of perturbations in α-β-ρ:

$$\begin{aligned}
\kappa &= \rho\, c^2 = \rho\left(\alpha^2 - \tfrac{4}{3}\beta^2\right), \\
\delta\kappa &= \delta\rho\left(\alpha^2 - \tfrac{4}{3}\beta^2\right) \; + \; \rho\left(2\alpha\;\delta\alpha - \tfrac{4}{3}\,2\beta\;\delta\beta\right) \\
&= \delta\rho\left(\alpha^2 - \tfrac{4}{3}\beta^2\right) \; + \; 2\rho\left(\alpha\;\delta\alpha - \tfrac{4}{3}\beta\;\delta\beta\right),
\end{aligned} \tag{B.10}$$

and

$$\begin{aligned} \mu &= \rho\,\beta^2\,, \\ \delta\mu &= \delta\rho\,\beta^2 \;+\; 2\rho\beta\;\delta\beta\,. \end{aligned} \tag{B.11}$$

B.2.2 Fractional Perturbations

We can derive the fractional perturbations $\delta \ln \kappa = \frac{\delta\kappa}{\kappa}$ and $\delta \ln \mu = \frac{\delta\mu}{\mu}$ in terms of fractional perturbations in α-β-ρ by working with Equations (B.10) and (B.11):

$$\begin{aligned} \left(\frac{1}{\kappa\,\rho\,\alpha\,\beta}\right)[\delta\kappa &= \delta\rho\left(\alpha^2 - \tfrac{4}{3}\beta^2\right) \;+\; 2\rho\left(\alpha\;\delta\alpha - \tfrac{4}{3}\beta\;\delta\beta\right)] \\ \left(\frac{1}{\rho\,\alpha\,\beta}\right)\frac{\delta\kappa}{\kappa} &= \frac{1}{\kappa\,\alpha\,\beta}\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)\frac{\delta\rho}{\rho} \;+\; 2\rho\,\alpha\;\frac{1}{\kappa\,\rho\,\beta}\;\frac{\delta\alpha}{\alpha} \;-\; 2\rho\,\tfrac{4}{3}\beta\;\frac{1}{\kappa\,\rho\,\alpha}\;\frac{\delta\beta}{\beta} \\ \frac{\delta\kappa}{\kappa} &= \frac{\rho}{\kappa}\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)\frac{\delta\rho}{\rho} \;+\; 2\rho\,\alpha\;\frac{\alpha}{\kappa}\;\frac{\delta\alpha}{\alpha} \;-\; 2\rho\,\tfrac{4}{3}\beta\;\frac{\beta}{\kappa}\;\frac{\delta\beta}{\beta} \\ &= \frac{\rho}{\kappa}\frac{\kappa}{\rho}\frac{\delta\rho}{\rho} \;+\; 2\,\frac{\rho}{\kappa}\,\alpha^2\,\frac{\delta\alpha}{\alpha} \;-\; 2\,\frac{\rho}{\kappa}\,\tfrac{4}{3}\,\beta^2\,\frac{\delta\beta}{\beta} \\ &= \frac{\delta\rho}{\rho} \;+\; 2\,\frac{\rho}{\kappa}\left(\alpha^2\;\frac{\delta\alpha}{\alpha} - \tfrac{4}{3}\beta^2\;\frac{\delta\beta}{\beta}\right) \\ &= \frac{\delta\rho}{\rho} \;+\; 2\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}\left(\alpha^2\;\frac{\delta\alpha}{\alpha} - \tfrac{4}{3}\beta^2\;\frac{\delta\beta}{\beta}\right) \end{aligned}$$

$$\begin{aligned} \left(\frac{1}{\rho\,\beta\,\mu}\right)[\delta\mu &= \delta\rho\,\beta^2 \;+\; 2\rho\beta\;\delta\beta] \\ \frac{1}{\rho\,\beta}\frac{\delta\mu}{\mu} &= \frac{\delta\rho}{\rho}\frac{\beta}{\mu} + \;\frac{2\beta}{\mu}\frac{\delta\beta}{\beta} \\ \frac{\delta\mu}{\mu} &= \frac{\delta\rho}{\rho}\frac{\rho\,\beta^2}{\rho\,\beta^2} \;+\; \frac{2\rho\,\beta^2}{\rho\,\beta^2}\,\frac{\delta\beta}{\beta} \\ \frac{\delta\mu}{\mu} &= \frac{\delta\rho}{\rho} \;+\; 2\,\frac{\delta\beta}{\beta}\,. \end{aligned}$$

Thus, we have

$$\frac{\delta\kappa}{\kappa} = \frac{\delta\rho}{\rho} \;+\; 2\left(\alpha^2 - \tfrac{4}{3}\beta^2\right)^{-1}\left(\alpha^2\;\frac{\delta\alpha}{\alpha} - \tfrac{4}{3}\beta^2\;\frac{\delta\beta}{\beta}\right), \tag{B.12}$$

$$\frac{\delta\mu}{\mu} = \frac{\delta\rho}{\rho} \;+\; 2\,\frac{\delta\beta}{\beta}\,. \tag{B.13}$$

Table B.1: Sensitivity kernel expressions for three different parameterizations of elastic structure. **D** is the deviatoric strain (e.g., *Liu and Tromp*, 2006, Eq. 28).

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

Appendix C

Multitaper measurements for adjoint tomography

C.1 Introduction

A multitaper measurements uses a set of optimal tapers design to extract frequency-dependent measurements of traveltime and amplitude differences. Theory and examples of multitaper measurements can be found in *Zhou et al.* (2004, 2005); *Ekström et al.* (1997); *Laske and Masters* (1996); *Thomson* (1982), and also *Percival and Walden* (1993, p. 333–347). The objective of this Appendix is to state the multitaper misfit functions (both traveltime and amplitude), and then to derive the corresponding adjoint source, ala *Tromp et al.* (2005).

C.2 Misfit functions, measurements, and adjoint sources

The conventions for the measurement, the Fourier transform, and the transfer function used for multitaper measurements are related. First we specify the measurement convention (Figure C.1), then we specify the Fourier convention, and these determine the convention for the transfer function (Section C.3.2).

C.2.1 The misfit function and measurement convention

Our misfit functions consists of a set of discrete time windows covering the seismic dataset of S earthquake sources, with N_s time windows for source index s. The total number of

measurements is

$$N = \sum_{s=1}^{S} N_s \ . \tag{C.1}$$

For traveltime and amplitude tomography we seek to minimize the objective functions

$$F_{\mathrm{P}}(\mathbf{m}) = \tfrac{1}{2} \sum_{s=1}^{S} \sum_{p=1}^{N_s} \frac{1}{F_{sp}} \int_{-\infty}^{\infty} W_{sp}(\omega) \left[\frac{\tau_{sp}^{\mathrm{obs}}(\omega) - \tau_{sp}(\omega, \mathbf{m})}{\sigma_{\mathrm{P}sp}(\omega)} \right]^2 \mathrm{d}\omega \ , \tag{C.2}$$

$$F_{\mathrm{Q}}(\mathbf{m}) = \tfrac{1}{2} \sum_{s=1}^{S} \sum_{p=1}^{N_s} \frac{1}{F_{sp}} \int_{-\infty}^{\infty} W_{sp}(\omega) \left[\frac{\ln A_{sp}^{\mathrm{obs}}(\omega) - \ln A_{sp}(\omega, \mathbf{m})}{\sigma_{\mathrm{Q}sp}(\omega)} \right]^2 \mathrm{d}\omega \ , \tag{C.3}$$

where p is the index for measurement window "pick," and P and Q are labels denoting measures of traveltime and amplitude, respectively. For example, $\tau_{sp}^{\mathrm{obs}}(\omega) - \tau_{sp}(\omega, \mathbf{m})$ represents the frequency-dependent traveltime difference between synthetics and data for one time-windowed waveform on a single seismogram for source s. The frequency-dependent uncertainty associated with the traveltime measurement is estimated by $\sigma_{\mathrm{P}sp}(\omega)$. The function $W_{sp}(\omega)$ denotes a windowing filter whose width corresponds to the frequency range over which the measurements are assumed reliable. A normalization factor G_{sp} is defined as[1]

$$G_{sp} = \int_{-\infty}^{\infty} W_{sp}(\omega) \, \mathrm{d}\omega. \tag{C.4}$$

We can incorporate the measurement uncertainty and normalization factor into the frequency filter by defining

$$W_{\mathrm{P}sp}(\omega) \equiv \frac{W_{sp}(\omega)}{G_{sp} \, \sigma_{\mathrm{P}sp}^2(\omega)} \ , \tag{C.5}$$

$$W_{\mathrm{Q}sp}(\omega) \equiv \frac{W_{sp}(\omega)}{G_{sp} \, \sigma_{\mathrm{Q}sp}^2(\omega)} \ , \tag{C.6}$$

[1] In practice we define our measurement window according to positive angular frequencies only. Thus we can write Equation (C.4) as

$$G_{sp} = \int_{-\infty}^{\infty} W_{sp}(\omega) \, \mathrm{d}\omega = 2 \int_{0}^{\infty} W_{sp}(\omega) \mathrm{d}\omega,$$

where $W_{sp}(\omega)$ is defined over positive frequencies only.

and then the misfit functions (Eqs. C.2 and C.3) become

$$F_{\mathrm{P}}(\mathbf{m}) = \tfrac{1}{2}\sum_{s=1}^{S}\sum_{p=1}^{N_s}\int_{-\infty}^{\infty} W_{\mathrm{P}sp}(\omega)\left[\tau_{sp}^{\mathrm{obs}}(\omega)-\tau_{sp}(\omega,\mathbf{m})\right]^2 \mathrm{d}\omega\ , \tag{C.7}$$

$$F_{\mathrm{Q}}(\mathbf{m}) = \tfrac{1}{2}\sum_{s=1}^{S}\sum_{p=1}^{N_s}\int_{-\infty}^{\infty} W_{\mathrm{Q}sp}(\omega)\left[\ln A_{sp}^{\mathrm{obs}}(\omega)-\ln A_{sp}(\omega,\mathbf{m})\right]^2 \mathrm{d}\omega\ . \tag{C.8}$$

The variations of Equations (C.7) and (C.8) are given by

$$\delta F_{\mathrm{P}}(\mathbf{m}) = -\tfrac{1}{2}\sum_{s=1}^{S}\sum_{p=1}^{N_s}\int_{-\infty}^{\infty} W_{\mathrm{P}sp}(\omega)\,\Delta\tau_{sp}(\omega,\mathbf{m})\,\delta\tau_{sp}(\omega,\mathbf{m})\,\mathrm{d}\omega\ , \tag{C.9}$$

$$\delta F_{\mathrm{Q}}(\mathbf{m}) = -\tfrac{1}{2}\sum_{s=1}^{S}\sum_{p=1}^{N_s}\int_{-\infty}^{\infty} W_{\mathrm{Q}sp}(\omega)\,\Delta\ln A_{sp}(\omega,\mathbf{m})\,\delta\ln A_{sp}(\omega,\mathbf{m})\,\mathrm{d}\omega\ , \tag{C.10}$$

where

$$\Delta\tau_{sp}(\omega,\mathbf{m}) \equiv \tau_{sp}^{\mathrm{obs}}(\omega)-\tau_{sp}(\omega,\mathbf{m})\ , \tag{C.11}$$

$$\Delta\ln A_{sp}(\omega,\mathbf{m}) \equiv \ln A_{sp}^{\mathrm{obs}}(\omega)-\ln A_{sp}(\omega,\mathbf{m})\ , \tag{C.12}$$

are the measured traveltime difference and amplitude difference between data and synthetics, and $\delta\tau(\omega,\mathbf{m})$ and $\delta\ln A(\omega,\mathbf{m})$ are the traveltime and amplitude perturbations with respect to changes in the model parameters. The conventions in Equations (C.11) and (C.12) are such that $\Delta\tau(\omega) > 0$ corresponds to a delay in the data, i.e., the data at frequency ω arrive late with respect to the synthetics at frequency ω (Figure C.1). Similarly, $\Delta\ln A(\omega) > 0$ corresponds to an amplification of the data with respect to the synthetics, for frequency ω. These are the same conventions used in defining the cross-correlation measurements of (*Dahlen et al.*, 2000; *Dahlen and Baig*, 2002).

Reduction for a single measurement ($N = 1$)

The notation is simpler if we consider a single event, a single receiver, a single component, and a single phase. In that case, the misfit functions are

$$F_{\mathrm{P}}(\mathbf{m}) = \tfrac{1}{2}\int_{-\infty}^{\infty} W_{\mathrm{P}}(\omega)\left[\tau^{\mathrm{obs}}(\omega)-\tau(\omega,\mathbf{m})\right]^2 \mathrm{d}\omega\ , \tag{C.13}$$

$$F_{\mathrm{Q}}(\mathbf{m}) = \tfrac{1}{2}\int_{-\infty}^{\infty} W_{\mathrm{Q}}(\omega)\left[\ln A^{\mathrm{obs}}(\omega)-\ln A(\omega,\mathbf{m})\right]^2 \mathrm{d}\omega\ , \tag{C.14}$$

where

$$W_{\rm P}(\omega) \equiv \frac{W(\omega)}{G\,\sigma_{\rm P}^2(\omega)}\,, \tag{C.15}$$

$$W_{\rm Q}(\omega) \equiv \frac{W(\omega)}{G\,\sigma_{\rm Q}^2(\omega)}\,. \tag{C.16}$$

The variations of Equations (C.13) and (C.14) are given by

$$\delta F_{\rm P}(\mathbf{m}) = -\int_{-\infty}^{\infty} W_{\rm P}(\omega)\,\Delta\tau(\omega,\mathbf{m})\,\delta\tau(\omega,\mathbf{m})\,{\rm d}\omega\,, \tag{C.17}$$

$$\delta F_{\rm Q}(\mathbf{m}) = -\int_{-\infty}^{\infty} W_{\rm Q}(\omega)\,\Delta\ln A(\omega,\mathbf{m})\,\delta\ln A(\omega,\mathbf{m})\,{\rm d}\omega\,, \tag{C.18}$$

where

$$\Delta\tau(\omega,\mathbf{m}) \equiv \tau^{\rm obs}(\omega) - \tau(\omega,\mathbf{m})\,, \tag{C.19}$$

$$\Delta\ln A(\omega,\mathbf{m}) \equiv \ln A^{\rm obs}(\omega) - \ln A(\omega,\mathbf{m})\,. \tag{C.20}$$

Reduction for frequency-independent measurements

For frequency-independent measurements (and uncertainties), we have

$$\Delta\tau(\omega) = \tau^{\rm obs} - \tau(\mathbf{m})\,, \tag{C.21}$$

$$\Delta\ln A(\omega) = \ln A^{\rm obs} - \ln A(\mathbf{m})\,, \tag{C.22}$$

$$\sigma_{\rm P}(\omega) = \sigma_{\rm P}\,, \tag{C.23}$$

$$\sigma_{\rm Q}(\omega) = \sigma_{\rm Q}\,. \tag{C.24}$$

The misfit functions (Eqs. C.13 and C.14) become

$$F_{\rm P}(\mathbf{m}) = \tfrac{1}{2}\Big[\tau^{\rm obs} - \tau(\mathbf{m})\Big]^2 \int_{-\infty}^{\infty} W_{\rm P}(\omega)\,{\rm d}\omega \;=\; \tfrac{1}{2}\left[\frac{\tau^{\rm obs} - \tau(\mathbf{m})}{\sigma_{\rm P}}\right]^2 \tag{C.25}$$

$$F_{\rm Q}(\mathbf{m}) = \tfrac{1}{2}\Big[\ln A^{\rm obs} - \ln A(\mathbf{m})\Big]^2 \int_{-\infty}^{\infty} W_{\rm Q}(\omega)\,{\rm d}\omega \;=\; \tfrac{1}{2}\left[\frac{\ln A^{\rm obs} - \ln A(\mathbf{m})}{\sigma_{\rm Q}}\right]^2 \tag{C.26}$$

which are the traveltime and amplitude cross-correlation misfit functions, $F_{\rm T}(\mathbf{m})$ and $F_{\rm A}(\mathbf{m})$, shown in *Tromp et al.* (2005), but also including measurement uncertainties.

The variations in Equations (C.25) and (C.26) are then

$$\delta F_{\mathrm{P}}(\mathbf{m}) = \left[-\frac{\Delta\tau(\mathbf{m})}{\sigma_{\mathrm{P}}^2}\right]\delta\tau(\mathbf{m}) , \tag{C.27}$$

$$\delta F_{\mathrm{Q}}(\mathbf{m}) = \left[-\frac{\Delta \ln A(\mathbf{m})}{\sigma_{\mathrm{Q}}^2}\right]\delta \ln A(\mathbf{m}) . \tag{C.28}$$

Uncertainty estimate (σ) based on cross-correlation measurements

Suppose we measure traveltime and amplitude anomalies based on the cross-correlation

$$\Gamma(\tau) = \int s(t-\tau)\, d(t)\, \mathrm{d}t, \tag{C.29}$$

where d denotes the observed seismogram and s the synthetic. Let δT denote the cross-correlation traveltime anomaly and $\delta \ln A$ the amplitude anomaly. We seek to determine σ_{T} and σ_A for these quantities. Therefore we write

$$d(t) = (1 + \delta \ln A \pm \sigma_A)\, s(t - \delta T \pm \sigma_{\mathrm{T}}) . \tag{C.30}$$

Expanding the second term to first order, we obtain

$$\begin{aligned} d(t) &\approx (1 + \delta \ln A \pm \sigma_A)\, [s(t-\delta T) \pm \sigma_{\mathrm{T}}\, \dot{s}(t-\delta T)] \\ &= (1 + \delta \ln A \pm \sigma_A)\, s(t-\delta T) \pm \sigma_{\mathrm{T}}\, (1 + \delta \ln A \pm \sigma_A)\, \dot{s}(t-\delta T) \\ &= (1 + \delta \ln A)\, s(t-\delta T) \pm \sigma_A\, s(t-\delta T) \pm \sigma_{\mathrm{T}}\, (1 + \delta \ln A)\, \dot{s}(t-\delta T) \pm \sigma_{\mathrm{T}}\, \sigma_A\, \dot{s}(t-\delta T). \end{aligned}$$

Thus, to first order in σ_{T} and σ_A, this may be written as

$$d(t) - (1 + \delta \ln A)\, s(t-\delta T) = \pm\sigma_{\mathrm{T}}(1 + \delta \ln A)\, \dot{s}(t-\delta T) \pm \sigma_A\, s(t-\delta T). \tag{C.31}$$

If we assume the errors are uncorrelated, we find that

$$\sigma_{\mathrm{T}}^2 = \frac{\int [d(t) - (1 + \delta \ln A)\, s(t-\delta T)]^2\, \mathrm{d}t}{\int [(1 + \delta \ln A)\, \dot{s}(t-\delta T)]^2\, \mathrm{d}t} , \tag{C.32}$$

$$\sigma_A^2 = \frac{\int [d(t) - (1 + \delta \ln A)\, s(t-\delta T)]^2\, \mathrm{d}t}{\int [s(t-\delta T)]^2\, \mathrm{d}t} . \tag{C.33}$$

Because σ_{T} and σ_A appear in the denominator of the adjoint source, one must specify a nonzero water-level value for each. Otherwise, for a perfect cross-correlation measurement, $\sigma_{\mathrm{T}} = \sigma_A = 0$, and the adjoint source (and therefore event kernel) will blow up. The water-level is an input parameter in `mt_measure_adj.f90`.

C.2.2 Multitaper measurements

Each windowed pulse on an individual seismogram is characterized by a (complex) transfer function from the modeled synthetics to the observed data:

$$T(\omega)\, s(\omega) = d(\omega). \tag{C.34}$$

Note that the transfer function is the same, whether the data and synthetics are in displacement, velocity, or acceleration, etc:

$$\begin{aligned} T(\omega)\, i\omega s(\omega) &= i\omega d(\omega)\ , \\ -T(\omega)\, \omega^2 s(\omega) &= -\omega^2 d(\omega)\ . \end{aligned}$$

Here, the convention $ds/dt \leftrightarrow i\omega s(\omega)$ is consistent with the Fourier convention in Section C.3.2.

Consider a single record of synthetics and data, $s(t)$ and $d(t)$, windowed in time over a particular phase and *both preprocessed in the same way*, e.g., filtered over a particular frequency window. The tapered versions are given by

$$s_j(t) = s(t)\, h_j(t)\ , \tag{C.35}$$

$$d_j(t) = d(t)\, h_j(t)\ , \tag{C.36}$$

where $h_j(t)$ is the taper.

Following *Laske and Masters* (1996), we use the multitaper method of *Thomson* (1982), which uses prolate spheroidal eigentapers (*Slepian*, 1978) for the $h_j(t)$. The transfer function $T(\omega)$ between the data and synthetics is given by (see Section C.3.1)

$$T(\omega) = \frac{\sum_j d_j(\omega)\, s_j^*(\omega)}{\sum_j s_j(\omega)\, s_j^*(\omega)}\ . \tag{C.37}$$

This function may be computed directly from the data and synthetics, and then represented

in terms of the real functions, $\Delta\tau(\omega)$ and $\Delta \ln A(\omega)$:

$$T(\omega) = \frac{\sum_j d_j(\omega)\, s_j^*(\omega)}{\sum_j s_j(\omega)\, s_j^*(\omega)} = \exp[-i\omega\, \Delta\tau(\omega)]\, [1 + \Delta \ln A(\omega)] \,, \tag{C.38}$$

$$\Delta\tau(\omega) = \frac{-1}{\omega} \tan^{-1}\left(\frac{\mathrm{Im}\,[T(\omega)]}{\mathrm{Re}\,[T(\omega)]}\right), \tag{C.39}$$

$$\Delta \ln A(\omega) = |T(\omega)| - 1 \,. \tag{C.40}$$

The sign convention in (C.38) is consistent with (C.19)–(C.20) and with the Fourier convention $\partial_t \leftrightarrow i\omega$ (Sections C.3.2 and C.3.3).

C.2.3 Multitaper adjoint sources

The units on various quantities in this section are shown in Table C.1.

Following *Tromp et al.* (2005), we must express the misfit function variations in (C.17)–(C.18) in terms of the perturbed seismograms δs. The tapered data, $d_j(t)$, can be expressed as

$$d_j = d\, h_j = (s + \delta s)\, h_j = s\, h_j + \delta s\, h_j = s_j + \delta s_j, \tag{C.41}$$

where we have defined the tapered, perturbed synthetics as

$$\delta s_j = \delta s\, h_j. \tag{C.42}$$

Substituting (C.41) into (C.37), we obtain

$$T(\omega) = \frac{\sum_j [s_j(\omega) + \delta s_j(\omega)]\; s_j^*(\omega)}{\sum_j s_j(\omega)\, s_j^*(\omega)} = 1 + \frac{\sum_j \delta s_j\, s_j^*}{\sum_j s_j\, s_j^*} \,. \tag{C.43}$$

If we write

$$T(\omega) = \exp[-i\omega\, \delta\tau(\omega)]\, [1 + \delta \ln A(\omega)] \,, \tag{C.44}$$

and make a first-order approximation for $\exp[-i\omega\, \delta\tau(\omega)]$, we have

$$T(\omega) \approx [1 - i\omega\, \delta\tau(\omega)]\, [1 + \delta \ln A(\omega)] \approx 1 - i\omega\, \delta\tau(\omega) + \delta \ln A(\omega) \,, \tag{C.45}$$

and thus, from (C.43) and (C.45),

$$T(\omega) - 1 = \frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \approx -i\omega \, \delta\tau(\omega) + \delta \ln A(\omega) \, , \tag{C.46}$$

$$\delta\tau(\omega) = \frac{-1}{\omega} \, \mathrm{Im} \left(\frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \right) , \tag{C.47}$$

$$\delta \ln A(\omega) = \mathrm{Re} \left(\frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \right) . \tag{C.48}$$

Using the identities in Appendix C.3.4, we obtain

$$\begin{aligned} \delta\tau(\omega) &= \frac{-1}{\omega} \, \mathrm{Im} \left(\frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \right) \\ &= \mathrm{Im} \left(\frac{-1}{\omega} \frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \right) \\ &= \mathrm{Re} \left(\frac{i}{\omega} \frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \right) \\ &= \mathrm{Re} \left(-\frac{i}{\omega} \frac{\sum_j s_j \, \delta s_j^*}{\sum_j s_j \, s_j^*} \right) , \end{aligned} \tag{C.49}$$

$$\begin{aligned} \delta \ln A(\omega) &= \mathrm{Re} \left(\frac{\sum_j \delta s_j \, s_j^*}{\sum_j s_j \, s_j^*} \right) \\ &= \mathrm{Re} \left(\frac{\sum_j s_j \, \delta s_j^*}{\sum_j s_j \, s_j^*} \right) . \end{aligned} \tag{C.50}$$

Inserting these expressions for the transfer function into (C.17)–(C.18), and omitting the explicit dependence on $\mathbf{m}$, we have

$$\begin{aligned} \delta F_\mathrm{P} &= -\int_{-\infty}^{\infty} W_\mathrm{P}(\omega) \, \Delta\tau(\omega) \, \mathrm{Re} \left(\frac{-i}{\omega} \frac{\sum_j s_j \, \delta s_j^*}{\sum_j s_j \, s_j^*} \right) \mathrm{d}\omega \\ &= \mathrm{Re} \left[\int_{-\infty}^{\infty} W_\mathrm{P}(\omega) \, \Delta\tau(\omega) \sum_j \left(\frac{i}{\omega} \frac{s_j}{\sum_k s_k s_k^*} \right) \delta s_j^*(\omega) \, \mathrm{d}\omega \right] \\ &= \mathrm{Re} \left[\int_{-\infty}^{\infty} W_\mathrm{P}(\omega) \, \Delta\tau(\omega) \sum_j p_j(\omega) \, \delta s_j^*(\omega) \, \mathrm{d}\omega \right] , \end{aligned} \tag{C.51}$$

$$\begin{aligned}
\delta F_{\mathrm{Q}} &= -\int_{-\infty}^{\infty} W_{\mathrm{Q}}(\omega)\,\Delta \ln A(\omega)\,\mathrm{Re}\left(\frac{\sum_j s_j\,\delta s_j^*}{\sum_j s_j\,s_j^*}\right)\mathrm{d}\omega \\
&= \mathrm{Re}\left[\int_{-\infty}^{\infty} W_{\mathrm{Q}}(\omega)\,\Delta \ln A(\omega)\sum_j\left(\frac{-s_j}{\sum_k s_k s_k^*}\right)\delta s_j^*(\omega)\,\mathrm{d}\omega\right] \\
&= \mathrm{Re}\left[\int_{-\infty}^{\infty} W_{\mathrm{Q}}(\omega)\,\Delta \ln A(\omega)\sum_j q_j(\omega)\,\delta s_j^*(\omega)\,\mathrm{d}\omega\right]\ , \qquad \text{(C.52)}
\end{aligned}$$

where

$$p_j(\omega) \equiv \frac{i}{\omega}\frac{s_j}{\sum_k s_k s_k^*} = \frac{i\omega\, s_j}{\sum_k (i\omega s_k)(-i\omega s_k^*)} = \frac{i\omega\, s_j}{\sum_k (i\omega s_k)(i\omega s_k)^*}\ , \qquad \text{(C.53)}$$

$$q_j(\omega) \equiv \frac{-s_j}{\sum_k s_k s_k^*} = i\omega\, p_j(\omega), \qquad \text{(C.54)}$$

where in (C.53) we have used the property (Eq. C.87) $-iz^* = (iz)^*$. Note that p_j and q_j are based on the (tapered) synthetics alone, and that, based on our Fourier convention in Section C.3.2,

$$q_j(t) = \dot{p}_j(t). \qquad \text{(C.55)}$$

Furthermore, note that the time-domain terms in (C.53) are all $\dot{s}_j(t)$, the derivative of the tapered synthetics.

We now use Plancherel's theorem (Section C.3.4), one version of which is

$$\int_{-\infty}^{\infty} f(\omega)\,g^*(\omega)\,\mathrm{d}\omega = 2\pi\int_{-\infty}^{\infty} f(t)\,g^*(t)\,\mathrm{d}t, \qquad \text{(C.56)}$$

to convert (C.51)–(C.52) into the time domain:

$$\begin{aligned}
\delta F_{\mathrm{P}} &= \mathrm{Re}\left[\int_{-\infty}^{\infty} W_{\mathrm{P}}(\omega)\,\Delta\tau(\omega)\sum_j p_j(\omega)\,\delta s_j^*(\omega)\,\mathrm{d}\omega\right] \\
&= \mathrm{Re}\left[\sum_j \int_{-\infty}^{\infty} \left[W_{\mathrm{P}}(\omega)\,\Delta\tau(\omega)\,p_j(\omega)\right]\delta s_j^*(\omega)\,\mathrm{d}\omega\right] \\
&= \sum_j \int_{-\infty}^{\infty} 2\pi\left[W_{\mathrm{P}}(t) * \Delta\tau(t) * p_j(t)\right]\delta s_j(t)\,\mathrm{d}t \\
&= \sum_j \int_{-\infty}^{\infty} 2\pi\left[W_{\mathrm{P}}(t) * \Delta\tau(t) * p_j(t)\right] h_j(t)\,\delta s(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \left\{2\pi\sum_j h_j(t)\left[W_{\mathrm{P}}(t) * \Delta\tau(t) * p_j(t)\right]\right\}\delta s(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} f_{\mathrm{P}}^{\dagger}(t)\;\delta s(t)\,\mathrm{d}t\;, \qquad \text{(C.57)}
\end{aligned}$$

$$\begin{aligned}
\delta F_{\mathrm{Q}} &= \mathrm{Re}\left[\int_{-\infty}^{\infty} W_{\mathrm{Q}}(\omega)\,\Delta\ln A(\omega)\sum_j q_j(\omega)\,\delta s_j^*(\omega)\,\mathrm{d}\omega\right] \\
&= \mathrm{Re}\left[\sum_j \int_{-\infty}^{\infty} \left[W_{\mathrm{Q}}(\omega)\,\Delta\ln A(\omega)\,q_j(\omega)\right]\delta s_j^*(\omega)\,\mathrm{d}\omega\right] \\
&= \sum_j \int_{-\infty}^{\infty} 2\pi\left[W_{\mathrm{Q}}(t) * \Delta\ln A(t) * q_j(t)\right]\delta s_j(t)\,\mathrm{d}t \\
&= \sum_j \int_{-\infty}^{\infty} 2\pi\left[W_{\mathrm{Q}}(t) * \Delta\ln A(t) * q_j(t)\right] h_j(t)\,\delta s(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \left\{2\pi\sum_j h_j(t)\left[W_{\mathrm{Q}}(t) * \Delta\ln A(t) * q_j(t)\right]\right\}\delta s(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{\infty} f_{\mathrm{Q}}^{\dagger}(t)\;\delta s(t)\,\mathrm{d}t\;, \qquad \text{(C.58)}
\end{aligned}$$

where $\delta s_j(t) = h_j(t)\,\delta s(t)$ is the tapered, perturbed time series, $\Delta\tau(t)$ and $\Delta\ln A(t)$ are the time domain versions of (C.39)–(C.40), and we have defined our adjoint sources for

multitaper traveltime measurements (P) and multitaper amplitude measurements (Q) as[2]

$$f^{\dagger}_{\mathrm{P}}(t) \equiv \sum_j h_j(t)\,[2\pi\, W_{\mathrm{P}}(t) * \Delta\tau(t) * p_j(t)] \ , \tag{C.59}$$

$$f^{\dagger}_{\mathrm{Q}}(t) \equiv \sum_j h_j(t)\,[2\pi\, W_{\mathrm{Q}}(t) * \Delta \ln A(t) * q_j(t)] \ . \tag{C.60}$$

The frequency domain versions of (C.59)–(C.60) are

$$f^{\dagger}_{\mathrm{P}}(\omega) = \sum_j h_j(\omega) * [2\pi\, W_{\mathrm{P}}(\omega)\,\Delta\tau(\omega)\,p_j(\omega)] \ , \tag{C.61}$$

$$f^{\dagger}_{\mathrm{Q}}(\omega) = \sum_j h_j(\omega) * [2\pi\, W_{\mathrm{Q}}(\omega)\,\Delta \ln A(\omega)\,q_j(\omega)] \ . \tag{C.62}$$

We also define the following functions:

$$P_j(t) \equiv 2\pi\, W_{\mathrm{P}}(t) * \Delta\tau(t) * p_j(t) \ , \tag{C.63}$$

$$Q_j(t) \equiv 2\pi\, W_{\mathrm{Q}}(t) * \Delta \ln A(t) * q_j(t) \ , \tag{C.64}$$

$$P_j(\omega) \equiv 2\pi\, W_{\mathrm{P}}(\omega)\,\Delta\tau(\omega)\,p_j(\omega) \ , \tag{C.65}$$

$$Q_j(\omega) \equiv 2\pi\, W_{\mathrm{Q}}(\omega)\,\Delta \ln A(\omega)\,q_j(\omega) \ . \tag{C.66}$$

These lead to succint expressions for the adjoint sources:

$$f^{\dagger}_{\mathrm{P}}(t) \equiv \sum_j h_j(t) P_j(t) \ , \tag{C.67}$$

$$f^{\dagger}_{\mathrm{Q}}(t) \equiv \sum_j h_j(t) Q_j(t) \ , \tag{C.68}$$

$$f^{\dagger}_{\mathrm{P}}(\omega) = \sum_j h_j(\omega) * P_j(\omega) \ , \tag{C.69}$$

$$f^{\dagger}_{\mathrm{Q}}(\omega) = \sum_j h_j(\omega) * Q_j(\omega) \ . \tag{C.70}$$

[2]Note that we have not written the time-dependence using the time-rerversal convention, i.e., $T - t$.

Table C.1: Units for adjoint quantities for multitaper measurements. The misfit functions $F_{\mathrm{P}}(\mathbf{m})$ and $F_{\mathrm{Q}}(\mathbf{m})$ (and $F_{\mathrm{T}}(\mathbf{m})$ and $F_{\mathrm{A}}(\mathbf{m})$) are unitless if we take into account the units for the σ terms. In practice, the adjoint sources have a m^{-3} or m^{-2} quantity as well, due to the 3D or 2D volume for the delta function, $\delta(\mathbf{x})$, that is applied at the source location. Bottom two rows are for adjoint sources based on cross-correlation measurements.

frequency domain $[\widetilde{h}(\omega)] = [h(t)]$ s		time domain $[h(t)] = [\widetilde{h}(\omega)]$ s^{-1}	
$s(\omega)$, $d(\omega)$, $\delta s(\omega)$	m s	$s(t)$, $d(t)$, $\delta s(t)$	m
$\dot{s}(\omega)$	m	$\dot{s}(t)$	m s^{-1}
$W_{\mathrm{P}}(\omega)$	s^{-1}	$W_{\mathrm{P}}(t)$	s^{-2}
$W_{\mathrm{Q}}(\omega)$	s	$W_{\mathrm{Q}}(t)$	none
$\Delta\tau(\omega)$	s	$\Delta\tau(t)$	none
$\Delta \ln A(\omega)$	none	$\Delta \ln A(t)$	s^{-1}
$p_j(\omega)$	m^{-1}	$p_j(t)$	m^{-1} s^{-1}
$q_j(\omega)$	m^{-1} s^{-1}	$q_j(t)$	m^{-1} s^{-2}
$f^{\dagger}_{\mathrm{P}}(\omega)$	m^{-1}	$f^{\dagger}_{\mathrm{P}}(t)$	m^{-1} s^{-1}
$f^{\dagger}_{\mathrm{Q}}(\omega)$	m^{-1}	$f^{\dagger}_{\mathrm{Q}}(t)$	m^{-1} s^{-1}
$f^{\dagger}_{\mathrm{T}}(\omega)$	m^{-1}	$f^{\dagger}_{\mathrm{T}}(t)$	m^{-1} s^{-1}
$f^{\dagger}_{\mathrm{A}}(\omega)$	m^{-1}	$f^{\dagger}_{\mathrm{A}}(t)$	m^{-1} s^{-1}

C.3 Miscellaneous

C.3.1 Deriving the transfer function

The transfer function, $T(\omega) = T_r(\omega) + i\, T_i(\omega)$ between data, d, and synthetics, s, is found by minimizing the objective function

$$F\left[T(\omega)\right] \;=\; \tfrac{1}{2} \sum_j \left| d_j(\omega) - T(\omega)\, s_j(\omega) \right|^2 \;,$$

where the sum over j represents the multitapers of Section C.2.2. Expanding this into real and imaginary parts, we have

$$\begin{aligned}
F(T) &= \tfrac{1}{2}\sum_j |d_j - T s_j|^2 \\
&= \tfrac{1}{2}\sum_j (d_j - T s_j)(d_j^* - T^* s_j^*) \\
&= \tfrac{1}{2}\sum_j \left[d_j - (T_r + i\,T_i)\,s_j\right]\left[d_j^* - (T_r - i\,T_i)\,s_j^*\right] \\
&= \tfrac{1}{2}\sum_j \left[d_j\,d_j^* - (T_r - i\,T_i)\,d_j\,s_j^* - (T_r + i\,T_i)\,d_j^*\,s_j + (T_r^2 + T_i^2)\,s_j\,s_j^*\right] \\
&= \tfrac{1}{2}\sum_j \left[d_j\,d_j^* - T_r\,d_j\,s_j^* + i\,T_i\,d_j\,s_j^* - T_r\,d_j^*\,s_j - i\,T_i\,d_j^*\,s_j + T_r^2\,s_j\,s_j^* + T_i^2\,s_j\,s_j^*\right] \\
&= \tfrac{1}{2}\sum_j \left[d_j\,d_j^* + T_r\left(-d_j\,s_j^* - d_j^*\,s_j\right) + T_i\left(i\,d_j\,s_j^* - i\,d_j^*\,s_j\right) + T_r^2\,s_j\,s_j^* + T_i^2\,s_j\,s_j^*\right] .
\end{aligned}$$

The derivatives with respect to the real and imaginary parts of the transfer function are given by

$$\begin{aligned}
\frac{\partial F}{\partial T_r} &= \tfrac{1}{2}\sum_j \left[-d_j\,s_j^* - d_j^*\,s_j + 2\,T_r\,s_j\,s_j^*\right] = -\tfrac{1}{2}\sum_j \left(d_j^*\,s_j + d_j\,s_j^* +\right) + T_r\sum_j s_j\,s_j^* , \\
\frac{\partial F}{\partial T_i} &= \tfrac{1}{2}\sum_j \left[i\,d_j\,s_j^* - i\,d_j^*\,s_j + 2\,T_i\,s_j\,s_j^*\right] = -\frac{i}{2}\sum_j \left(d_j^*\,s_j - d_j\,s_j^*\right) + T_i\sum_j s_j\,s_j^* .
\end{aligned}$$

Setting each equation equal to zero and solving for T_r and T_i gives:

$$\begin{aligned}
T_r &= \frac{\frac{1}{2}\sum_j \left(d_j\,s_j^* + d_j^*\,s_j\right)}{\sum_j s_j\,s_j^*} , \\
T_i &= \frac{\frac{i}{2}\sum_j \left(d_j^*\,s_j - d_j\,s_j^*\right)}{\sum_j s_j\,s_j^*} .
\end{aligned}$$

Thus, we obtain (C.37):

$$\begin{aligned}
T &= T_r + i\,T_i = \frac{\frac{1}{2}\sum_j \left(d_j\,s_j^* + d_j^*\,s_j\right)}{\sum_j s_j\,s_j^*} - \frac{\frac{1}{2}\sum_j \left(d_j^*\,s_j - d_j\,s_j^*\right)}{\sum_j s_j\,s_j^*} \\
&= \frac{\frac{1}{2}\sum_j \left(d_j\,s_j^* + d_j^*\,s_j - d_j^*\,s_j + d_j\,s_j^*\right)}{\sum_j s_j\,s_j^*} \\
&= \frac{\sum_j d_j\,s_j^*}{\sum_j s_j\,s_j^*} .
\end{aligned}$$

C.3.2 Conventions for measurements, Fourier transform, and transfer function

We define our measurements according to the conventions in (C.19)–(C.20), such that a positive traveltime measurement, $\Delta\tau > 0$, corresponds to a delay in the data with respect to the synthetics.

We define forward and inverse Fourier transforms

$$\mathcal{F}\left[h(t)\right] \;=\; \widetilde{h}(\omega) \;=\; \int_{-\infty}^{\infty} h(t)\, e^{-i\omega t}\, \mathrm{d}t\;, \tag{C.71}$$

$$\mathcal{F}^{-1}\left[\widetilde{h}(\omega)\right] \;=\; h(t) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{h}(\omega)\, e^{i\omega t}\, \mathrm{d}\omega\;. \tag{C.72}$$

Note that this convention follows that of *Dahlen and Tromp* (1998, p. 109), and that the units conversion is

$$\left[\widetilde{h}(\omega)\right] \;=\; [h(t)]\ \mathrm{s},$$

which is reflected in Table C.1.

The Fourier transform of $\dot{h}(t)$ can be determined using integration by parts,

$$\int u\, dv = [u\, v] - \int v\, du,$$

with $dv = \dot{h}(t)\, dt$, $u = e^{-i\omega t}$, $v = h(t)$, and $du = -i\omega\, e^{-i\omega t} dt$. Thus, we can write

$$\begin{aligned}
\mathcal{F}\left[\dot{h}(t)\right] &= \int_{-\infty}^{\infty} \dot{h}(t)\, e^{-i\omega t}\, \mathrm{d}t \\
&= \left[h(t)\, e^{-i\omega t}\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} h(t)\, (-i\omega)\, e^{-i\omega t}\, \mathrm{d}t \\
&= i\omega \int_{-\infty}^{\infty} h(t)\, e^{-i\omega t}\, \mathrm{d}t \\
&= i\omega\, \mathcal{F}\left[h(t)\right]\;.
\end{aligned} \tag{C.73}$$

This process can be iterated for the nth derivative to yield

$$\mathcal{F}\left[h^{(n)}(t)\right] \;=\; (i\omega)^n\, \mathcal{F}\left[h(t)\right]\;. \tag{C.74}$$

Note that (C.73)–(C.74) will depend on the Fourier convention.

The transfer function is defined according to $T(\omega)\, s(\omega) = d(\omega)$ (Eq. C.34). The convention for the measurements and the Fourier convention imply that the transfer function is to be written as

$$T(\omega) = e^{-i\omega\Delta\tau},$$

where we have ignored the amplitude measurement and have assumed that $\Delta\tau$ is constant over all ω. Then, the data in the frequency and time domain are given by

$$d(\omega) \;=\; s(\omega)\, e^{-i\omega\Delta\tau}\,, \tag{C.75}$$

$$d(t) \;=\; \mathcal{F}^{-1}\left[d(\omega)\right] \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} s(\omega)\, e^{-i\omega\Delta\tau}\, e^{i\omega t}\,\mathrm{d}\omega \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} s(\omega)\, e^{i\omega(t-\Delta\tau)}\,\mathrm{d}\omega \;=\; s(t-\Delta\tau). \tag{C.76}$$

For example, for data arriving early with $\Delta\tau = -3$, we have $d(t) = s(t+3)$, indicating that the data are advanced by 3 seconds with respect to the synthetics.

C.3.3 Implementation of conventions for measurement, Fourier, and transfer function

In practice, the original data are shifted by the cross-correlation measurement, ΔT, prior to making the multitaper measurement. In other words, $d(\omega) = d_0(\omega)\, e^{i\omega\,\Delta T}$, where d_0 are the original, unshifted data. This convention is checked as follows:

$$d(t) \;=\; \mathcal{F}^{-1}\left[d(\omega)\right] \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} d_0(\omega)\, e^{i\omega\Delta T}\, e^{i\omega t}\,\mathrm{d}\omega \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty} d_0(\omega)\, e^{i\omega\,(t+\Delta T)}\,\mathrm{d}\omega \;=\; d_0(t+\Delta T)\,, \tag{C.77}$$

that is, $d_0(t) = d(t-\Delta T)$. For example, for (original) data arriving early with respect to the synthetics, with $\Delta T = -3$, then we have $d_0(t) = d(t+3)$, indicating that the original data are advanced by 3 seconds with respect to the shifted data.

In the measurement code, the transfer function we compute is $T'(\omega) = e^{-i\omega\Delta\tau'}$, with $T'(\omega)\, s(\omega) = d(\omega)$. (Again, we ignore amplitudes for clarity.) Thus, the shifted data, $d(t)$, are aligned in phase with the synthetics by a uniform shift with magnitude $|\Delta T|$. Then we

can express the *unshifted* data as

$$\begin{aligned} d_0(\omega) &= e^{-i\omega\,\Delta T}\,d(\omega) \\ &= e^{-i\omega\,\Delta T}\,T'(\omega)\,s(\omega) \;=\; e^{-i\omega\,\Delta T}\,e^{-i\omega\Delta\tau'}\,s(\omega) \;=\; e^{-i\omega\,(\Delta T+\Delta\tau')}\,s(\omega) \;=\; T(\omega)\,s(\omega)\;, \end{aligned} \tag{C.78}$$

where $T(\omega)$ represents the transfer function from the synthetics to the unshifted data, and $\Delta\tau'$ represents the frequency-dependent perturbations from the frequency-independent cross-correlation measurement ΔT, such that

$$\Delta\tau(\omega) = \Delta\tau'(\omega) + \Delta T. \tag{C.79}$$

Thus, we can write the phases as

$$-\omega\,\Delta\tau'(\omega) = \tan^{-1}\left(\frac{\mathrm{Im}\,[T'(\omega)]}{\mathrm{Re}\,[T'(\omega)]}\right), \tag{C.80}$$

$$-\omega\,\Delta\tau(\omega) = -\omega\left(\Delta\tau'(\omega)+\Delta T\right) = \tan^{-1}\left(\frac{\mathrm{Im}\,[T'(\omega)]}{\mathrm{Re}\,[T'(\omega)]}\right) - \omega\,\Delta T\;, \tag{C.81}$$

and the corresponding traveltimes as

$$\Delta\tau'(\omega) = \frac{-1}{\omega}\tan^{-1}\left(\frac{\mathrm{Im}\,[T'(\omega)]}{\mathrm{Re}\,[T'(\omega)]}\right), \tag{C.82}$$

$$\Delta\tau(\omega) = \frac{-1}{\omega}\tan^{-1}\left(\frac{\mathrm{Im}\,[T'(\omega)]}{\mathrm{Re}\,[T'(\omega)]}\right) + \Delta T\;. \tag{C.83}$$

Compare (C.83) with (C.39). The use of $T'(\omega)$ instead of $T(\omega)$ gives rise to the inclusion of ΔT.

C.3.4 Plancherel's theorem

Parseval's theorem applied to Fourier series is known as *Rayleigh's theorem* or *Plancherel's theorem*. The following derivation of Plancherel's theorem is adapted from the Mathworld website.

The exact representation of the theorem depends on the Fourier convention. Using the

Fourier conventions in Equations (C.71) and (C.72), we have

$$\begin{aligned} f(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{f}(\omega)\, e^{i\omega t}\,\mathrm{d}\omega\,, \\ f^*(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{f}^*(\omega)\, e^{-i\omega t}\,\mathrm{d}\omega\,, \\ \widetilde{f}(\omega) &= \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t}\,\mathrm{d}t\,, \\ \widetilde{f}^*(\omega) &= \int_{-\infty}^{\infty} f^*(t)\, e^{i\omega t}\,\mathrm{d}t\,. \end{aligned}$$

Consider the following derivation:

$$\begin{aligned} \int_{-\infty}^{\infty} f(t)\, g^*(t)\,\mathrm{d}t &= \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \widetilde{f}(\omega)\, e^{i\omega t}\,\mathrm{d}\omega\right] \left[\int_{-\infty}^{\infty} \widetilde{g}^*(\omega')\, e^{-i\omega' t}\,\mathrm{d}\omega'\right] \mathrm{d}t \\ &= \left(\frac{1}{2\pi}\right)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{f}(\omega)\,\widetilde{g}^*(\omega')\, e^{it(\omega-\omega')}\mathrm{d}\omega\,\mathrm{d}\omega'\,\mathrm{d}t \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \widetilde{f}(\omega)\,\widetilde{g}^*(\omega') \left[\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{it(\omega-\omega')}\mathrm{d}t\right] \mathrm{d}\omega\,\mathrm{d}\omega' \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(\omega-\omega')\,\widetilde{f}(\omega)\,\widetilde{g}^*(\omega')\,\mathrm{d}\omega\,\mathrm{d}\omega' \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \widetilde{f}(\omega)\,\widetilde{g}^*(\omega)\,\mathrm{d}\omega. \end{aligned}$$

If $g(t) = f(t)$ (and thus $g^*(t) = f^*(t)$), then we obtain Plancherel's Theorem,

$$\int_{-\infty}^{\infty} |f(t)|^2\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\infty}^{\infty} \left|\widetilde{f}(\omega)\right|^2\,\mathrm{d}\omega, \tag{C.84}$$

which states that the integral of the squared modulus of a function is equal to the integral of the squared modulus of its spectrum.

Miscellaneous formulas

Considering two complex numbers,

$$\begin{aligned} z &= a + bi \,, \\ w &= c + di \,, \end{aligned}$$

we can derive the following expressions:

$$\begin{aligned} \mathrm{Re}\,[iz] &= \mathrm{Re}\,[i(a+bi)] = \mathrm{Re}\,[ai-b] = -b = \mathrm{Im}[-ai-b] = \mathrm{Im}[-z] \,, && \text{(C.85)} \\ \mathrm{Re}\left[i^{-1}z\right] &= \mathrm{Re}\,[-iz] = \mathrm{Im}[z] \,, && \text{(C.86)} \\ -iz^* &= -i(a-bi) = -b+ai = (-b+ai)^* = [i(a+bi)]^* = (iz)^* \,, && \text{(C.87)} \\ \mathrm{Re}\,[zw^*] &= \mathrm{Re}\,[(a+bi)(c-di)] \\ &= \mathrm{Re}\,[ac-adi+bci+bd] \\ &= ac+bd \\ &= \mathrm{Re}\,[ac+adi-bci+bd] \\ &= \mathrm{Re}\,[(a-bi)(c+di)] \\ &= \mathrm{Re}\,[z^*w] \;. && \text{(C.88)} \end{aligned}$$

Using Equations (C.87) and (C.88), we obtain

$$\mathrm{Re}\,[izw^*] = \mathrm{Re}\,[(iz)^*w] = \mathrm{Re}\,[-iz^*w] \;. \qquad \text{(C.89)}$$

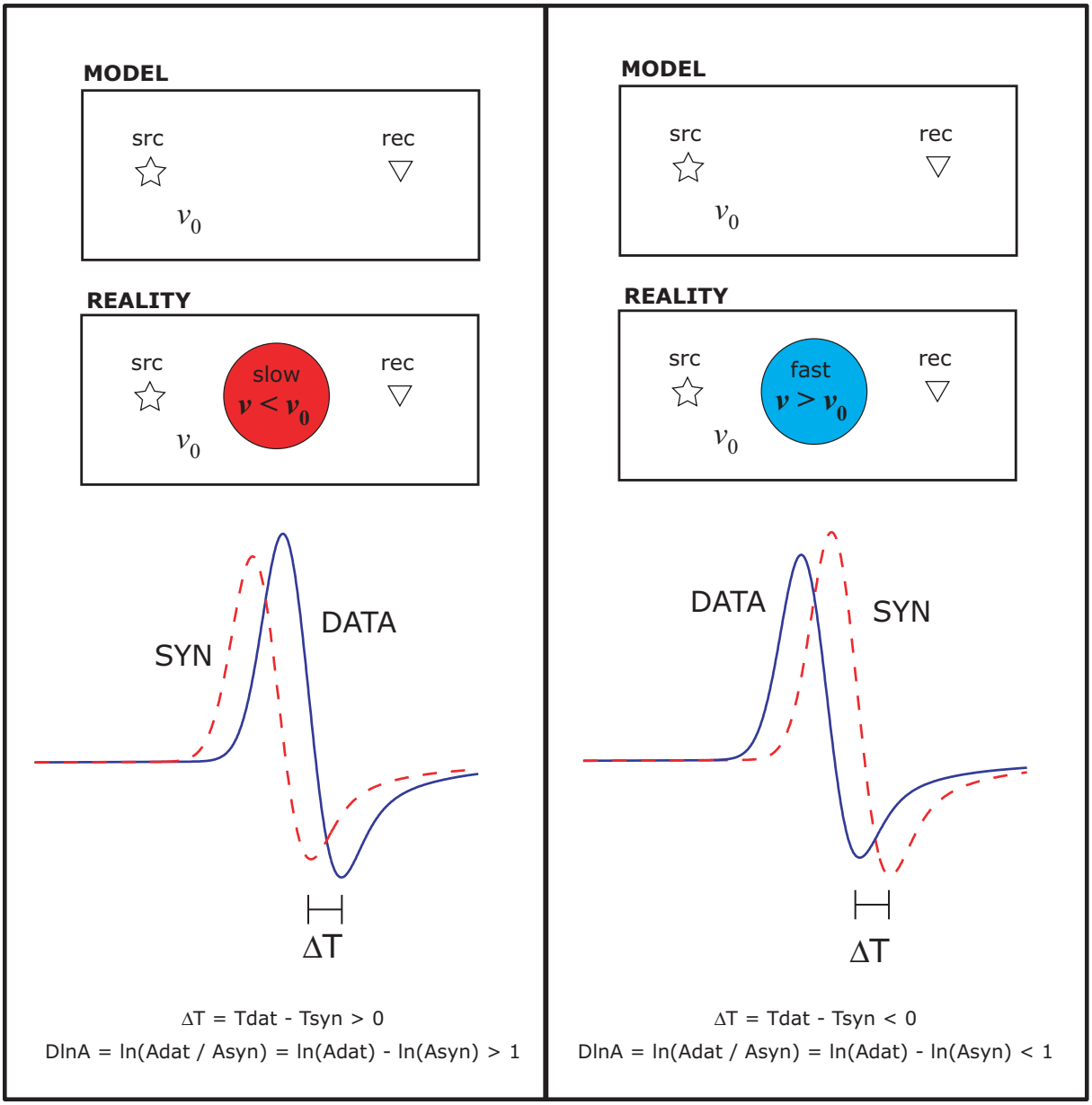


Figure C.1: The measurement convention for traveltime differences, ΔT, and amplitude differences $\Delta \ln A$. See Section C.2.1.

Appendix D

Earthquake source parameters for southern California tomography

Note

This appendix is devoted to explaining the details of the 294 earthquakes listed in Figures D.1–D.37. Details regarding the selection of sources are discussed in Chapter 6.

Description of Figures D.1–D.37

We have two objectives in assembling focal mechanisms from previously studied earthquakes:

1. to obtain the best possible source parameters for our SEM-based source inversions or for our tomographic inversion;

2. to test the differences among various source-inversion approaches (Section 6.2.2) by comparing 3D synthetics directly with data.

Figures D.1–D.37 is a compilation of focal mechanisms for 294 earthquakes in southern California. The earthquakes are sorted by region, and, within each region, by origin time. Many of the earthquakes occur in aftershock sequences, so this ordering allows one to readily identify differences within the same sequence.

Each of the 294 earthquakes is classified into one of the six groups in Table D.1. The majority of the "extra" and "outside" sources are primarily from a set of 159 well-studied

Table D.1: Classification groups for all 294 earthquakes in Figures D.1–D.37.

label	number	description
TOMO	143	used in at least one iteration of the tomographic inversion
EXTRA	91	not used in the tomographic inversion
LOW SNR	28	low signal-to-noise ratio
OUTSIDE	16	outside simulation region
REJECTED	9	rejected
BAD SOURCE	7	wrong source mechanism

earthquakes of *Tan* (2006, Appendix A).

The "low SNR" earthquakes are primarily events that generate synthetic seismograms that have measurement windows (*Maggi et al.*, 2009) at fewer than 10 stations. In regions that are very challenging to determine source parameters—for example, Continental Borderlands and Salton trough—I have moved some low signal-to-noise earthquakes in the "extra" group, if the comparison with data suggested that the focal mechanism was "in the ballpark". I have left the "low SNR" earthquakes in the compilation for completeness, but most of these earthquakes are not quality events and are probably not worth investigating any further.

The "rejected" earthquakes are primarily events that occurred close in space, time, and magnitude to other events. These were typically determined by analyzing near-source records and identifying coherent seismic energy later in the same seismograms. Each secondary event was confirmed using the Southern California Earthquake Data Center catalog. I also rejected earthquakes that were clearly too large for a point-source approximation, given our period range of interest (2–30 s), such as the 2004.09.28 M_{w} 6.2 Parkfield earthquake (14094992).

The "bad source" earthquakes are events that appear to have data with high enough signal-to-noise ratio, but that clearly have the wrong source parameters. It is possible that better mechanisms could be determined with additional source inversion attempts, such as *Liu et al.* (2004).

Figures D.1–D.37 contains eight columns, which are described in Table D.2.

The "CAP" mechanisms are primarily from the set of 159 in *Tan* (2006), with 20 additional mechanisms provided by Shengji Wei (Caltech, December 2008). These 20 events are: 10006857, 10148421, 11671240, 12659440, 14073800, 14077668, 14138080, 14178236,

Table D.2: Eight columns of Figures D.1–D.37

column	label	description	reference
1	CAP	cut-and-paste method	*Tan* (2006)
2	JH	P/S amplitude ratio method	*Hardebeck and Shearer* (2003)
3	SCEDC	SCEDC	*Clinton et al.* (2006)
4	mod	SCEDC-modified	*Clinton et al.* (2006)
5	SEMm00	SEM inversion using $\mathbf{m}_{00}$	*Liu et al.* (2004)
6	m12	source parameters for $\mathbf{m}_{12}$	
7	SEMm12	SEM inversion using $\mathbf{m}_{12}$	*Liu et al.* (2004)
8	m16	source parameters for $\mathbf{m}_{16}$	

14179288, 14179292, 14186612, 14239184, 3320736, 9111353, 9112735, 9117942, 9154092, 9967901, 14383980, and 14408052.

The "mod" mechanisms only differ from "SCEDC" in cases where Egill Hauksson tried the inversion of *Clinton et al.* (2006) using different stations. These were cases where I identified poor agreement between data and 3D synthetics generated using the SCEDC mechanism. These events include: 10230869, 13970876, 13966672, 14072464, 9944301, 14179288, 14179292, 14263712, 9753485, 9755013, and 14178236.

The "SEMm00" inversions required an initial-guess focal mechanisms, which was taken to be the SCEDC mechanism in all cases except for 14263712, which used the modified SCEDC mechanism ("mod").

For the labels at the right, the numbers N1, N2, and N3 in "m16 : N1 (N2, N3, N4)" are given by:

N1	total number of stations with measurements for model $\mathbf{m}_{16}$
N2	number of stations with measurements for periods 6–30 s for $\mathbf{m}_{16}$
N3	number of stations with measurements for periods 3–30 s for $\mathbf{m}_{16}$
N4	number of stations with measurements for periods 2–30 s for $\mathbf{m}_{16}$

The label for each earthquake at the left of each row contains the event ID with a tag denoting two items: (1) the dataset providing the hypocenter and origin time; (2) the dataset providing the focal mechanism. The datasets for the hypocenters and origin times are:

label	reference
Salton	*Lohman and McGuire* (2007)
Parkfield	*Thurber et al.* (2006)
SanSimeon	*McLaren et al.* (2008), courtesy of Jeanne Hardebeck
Lin	*Lin et al.* (2007a), plus 18 from Guoqing Lin
NCEDC	NCEDC catalog
SCEDC-Loc	SCEDC local catalog
SCEDC-Reg	SCEDC regional catalog

The data sources for the focal mechanisms are:

label	reference
SEMm00	SEM inversion with model $\mathbf{m}_{00}$ (unpublished)
CAP	*Tan* (2006), plus 20 by Shengji Wei
JH	*Hardebeck and Shearer* (2003)
SCEDCmod	SCEDC with Hauksson modifications, if available
SEMm12	SEM inversion with model $\mathbf{m}_{00}$ (unpublished)
CHT	Carl's replacement after synthetic tests (unpublished)

For example, the label `9718013_SEMm12_Lin` denotes event 9718013, focal mechanism and modified depth from the SEM inversion using $\mathbf{m}_{12}$, and origin time, epicenter, and initial depth from *Lin et al.* (2007a).

The eight "CHT" events (10061489, 9119414, 14139160, 9154233, 9722669, 9817605, 13966672, 9660449) are events that initially generated poor fits to the data, but which had other proximal events (in time, space, and magnitude) with *different* mechanisms that produced much better fits. In these cases, I assigned the "other" event's focal mechanism to the CHT event, then generated 3D synthetics to verify that the new mechanism was better. One dramatic example of improvement is for 9817605, an event in the Salton trough. Mechanisms from CAP and SCEDC are similar and, based on the 3D synthetics fits to data, are clearly not correct. I assigned the focal mechanism of 9722633, a well-fit earthquake that occurred in the same region less than one year earlier. Using the new mechanism for 9817605, I produced 3D synthetics with measurements at 112 stations, indicating a very well-recorded earthquake. None of the CHT events were used in the tomographic inversion, and they await SEM-based inversion (*Liu et al.*, 2004) using the final model ($\mathbf{m}_{16}$).

Additional labels are associated with the focal mechanisms in each column:

- For the CAP focal mechanisms, we list the depth and also the magnitude.
- For the JH focal mechanism, "A: P29, R4" would denote quality A, 29 P-wave polarities used, and 4 S/P amplitude ratios used.
- For the SCEDC focal mechanisms, the variance reduction is listed. The variance reduction determines the "quality factor" as follows:

VR interval	quality factor
$VR > 60$	A : "Mw, MT good enough for distribution"
$40 < VR < 60$	B : "Mw only good enough for distribution"
$VR < 40$	C : "Solution needs review before distribution"

- For the SEMm00 focal mechanisms, we list the percent non-double couple, which can range from 0 to 100.
- For the SEMm12 focal mechanisms, we list the depth and also the magnitude.

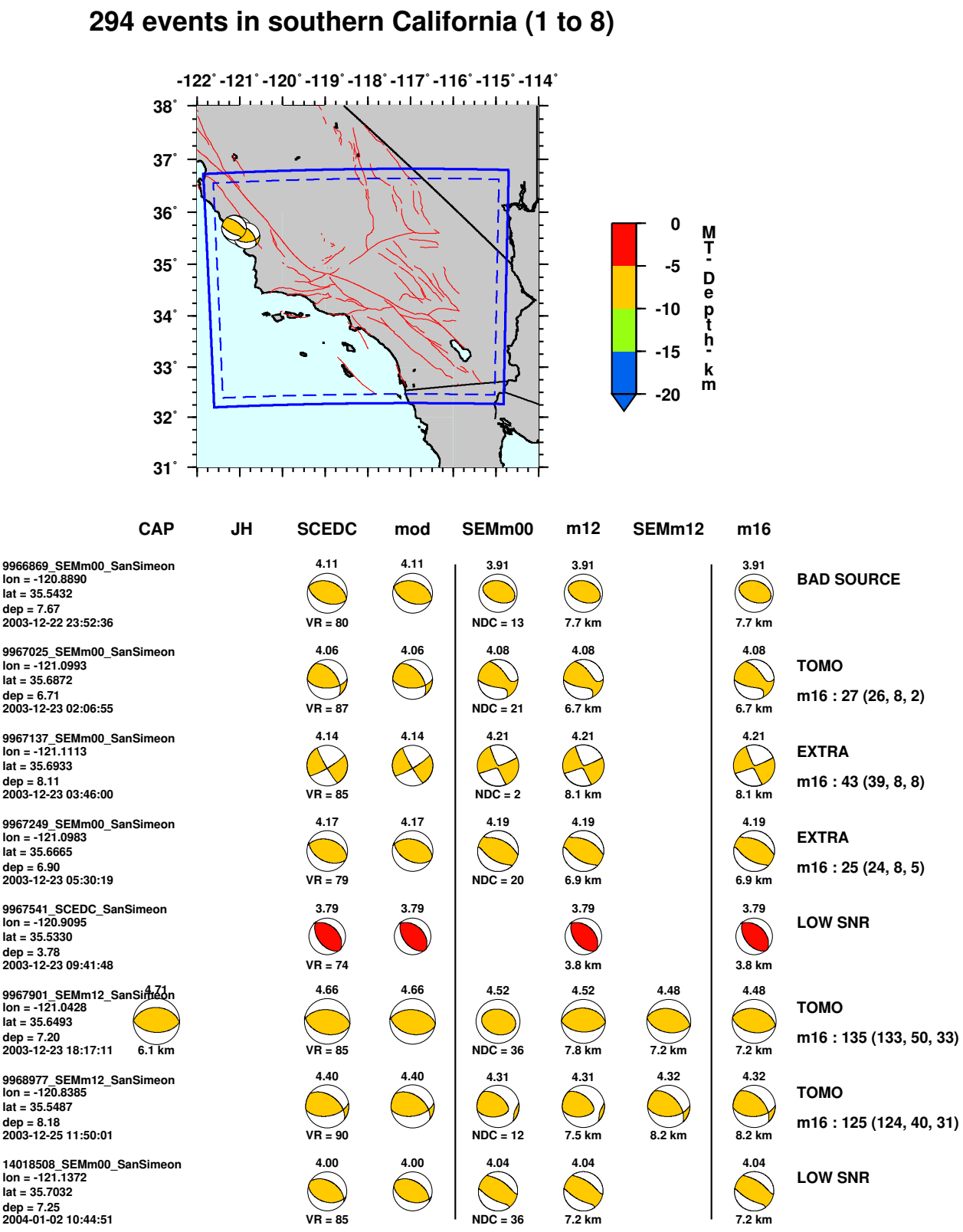


Figure D.1: Source mechanisms considered in the southern California tomography study (1 through 8 out of 294).

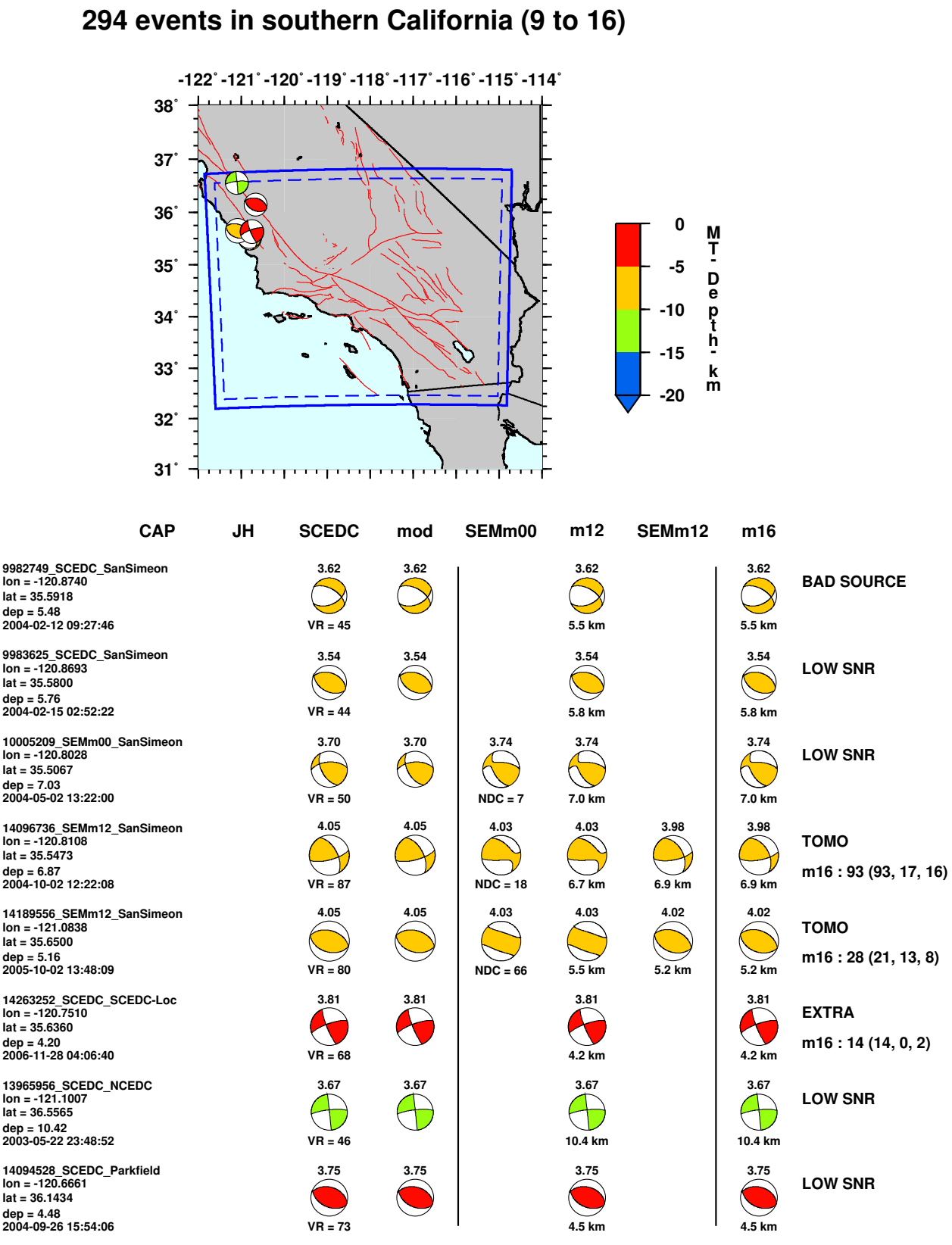


Figure D.2: Source mechanisms considered in the southern California tomography study (9 through 16 out of 294).

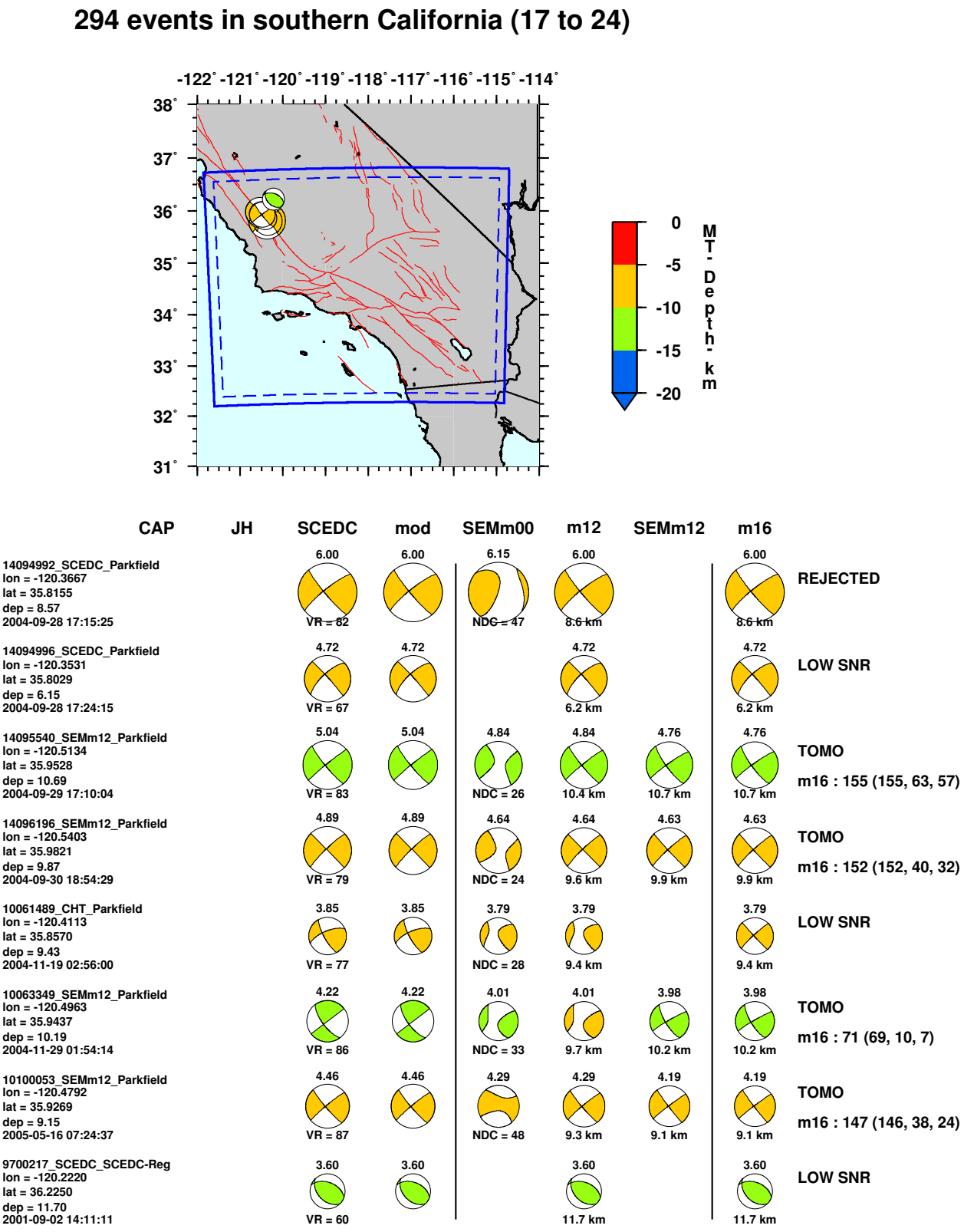


Figure D.3: Source mechanisms considered in the southern California tomography study (17 through 24 out of 294).

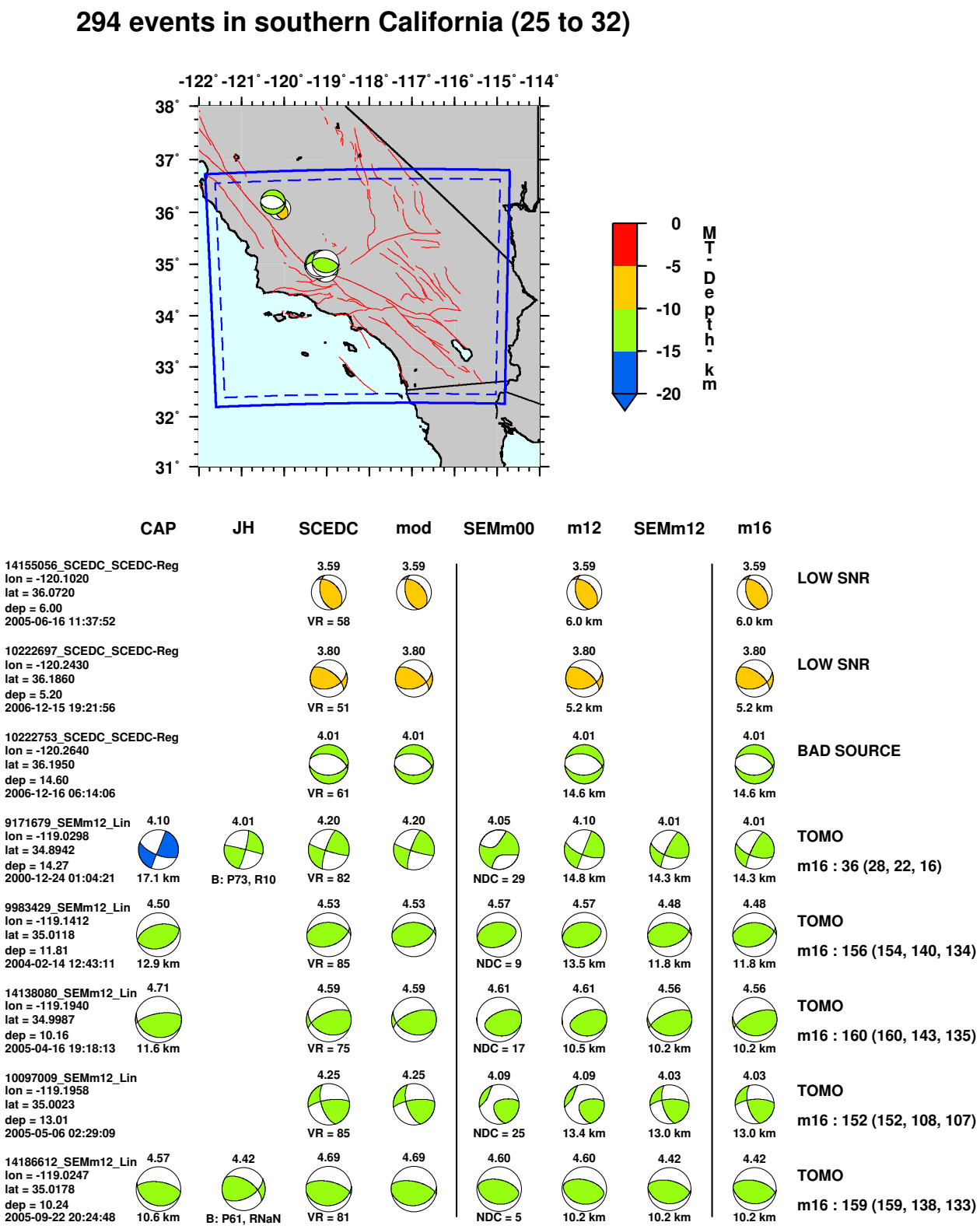


Figure D.4: Source mechanisms considered in the southern California tomography study (25 through 32 out of 294).

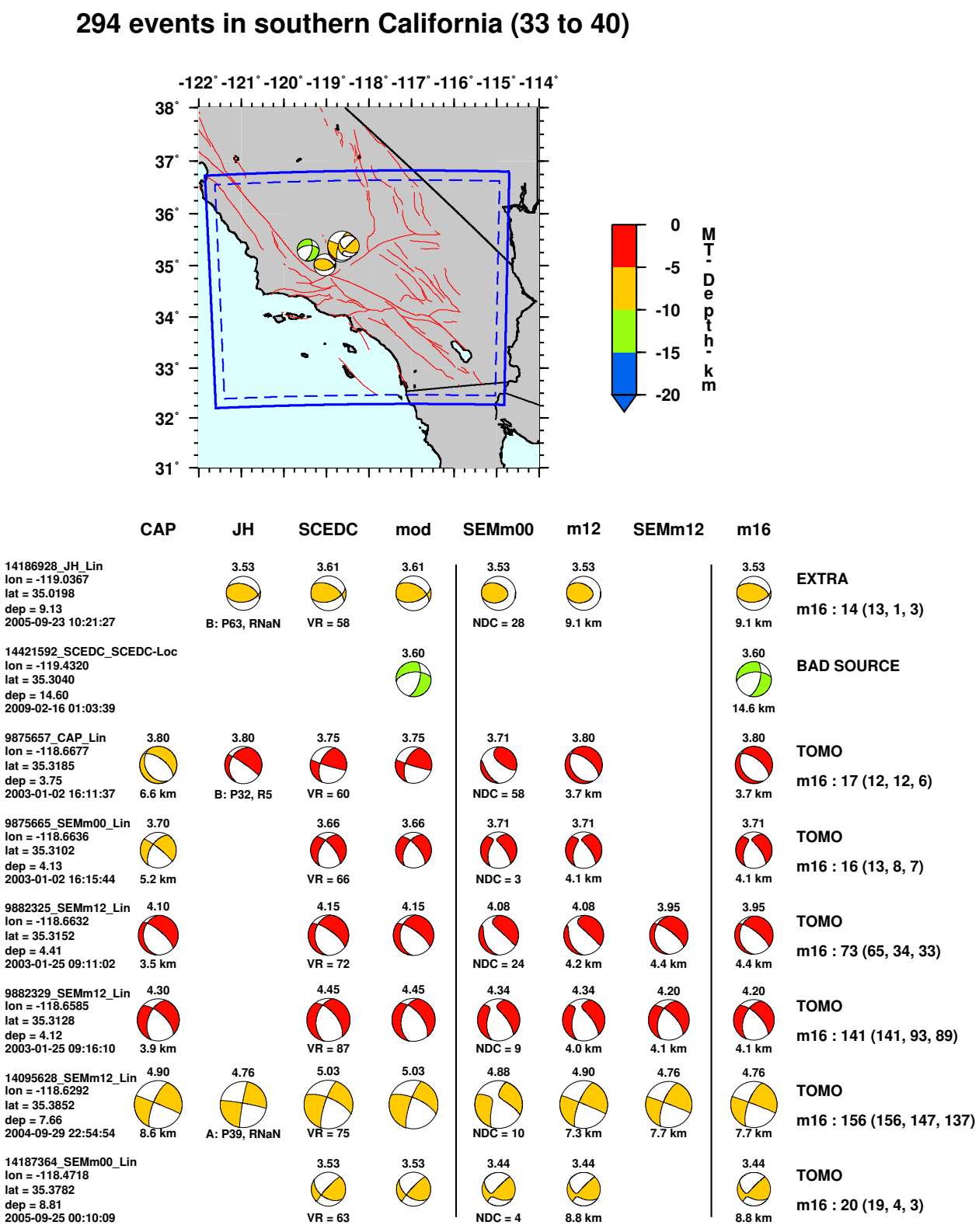


Figure D.5: Source mechanisms considered in the southern California tomography study (33 through 40 out of 294).

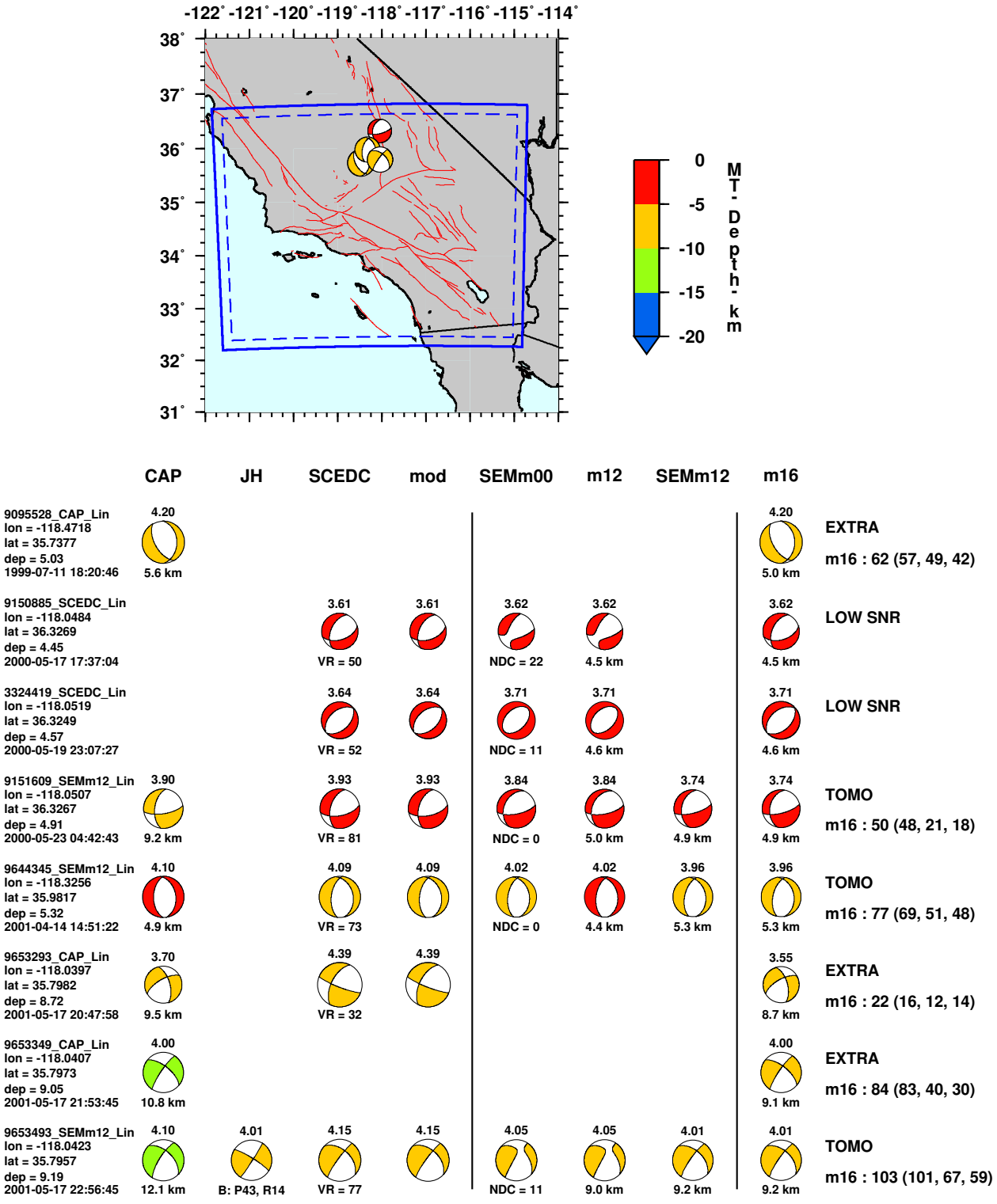


Figure D.6: Source mechanisms considered in the southern California tomography study (41 through 48 out of 294).

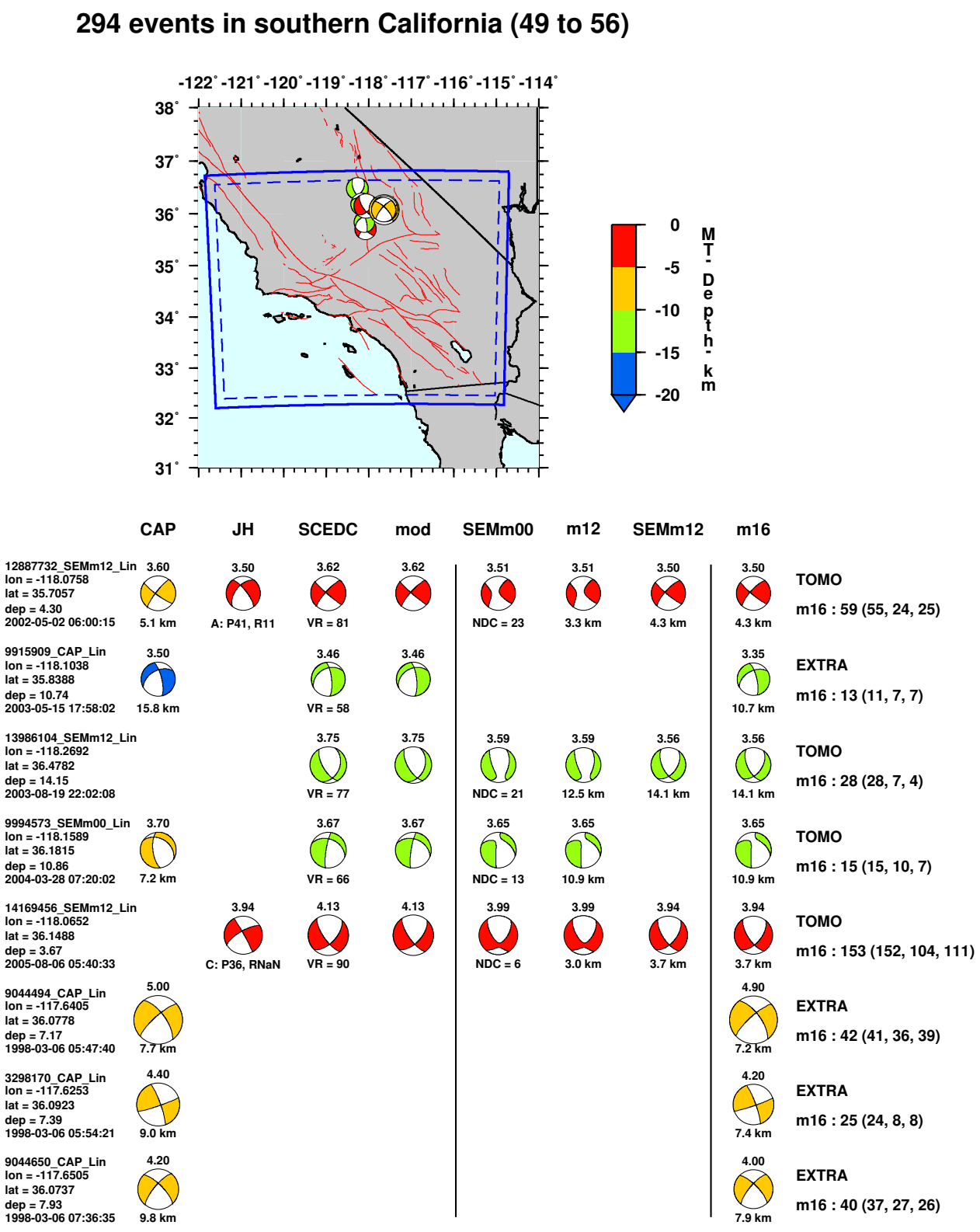


Figure D.7: Source mechanisms considered in the southern California tomography study (49 through 56 out of 294).

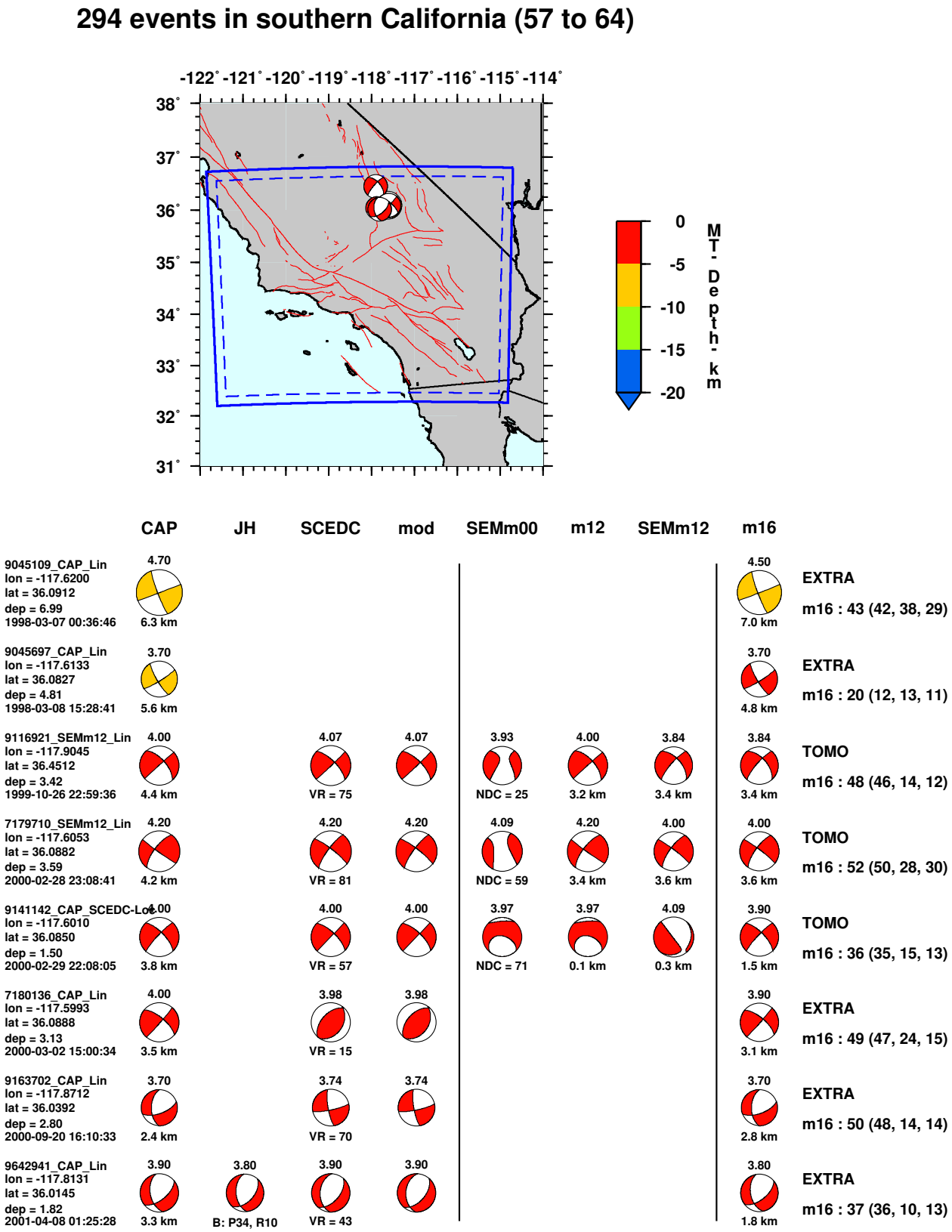


Figure D.8: Source mechanisms considered in the southern California tomography study (57 through 64 out of 294).

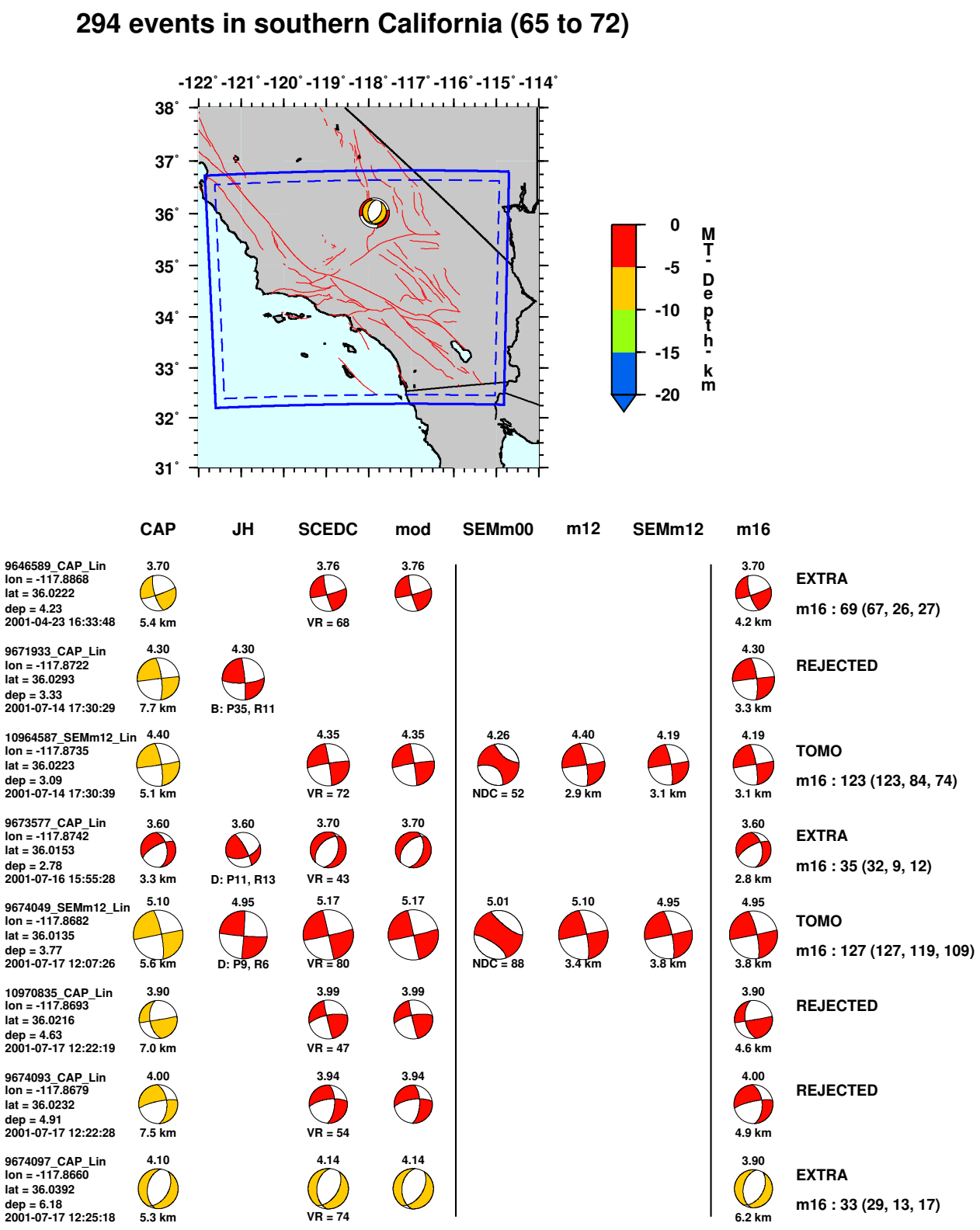


Figure D.9: Source mechanisms considered in the southern California tomography study (65 through 72 out of 294).

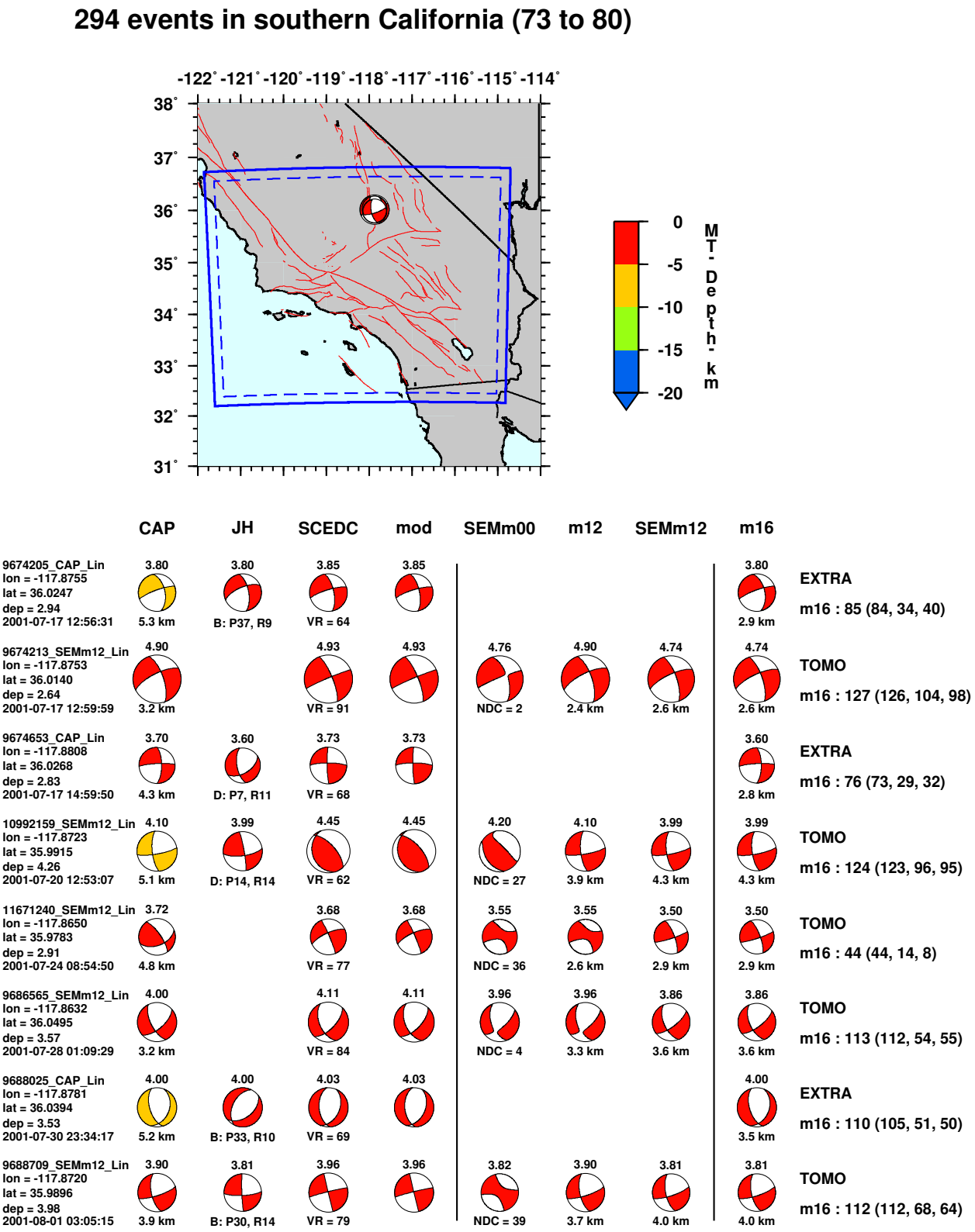


Figure D.10: Source mechanisms considered in the southern California tomography study (73 through 80 out of 294).

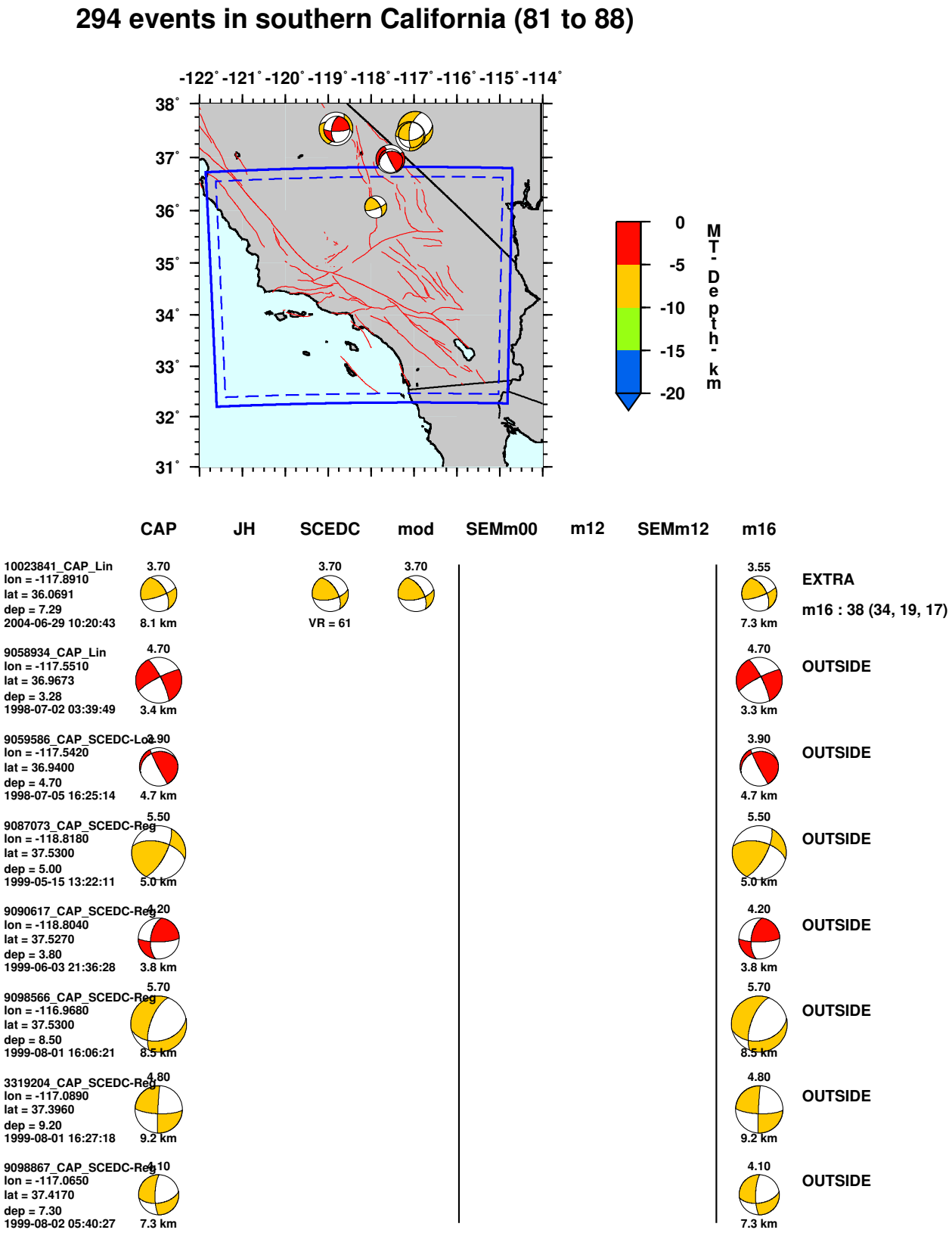


Figure D.11: Source mechanisms considered in the southern California tomography study (81 through 88 out of 294).

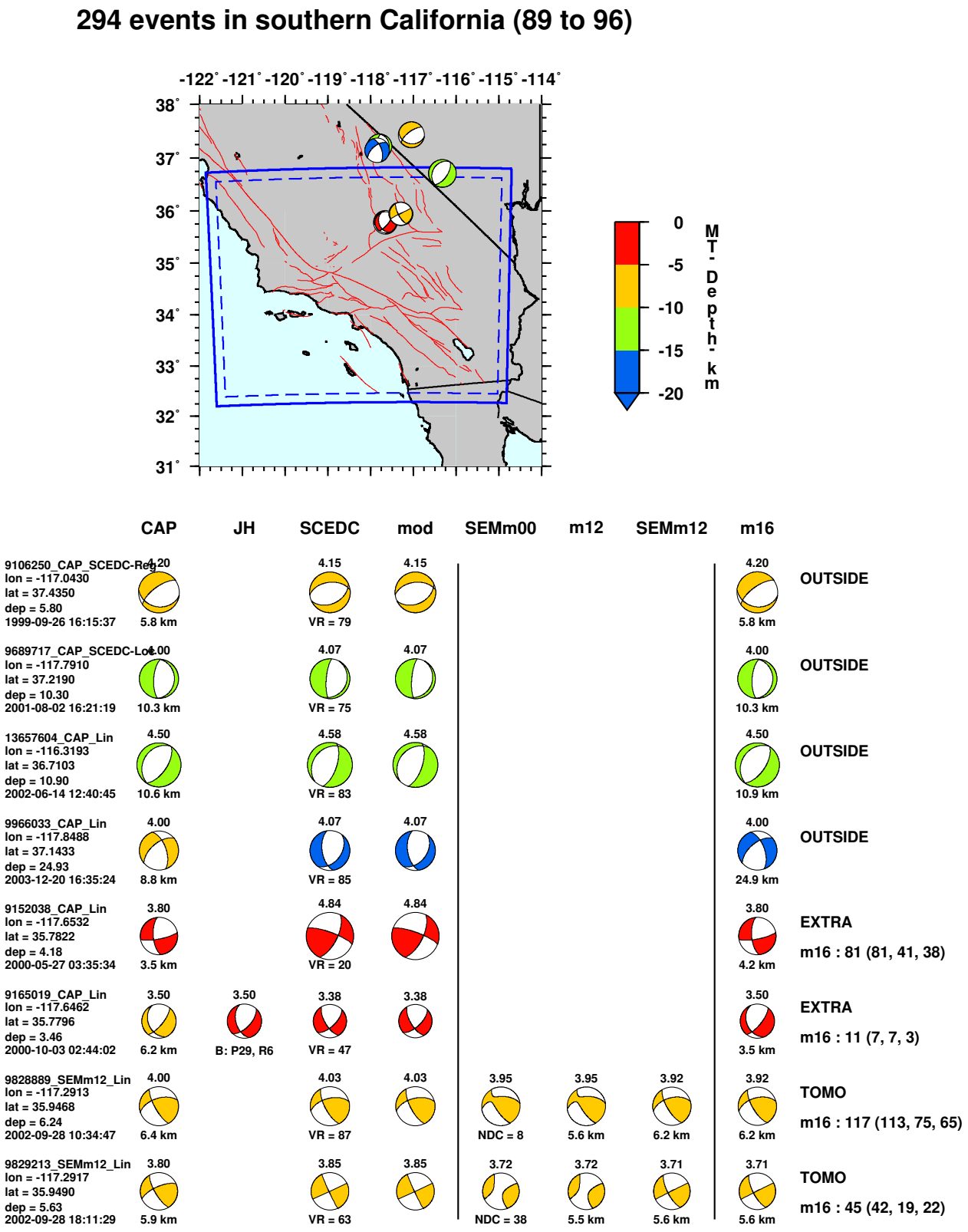


Figure D.12: Source mechanisms considered in the southern California tomography study (89 through 96 out of 294).

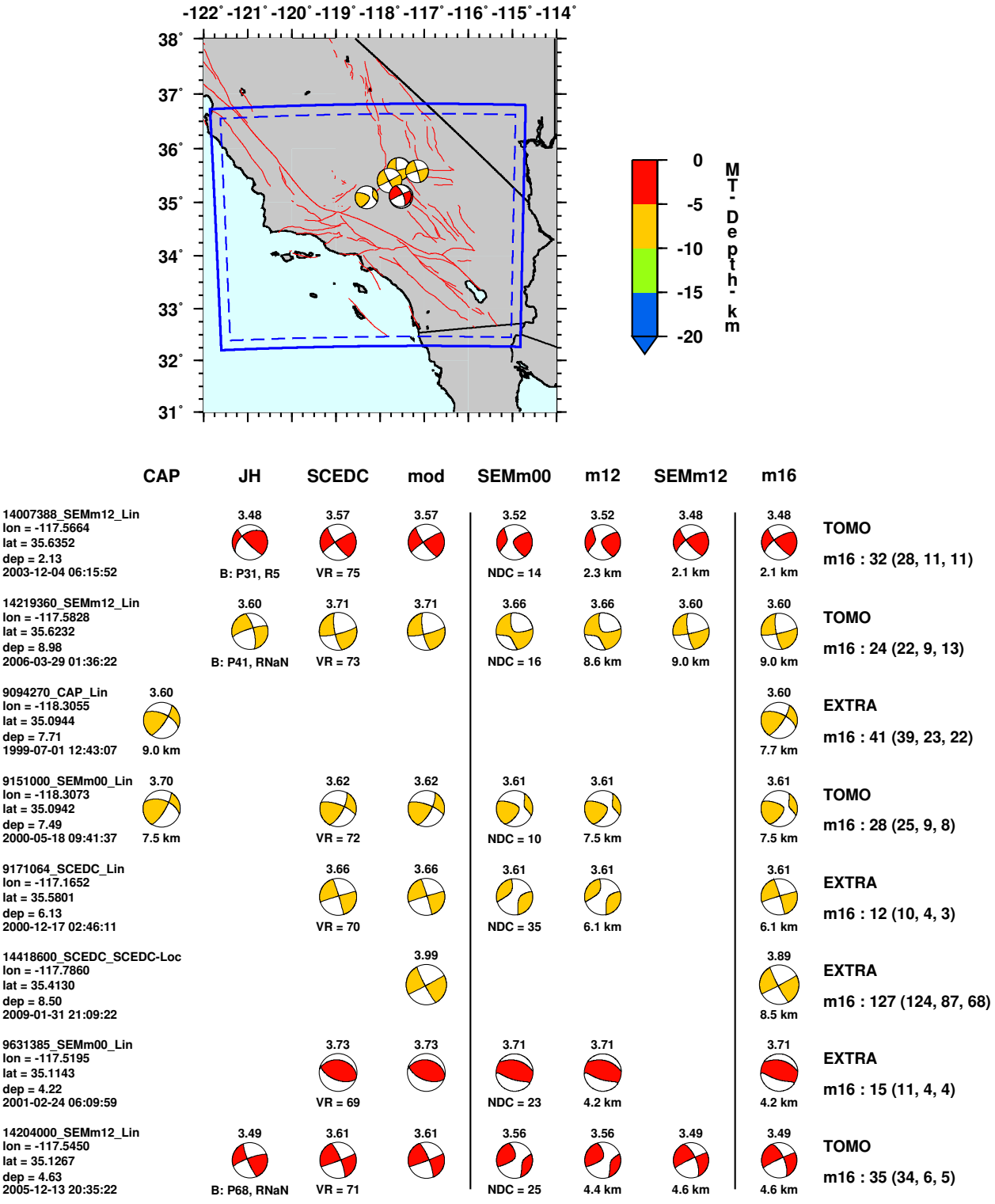


Figure D.13: Source mechanisms considered in the southern California tomography study (97 through 104 out of 294).

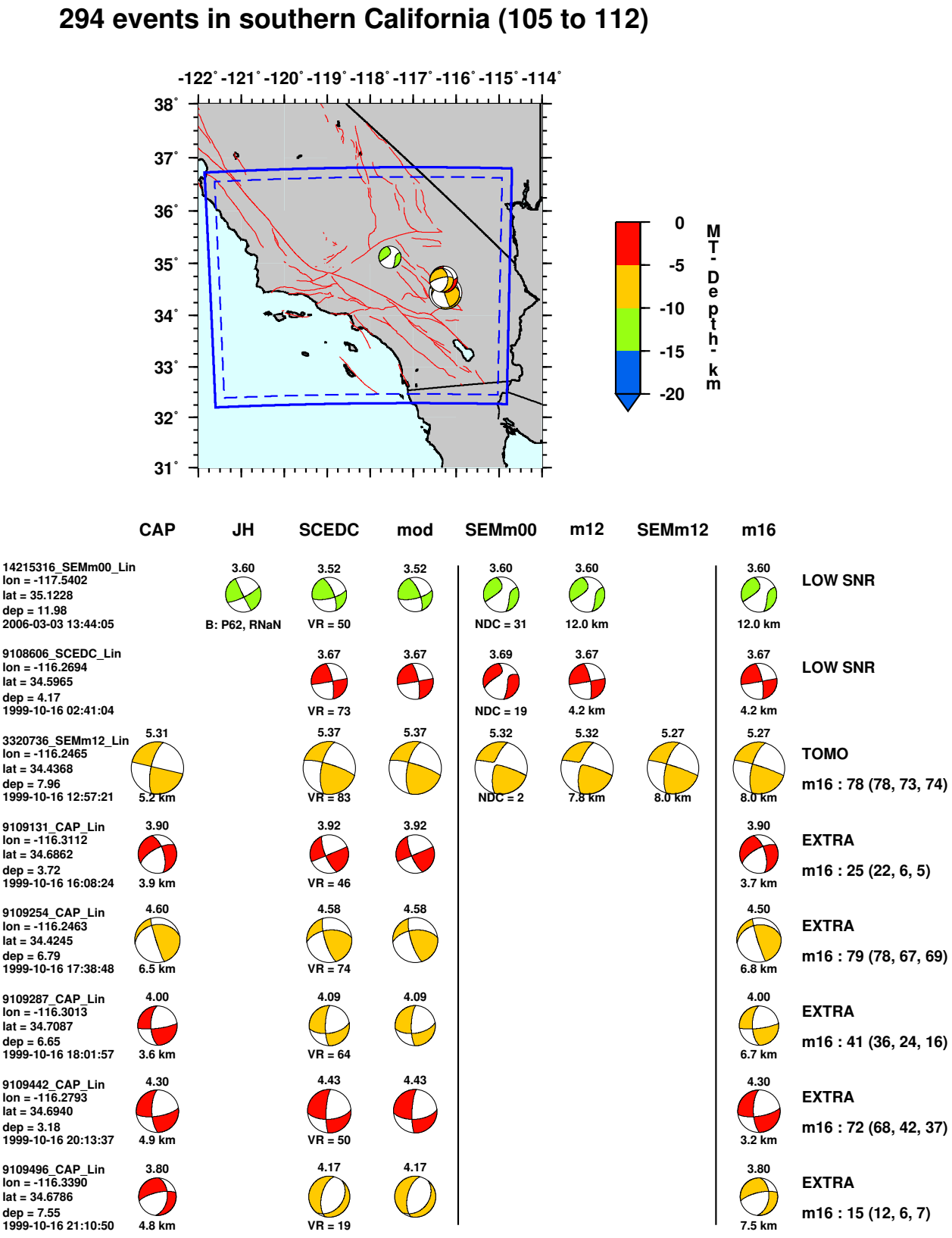


Figure D.14: Source mechanisms considered in the southern California tomography study (105 through 112 out of 294).

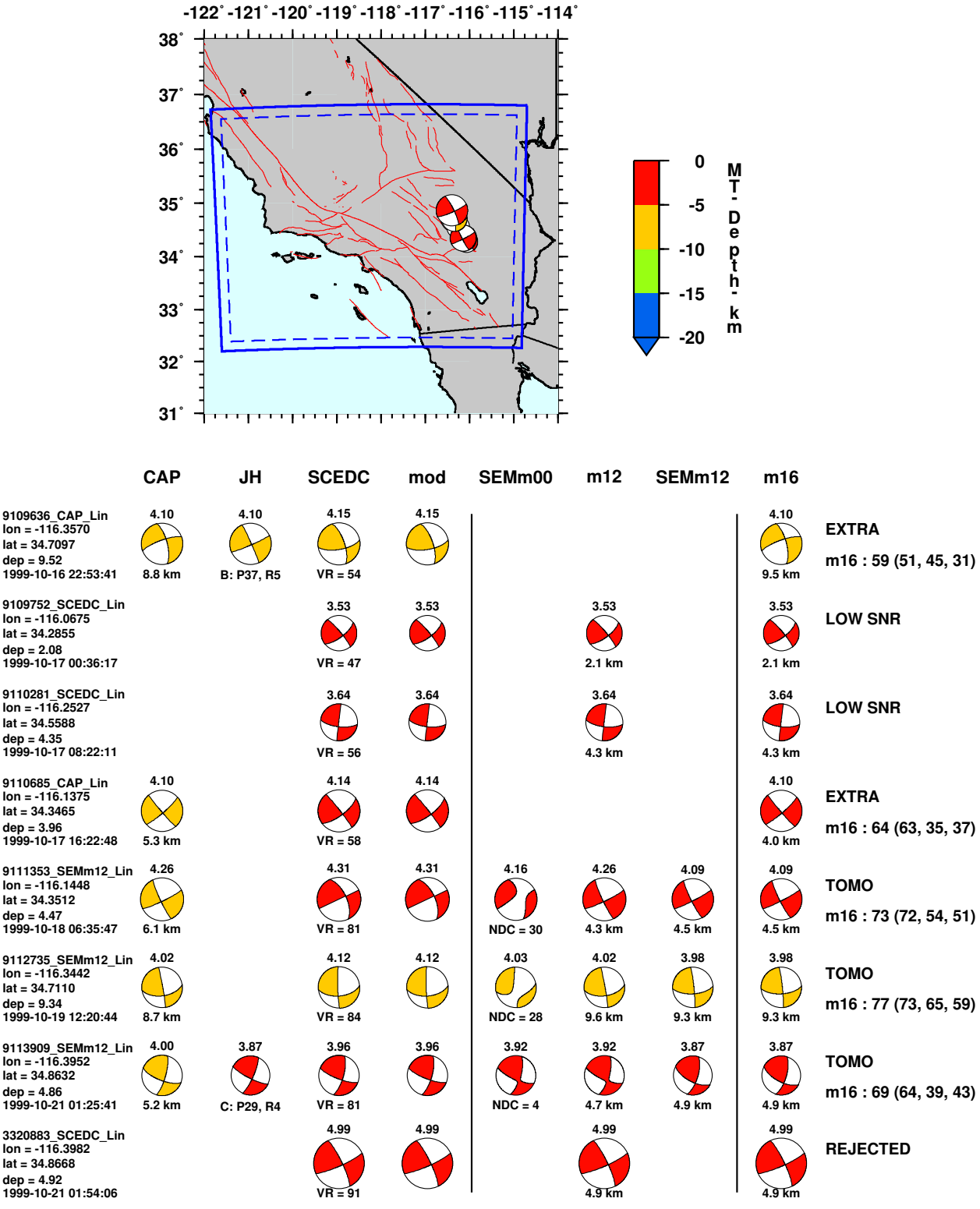


Figure D.15: Source mechanisms considered in the southern California tomography study (113 through 120 out of 294).

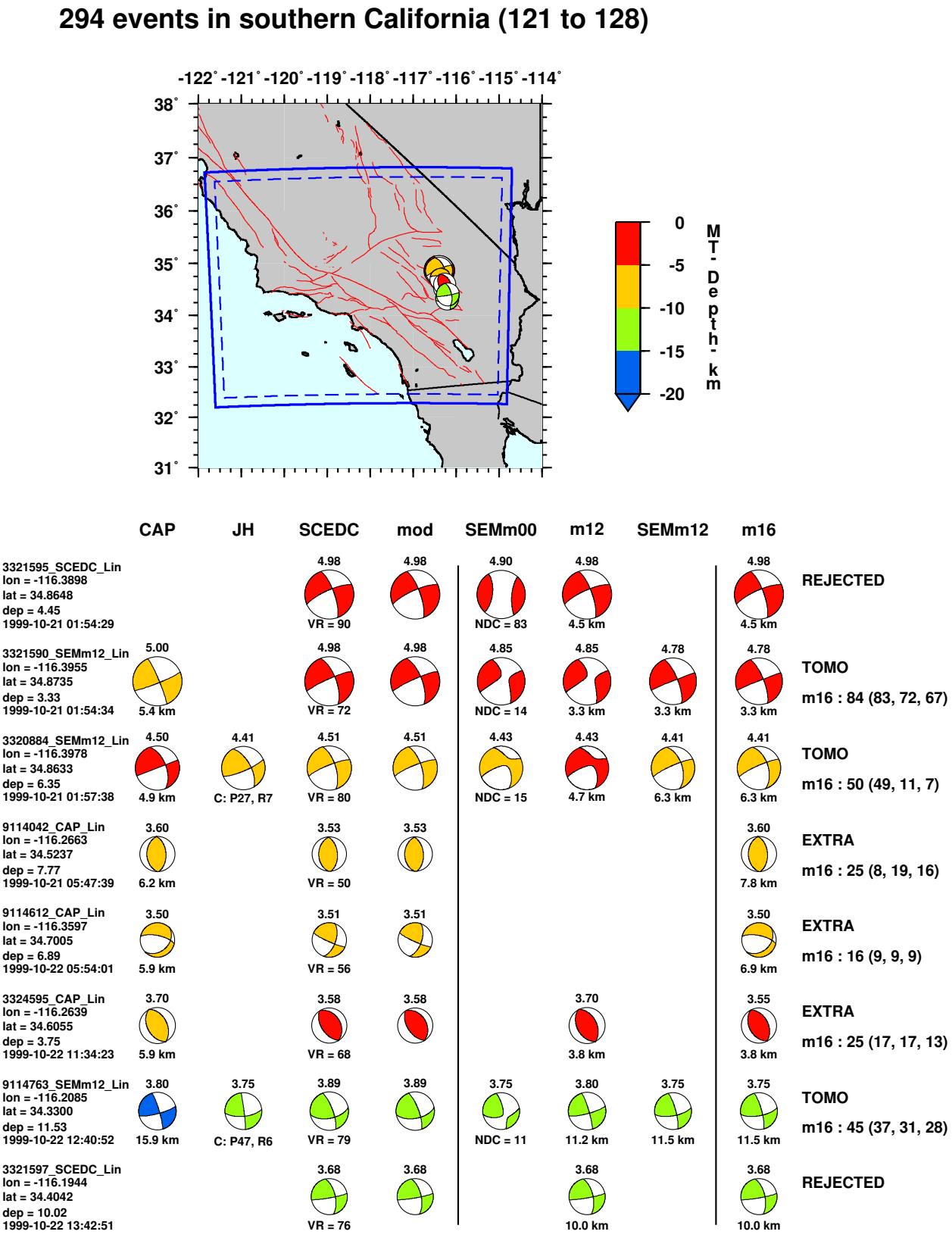


Figure D.16: Source mechanisms considered in the southern California tomography study (121 through 128 out of 294).

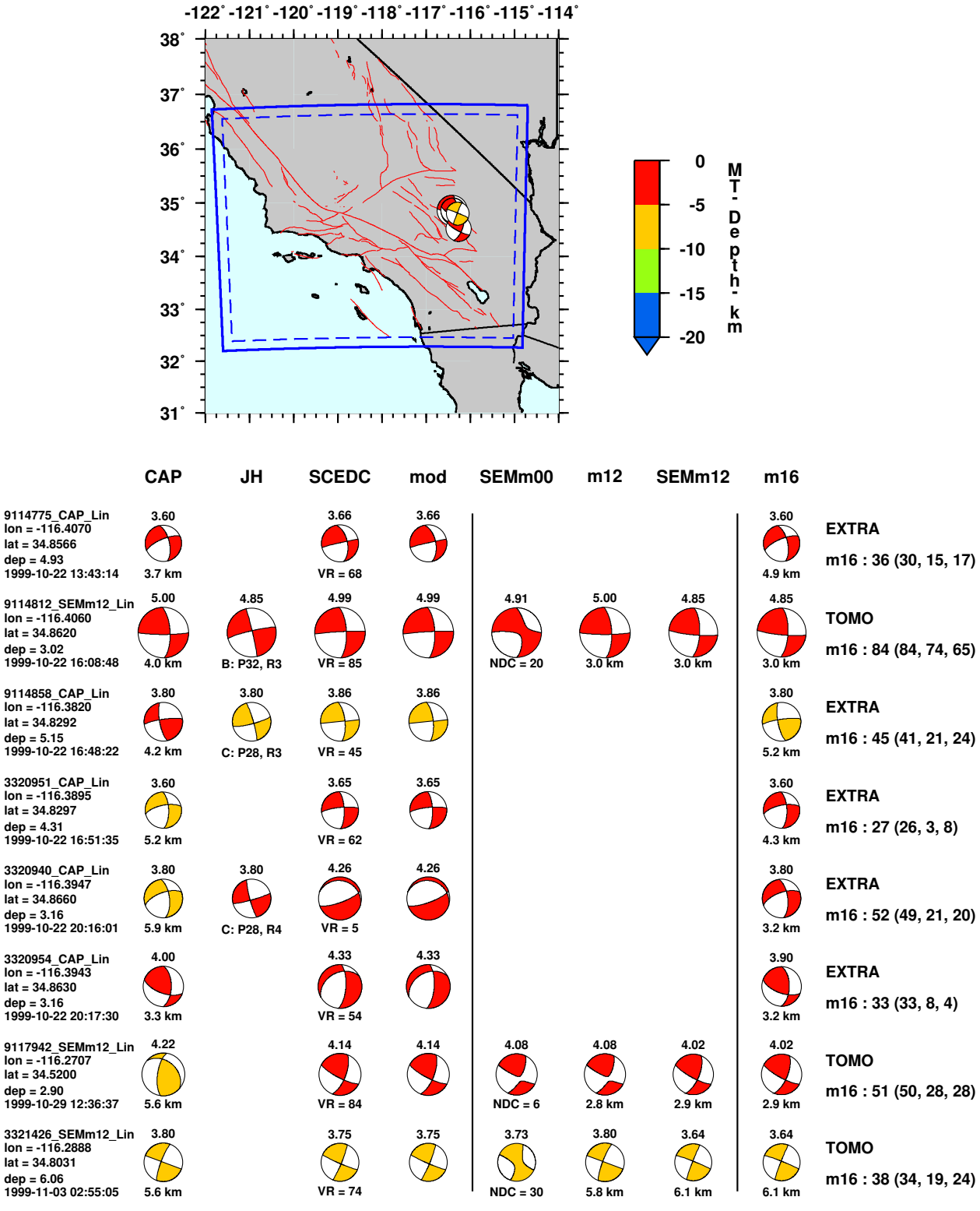


Figure D.17: Source mechanisms considered in the southern California tomography study (129 through 136 out of 294).

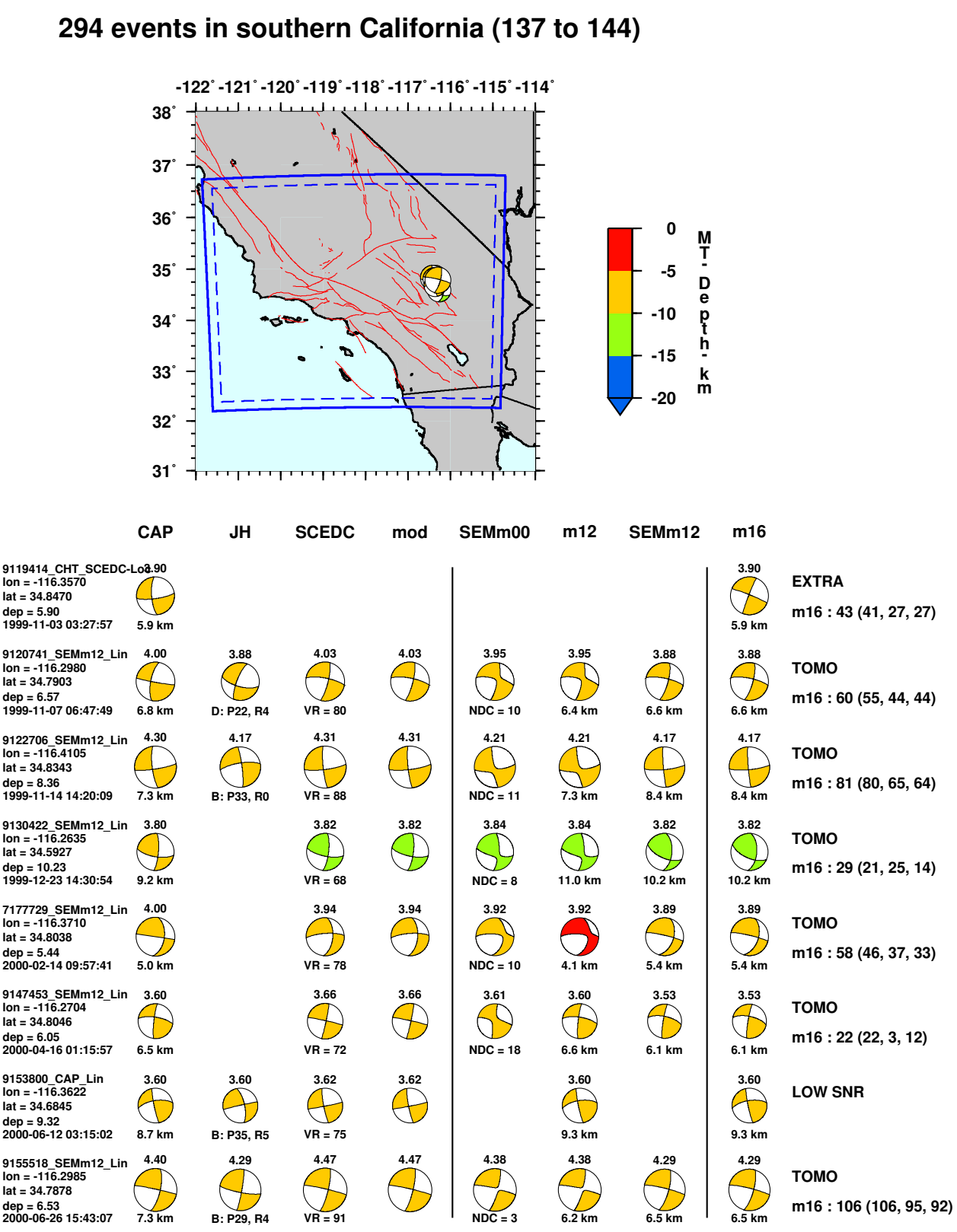


Figure D.18: Source mechanisms considered in the southern California tomography study (137 through 144 out of 294).

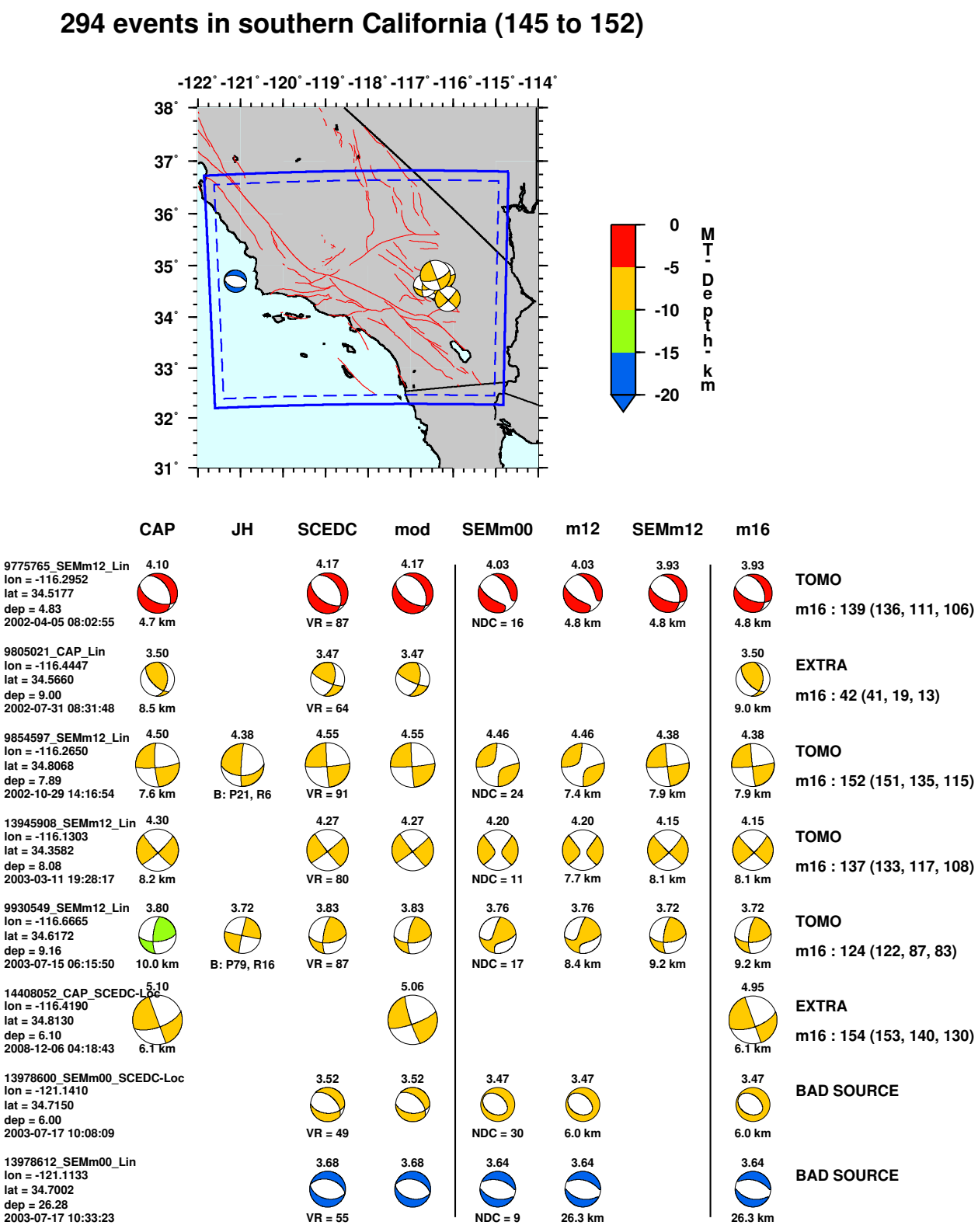


Figure D.19: Source mechanisms considered in the southern California tomography study (145 through 152 out of 294).

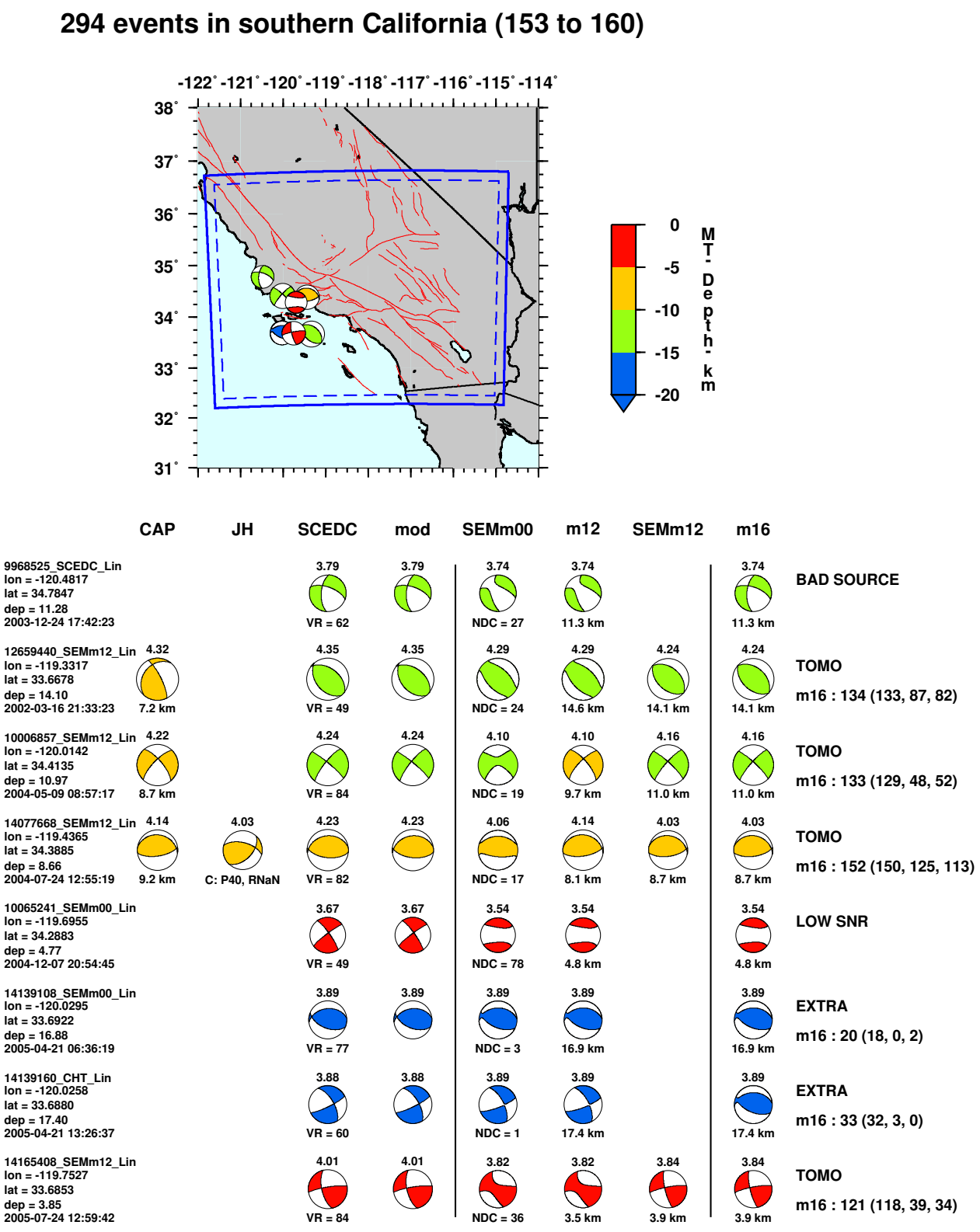


Figure D.20: Source mechanisms considered in the southern California tomography study (153 through 160 out of 294).

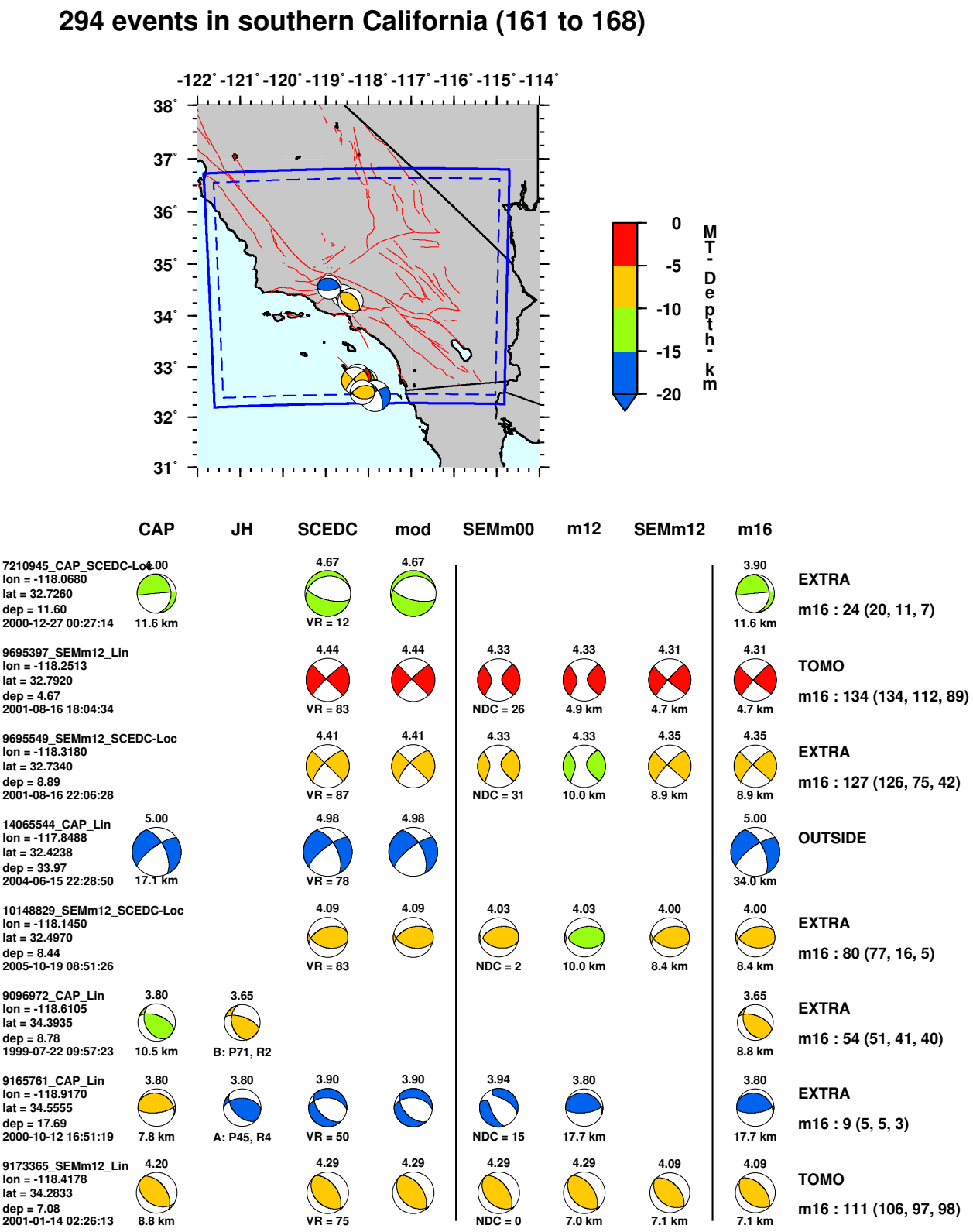


Figure D.21: Source mechanisms considered in the southern California tomography study (161 through 168 out of 294).

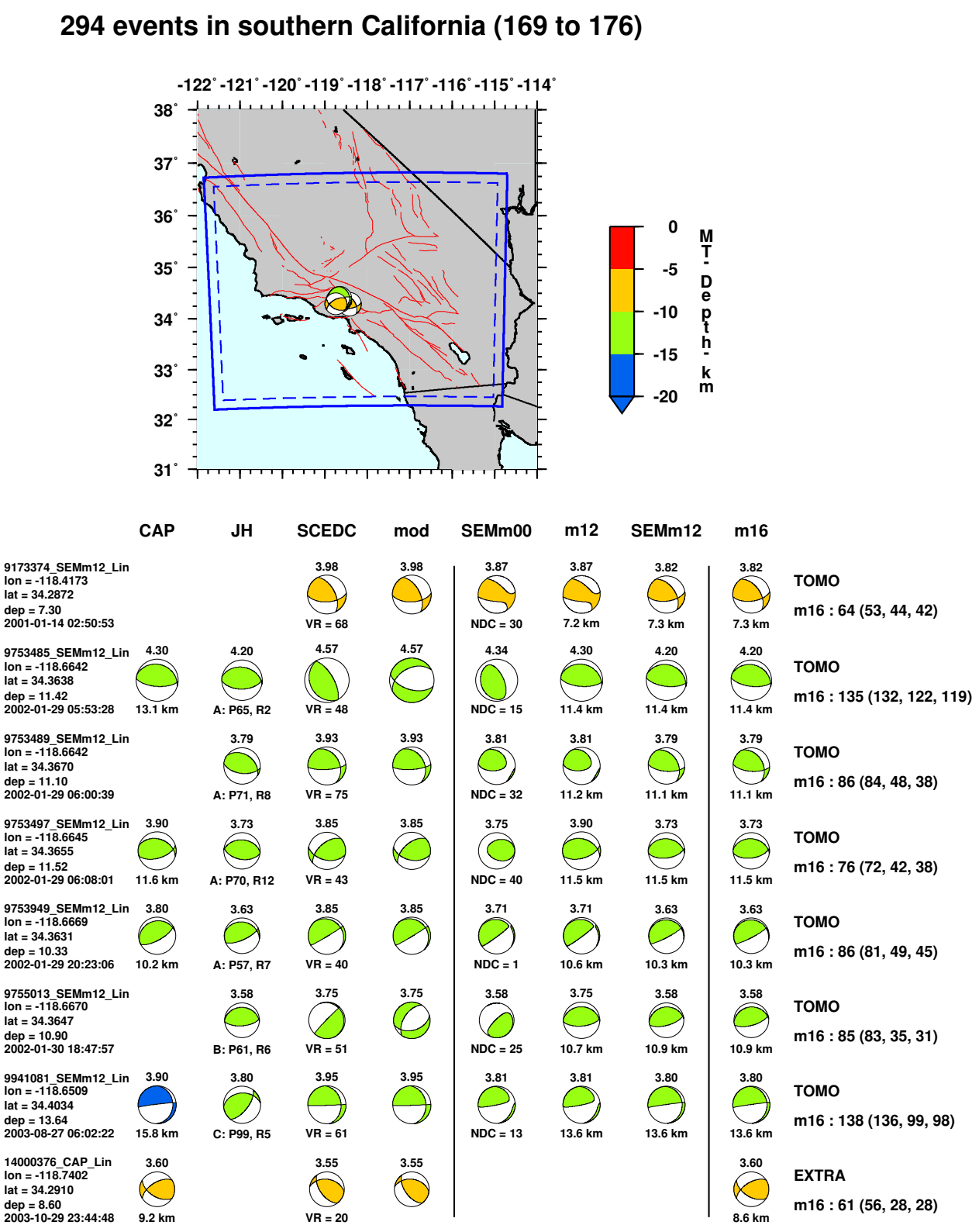


Figure D.22: Source mechanisms considered in the southern California tomography study (169 through 176 out of 294).

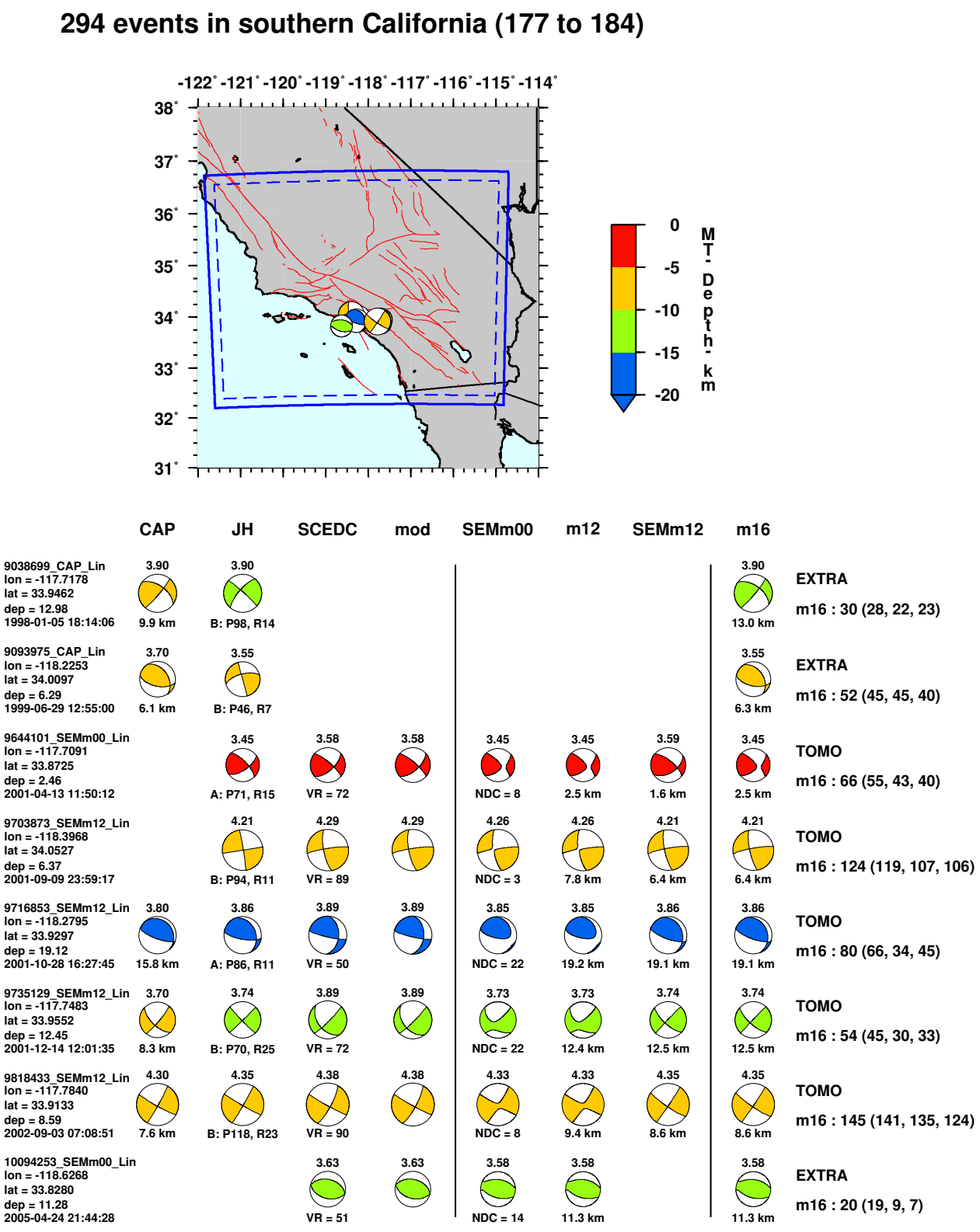


Figure D.23: Source mechanisms considered in the southern California tomography study (177 through 184 out of 294).

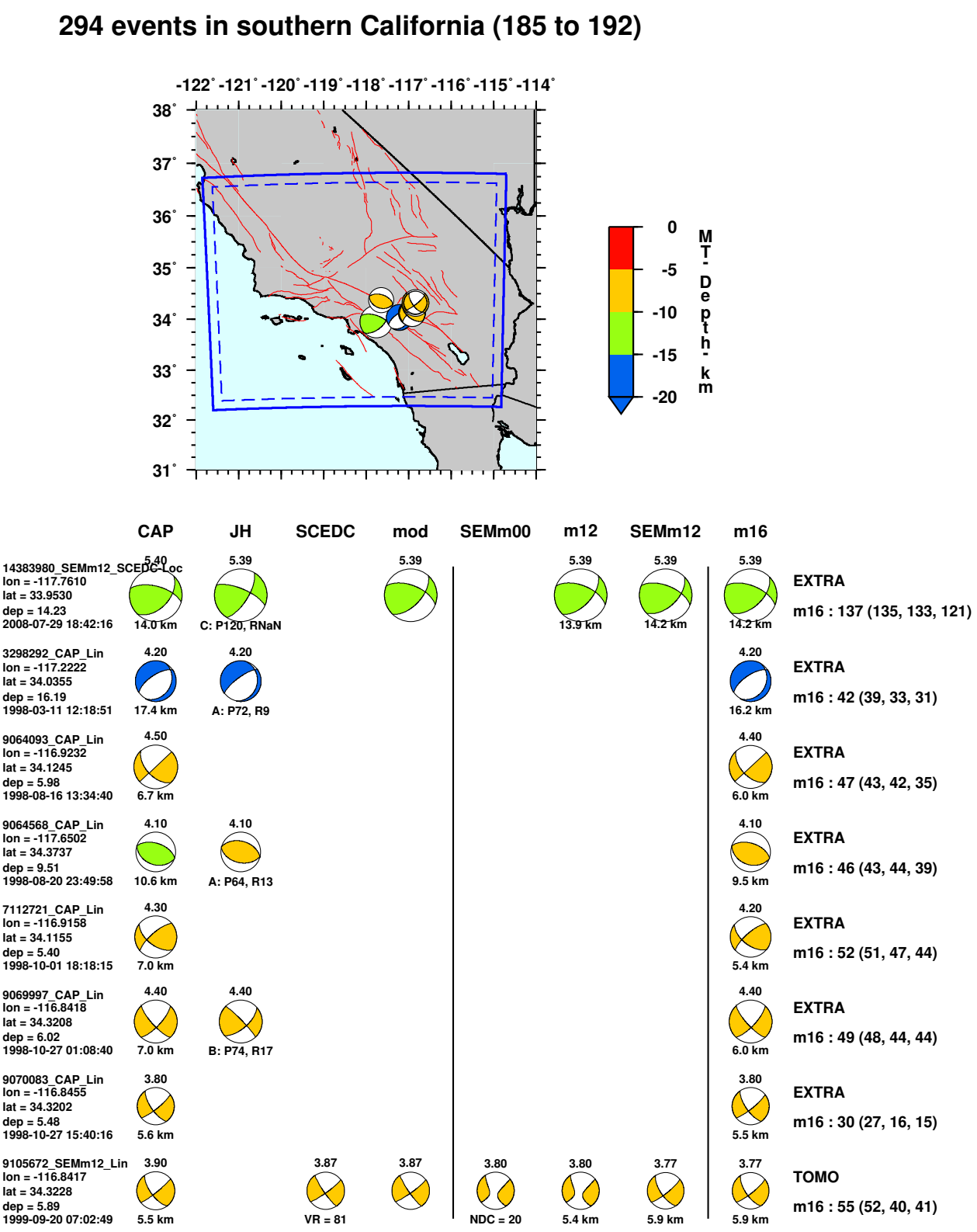


Figure D.24: Source mechanisms considered in the southern California tomography study (185 through 192 out of 294).

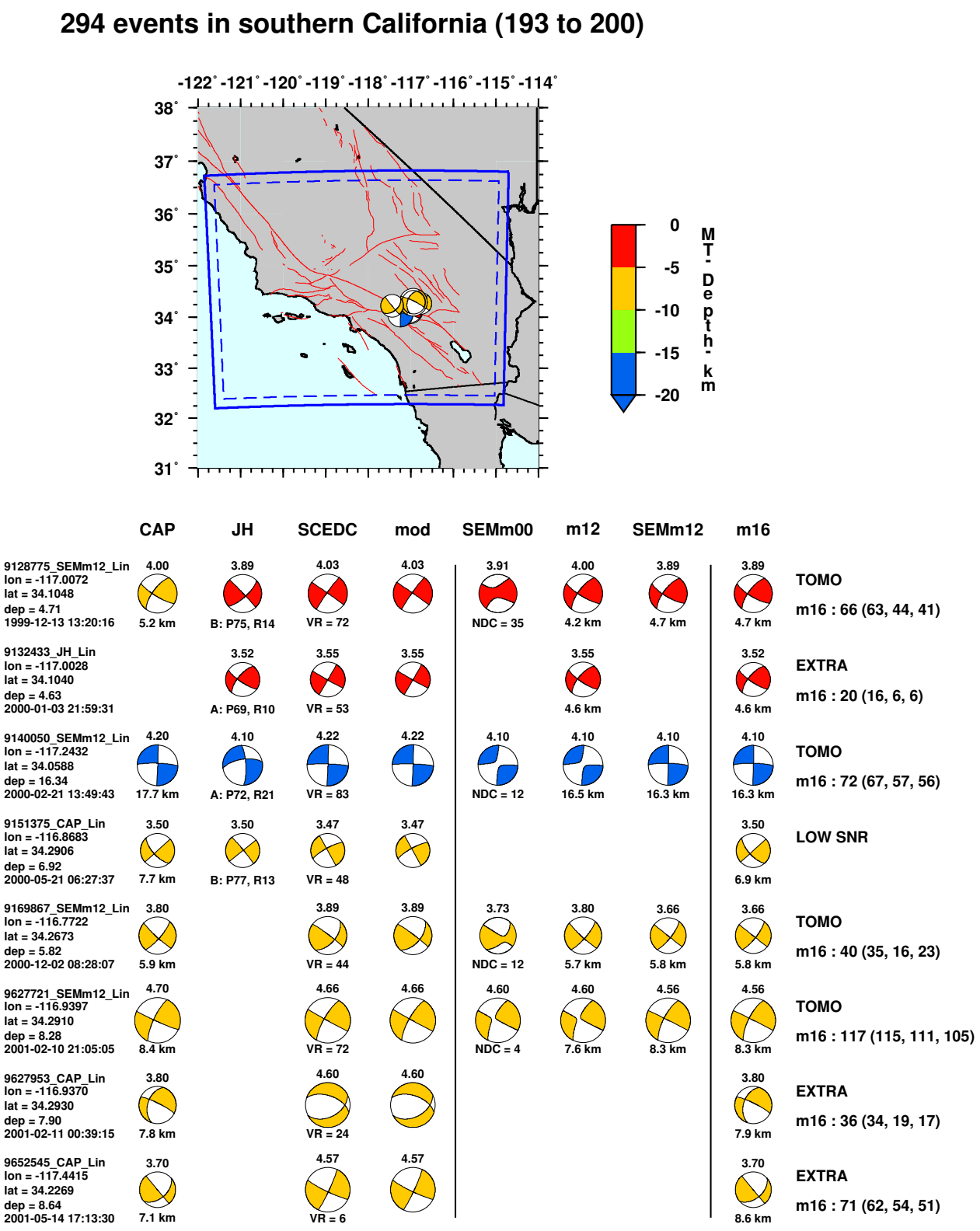


Figure D.25: Source mechanisms considered in the southern California tomography study (193 through 200 out of 294).

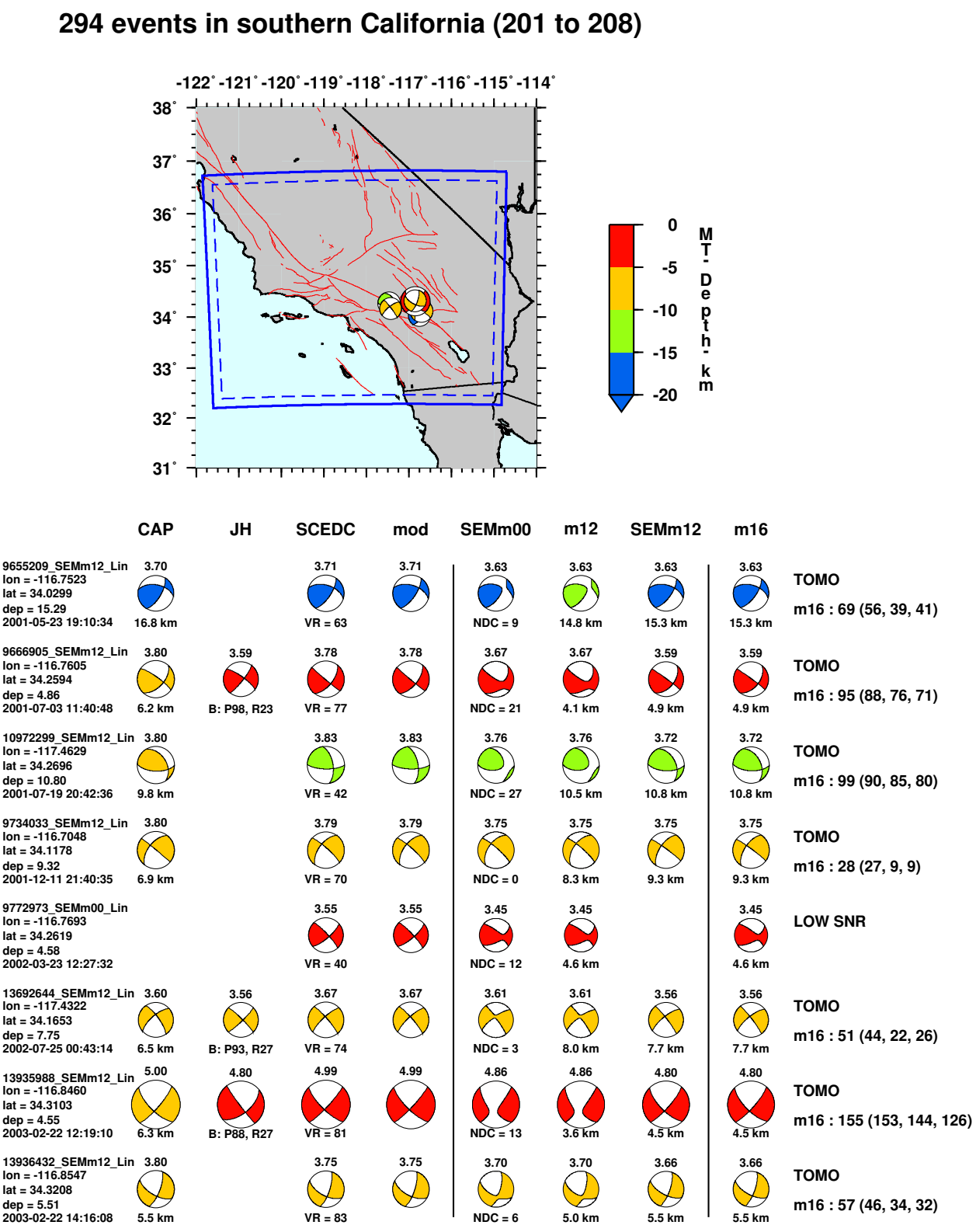


Figure D.26: Source mechanisms considered in the southern California tomography study (201 through 208 out of 294).

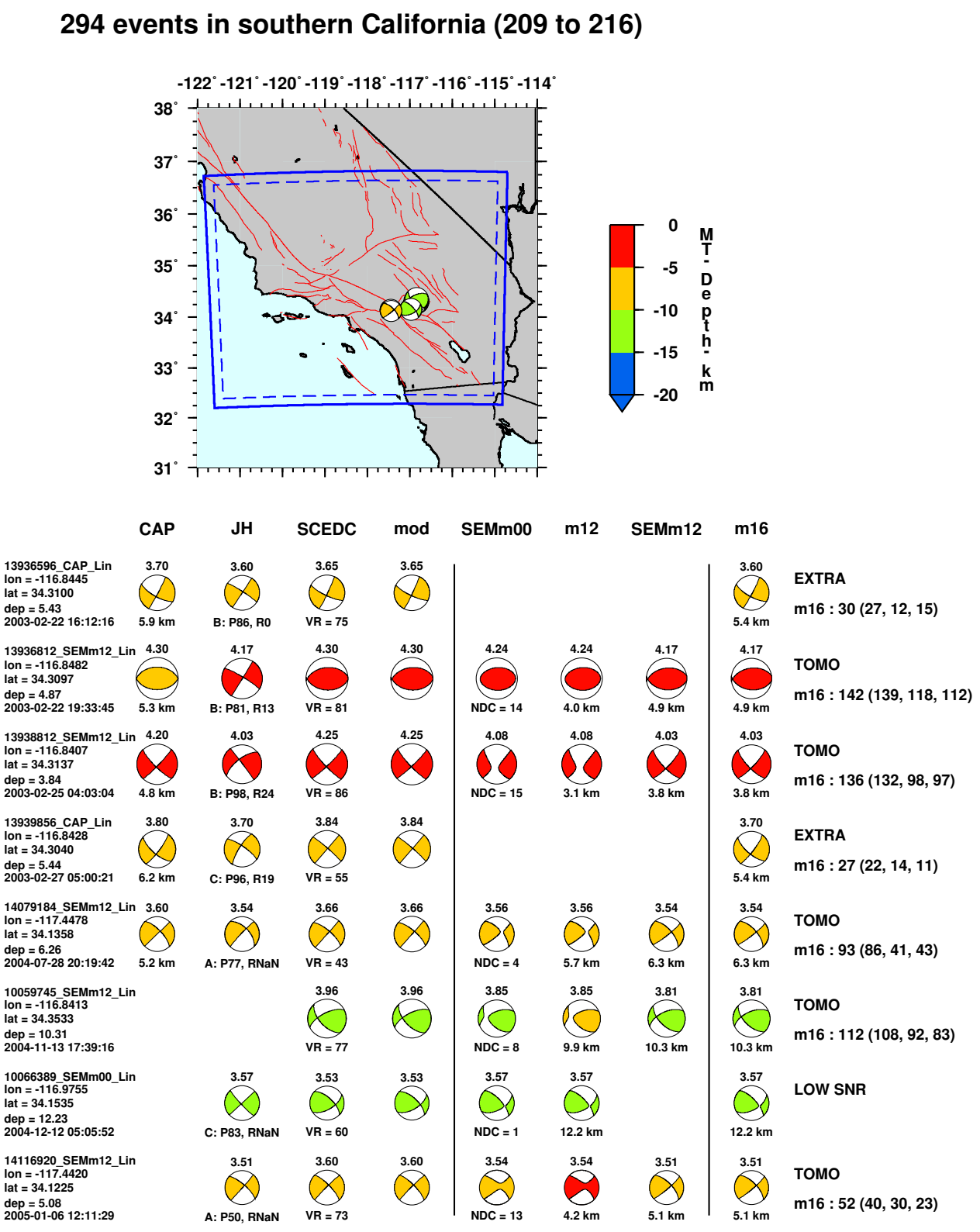


Figure D.27: Source mechanisms considered in the southern California tomography study (209 through 216 out of 294).

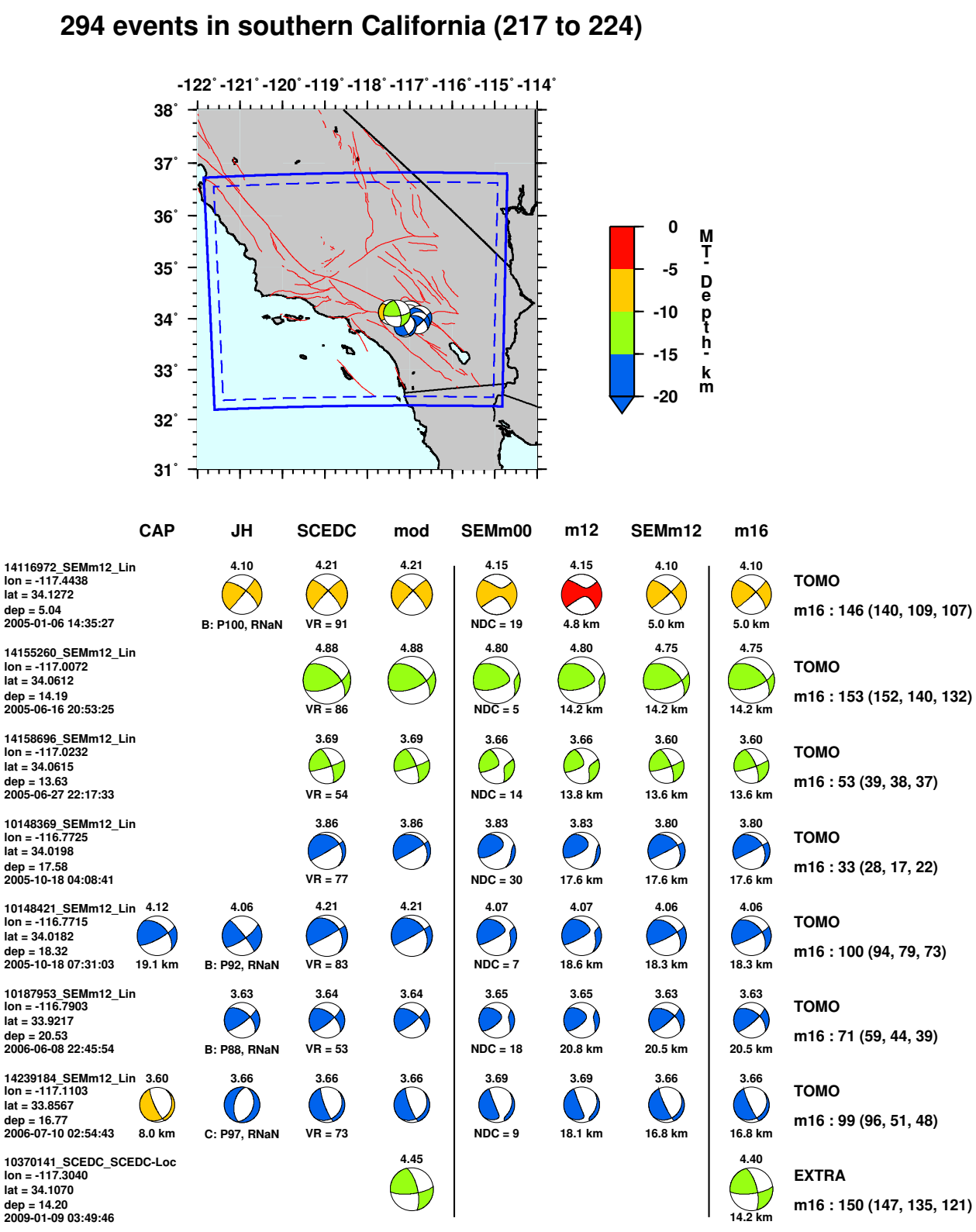


Figure D.28: Source mechanisms considered in the southern California tomography study (217 through 224 out of 294).

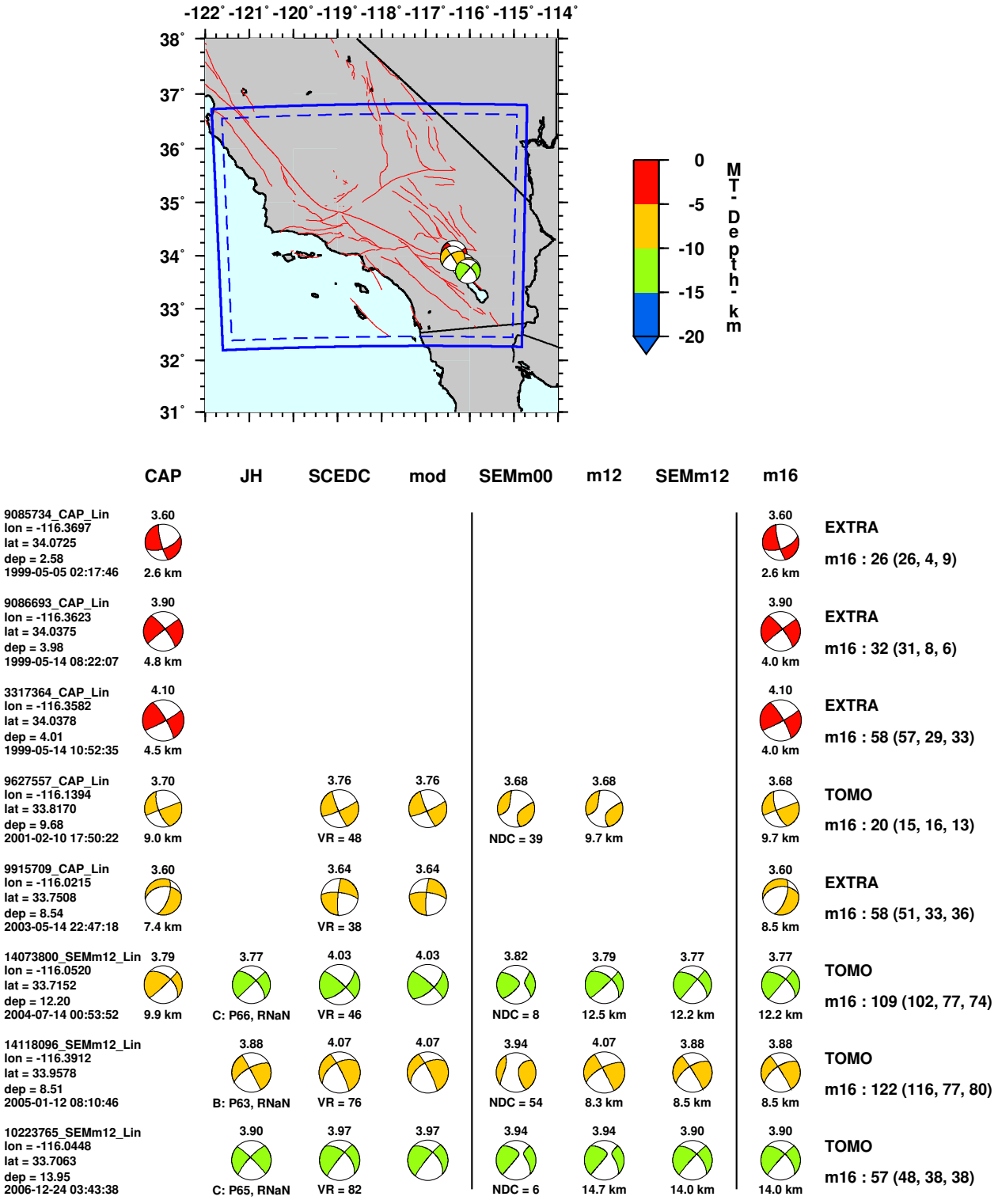


Figure D.29: Source mechanisms considered in the southern California tomography study (225 through 232 out of 294).

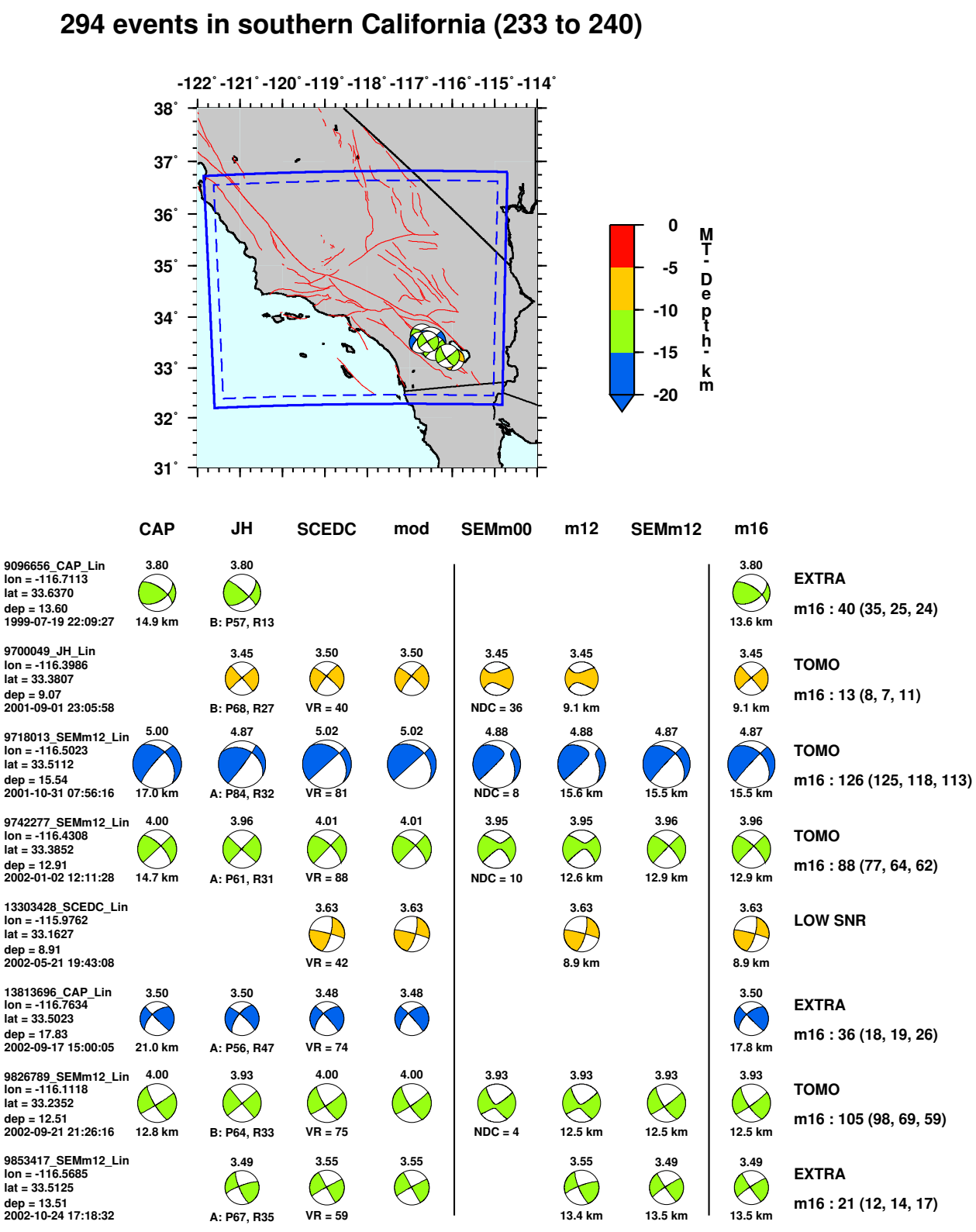


Figure D.30: Source mechanisms considered in the southern California tomography study (233 through 240 out of 294).

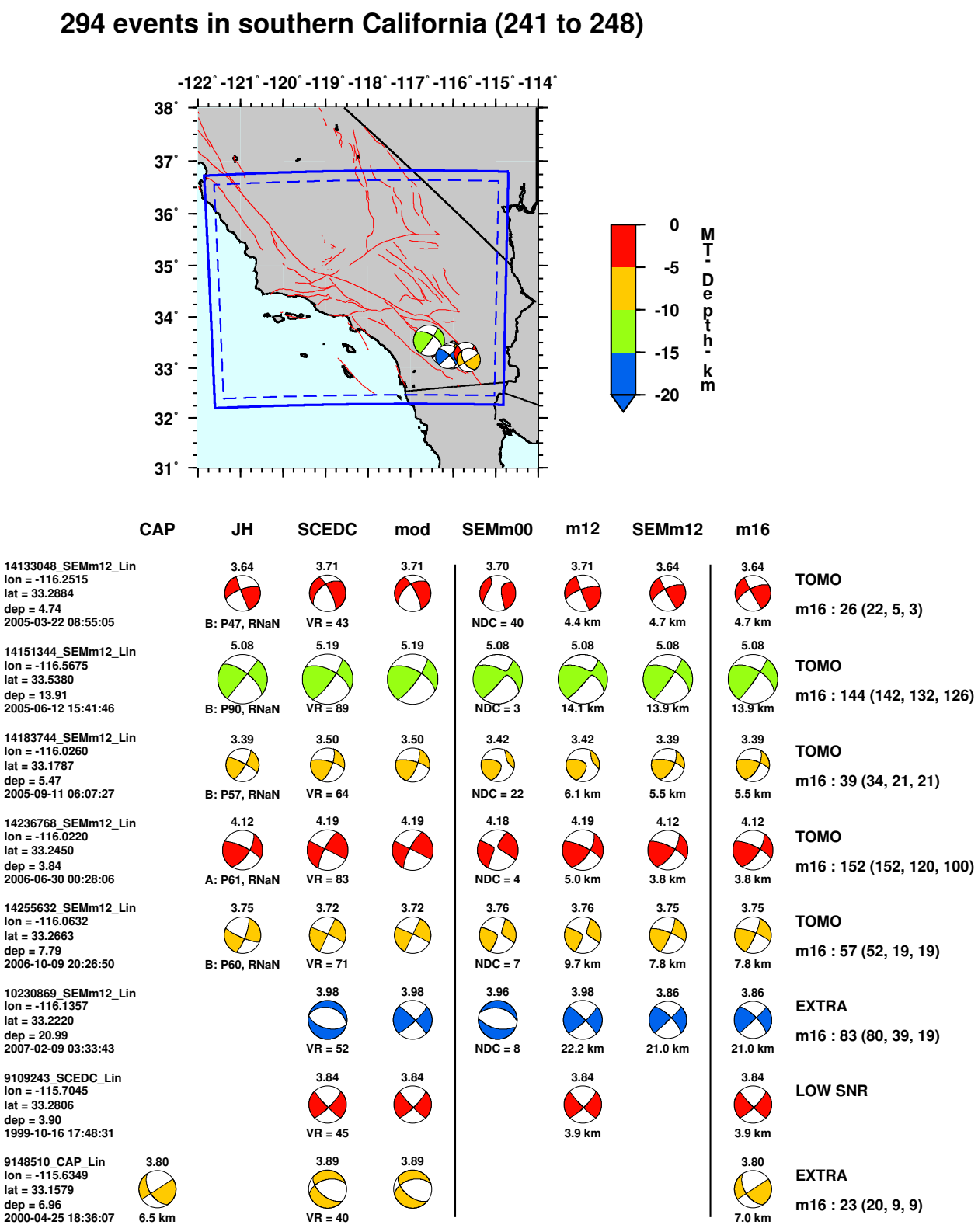


Figure D.31: Source mechanisms considered in the southern California tomography study (241 through 248 out of 294).

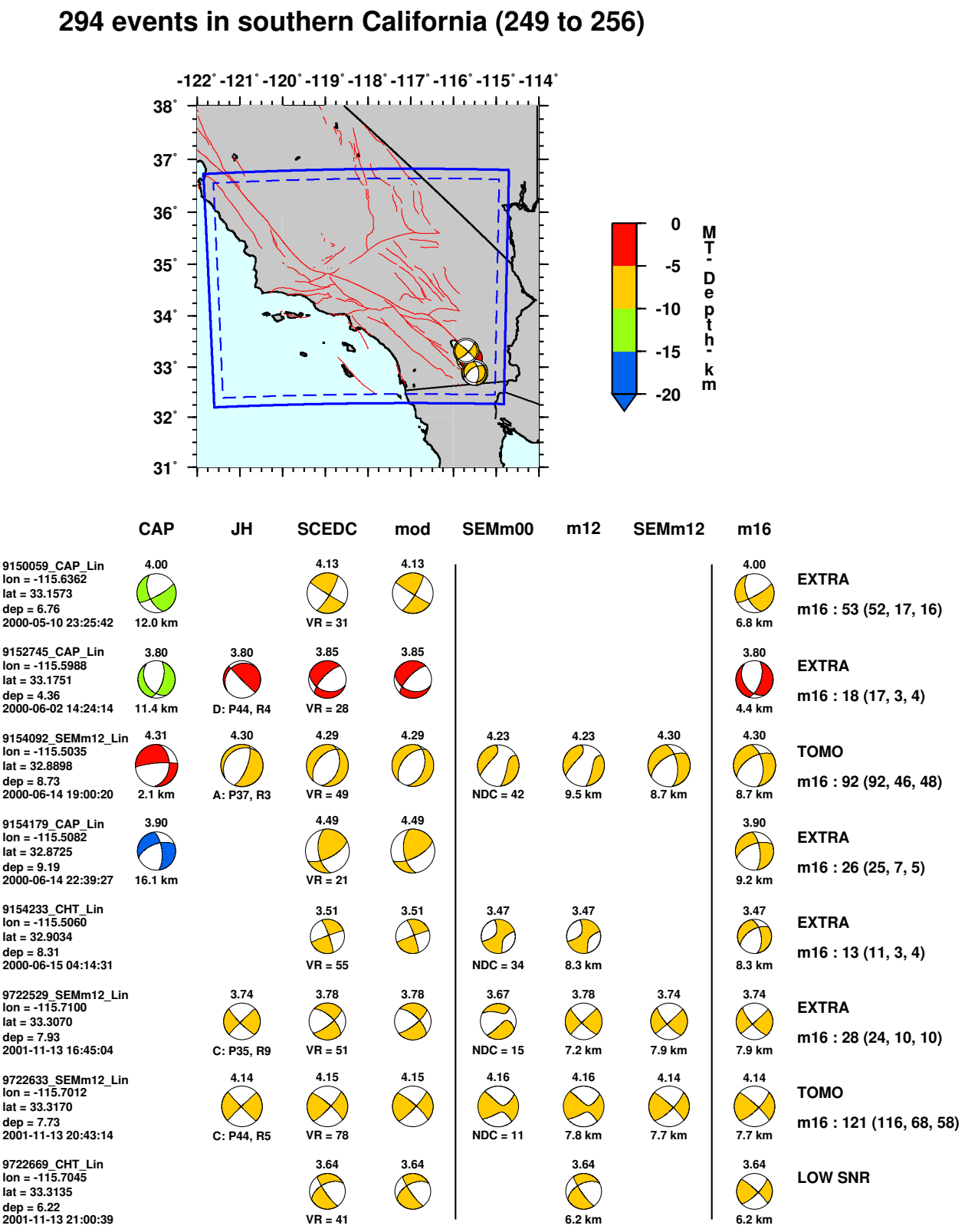


Figure D.32: Source mechanisms considered in the southern California tomography study (249 through 256 out of 294).

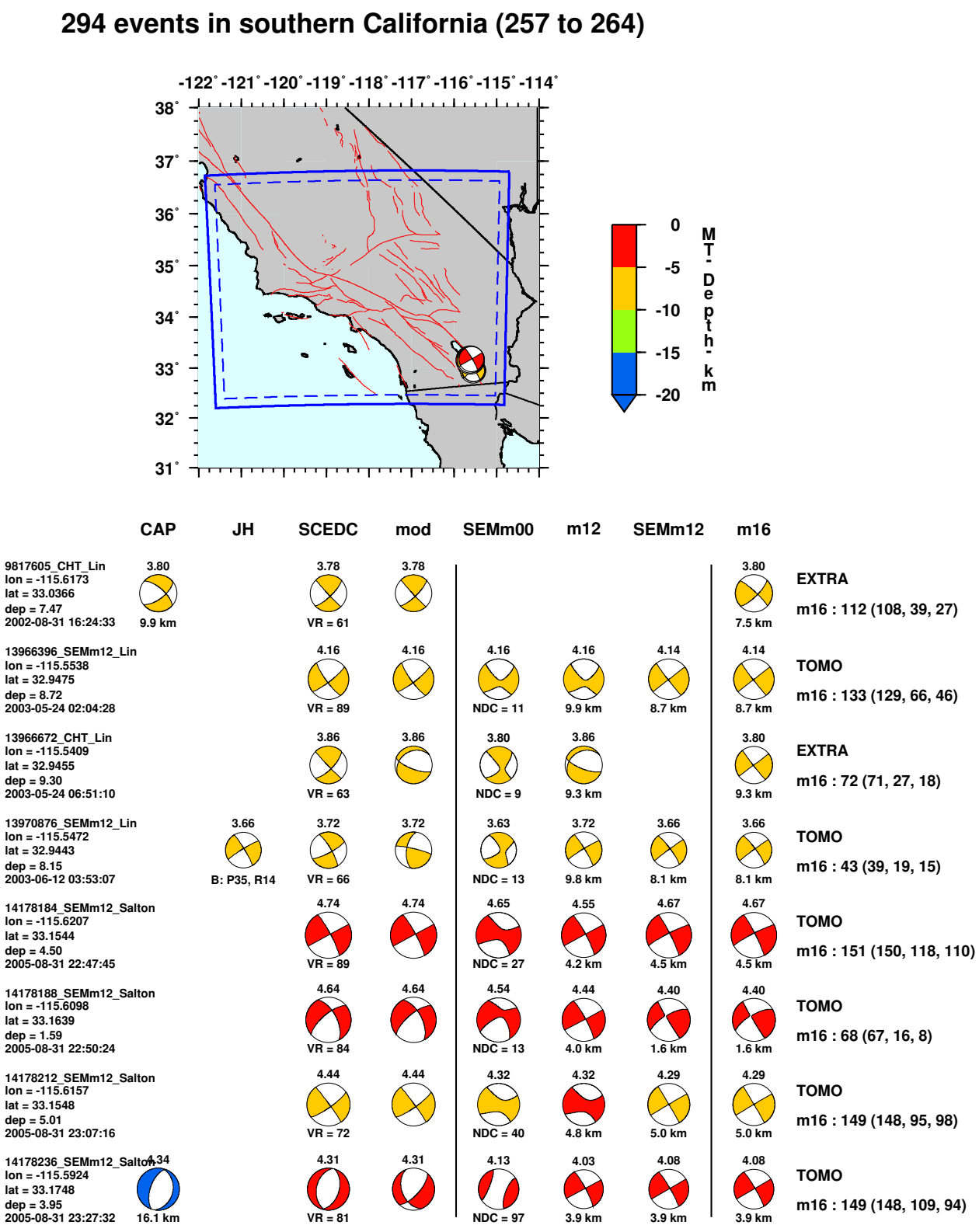


Figure D.33: Source mechanisms considered in the southern California tomography study (257 through 264 out of 294).

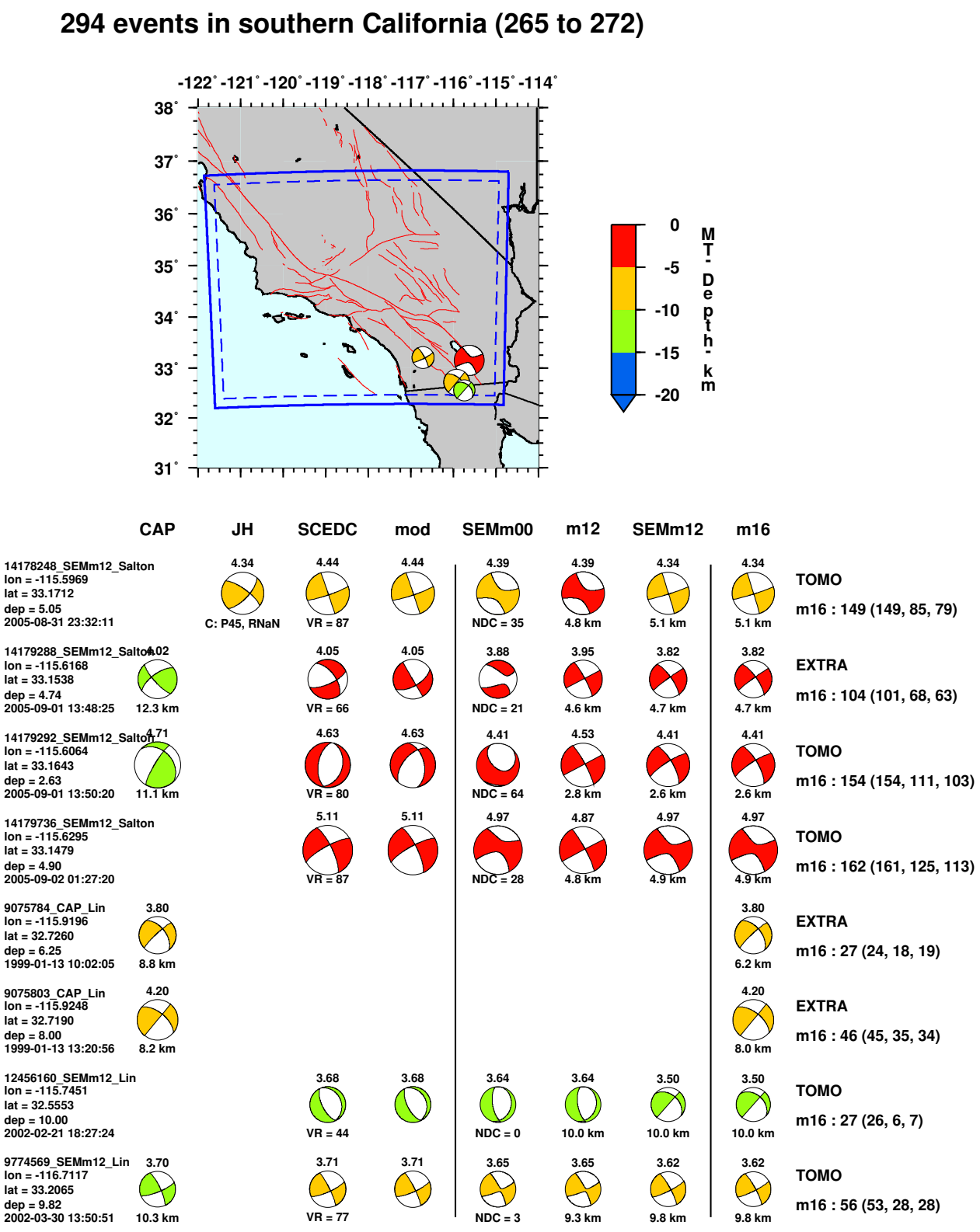


Figure D.34: Source mechanisms considered in the southern California tomography study (265 through 272 out of 294).

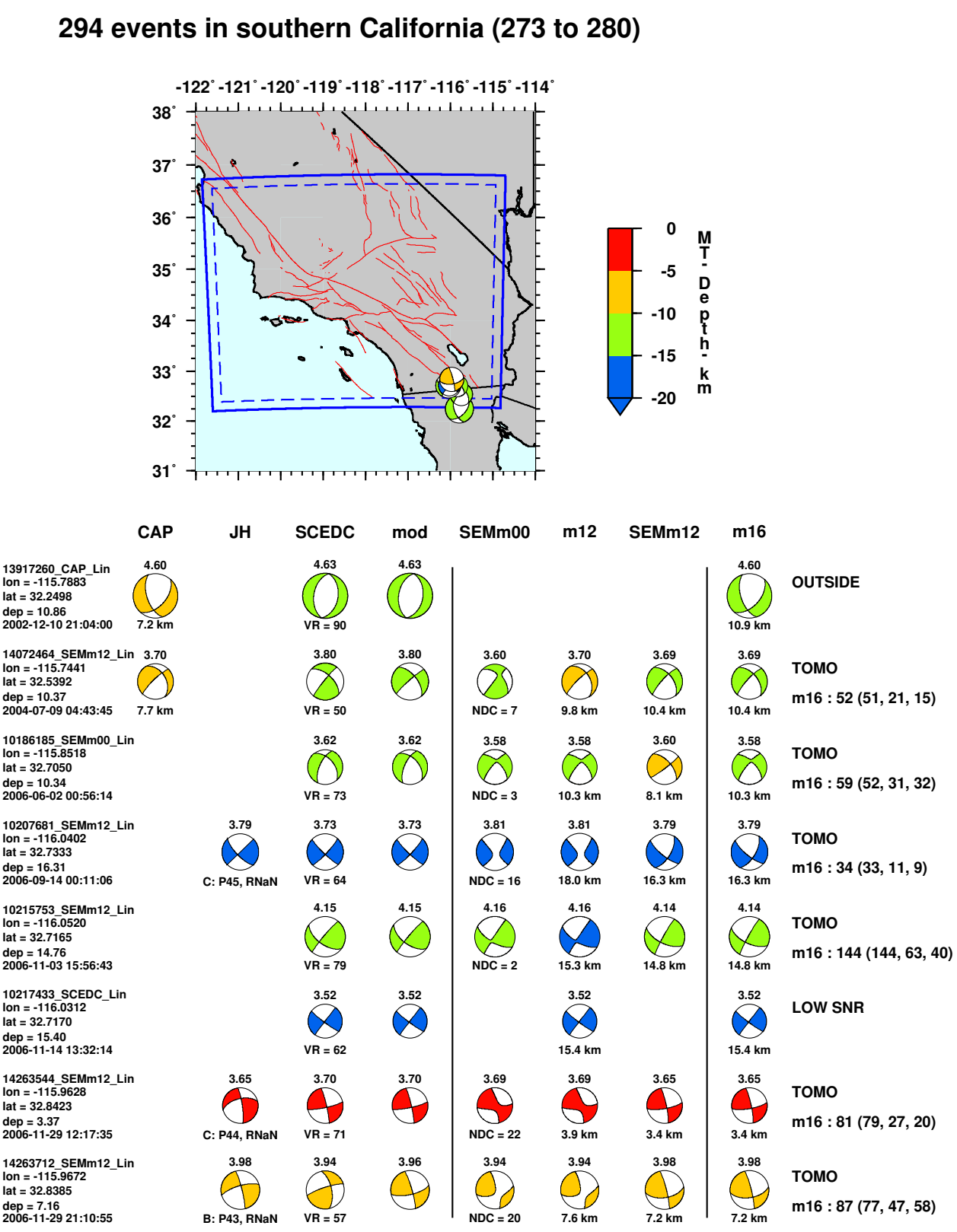


Figure D.35: Source mechanisms considered in the southern California tomography study (273 through 280 out of 294).

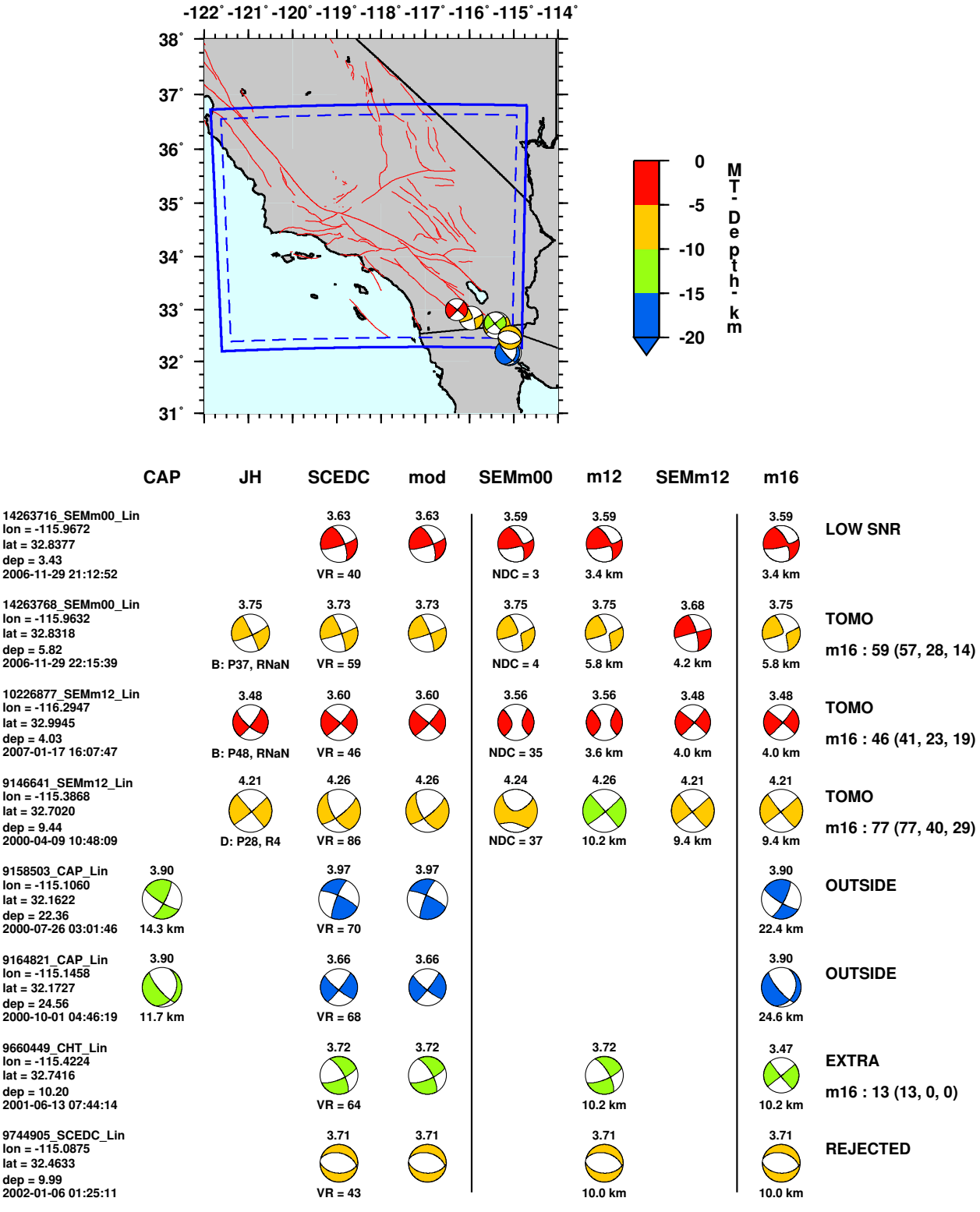


Figure D.36: Source mechanisms considered in the southern California tomography study (281 through 288 out of 294).

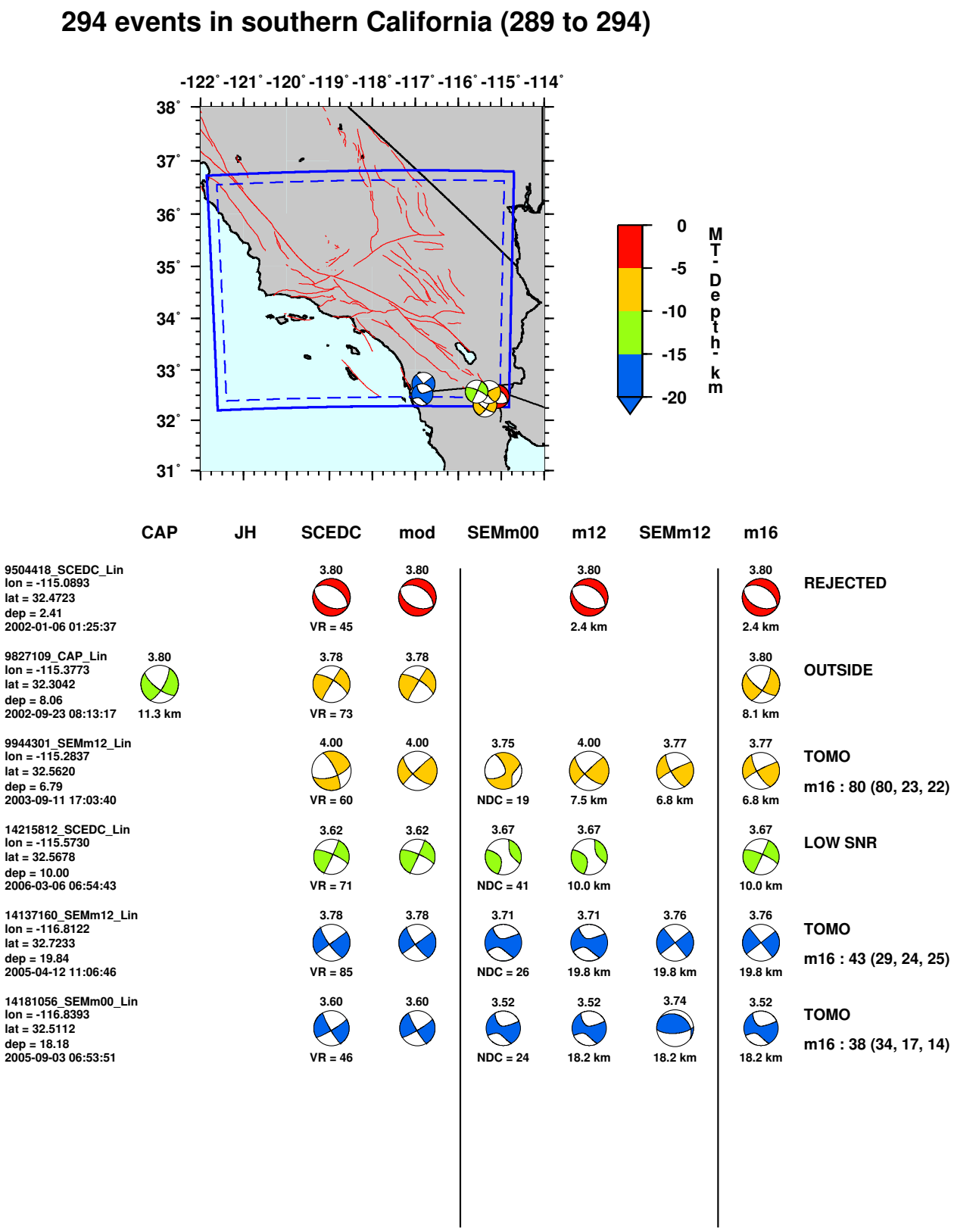


Figure D.37: Source mechanisms considered in the southern California tomography study (289 through 294 out of 294).

Appendix E

Polarity problems at selected stations in southern California

Note

I am fortunate to have had close contact with the Southern California Data Center. I thank Ellen Yu, Egill Hauksson, and Kate Hutton for discussions regarding the matters presented in this appendix. The key results are summarized in Table E.1.

E.1 Overview

Our tomography study for southern California has aimed to incorporate three-component waveform data from all available broadband stations for 234 earthquakes, $M_{\rm w}$ = 3.5–5.5, over the time period 1998–2009. For these earthquakes we have generated synthetic seismograms using a 3D crustal model provided by the Southern California Earthquake Center, which we improved with 16 iterations in a tomographic inversion. Based on thousands of comparisons between synthetic and recorded seismograms, I have discovered a problem with the polarity of certain stations for specific epochs. The polarity problem is summarized in Table E.1 and in the following figures. I demonstrate the problematic records using the seismograms filtered at relatively long periods (bandpass 6–30 s), which are not as sensitive to possible "site effect," i.e., strong heterogeneity in the immediate region of the station. Because the station coverage in southern California is dense, it is usually possible to find a nearby station to the problematic one, in order to demonstrate the problem. One example

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.

	Earthquake dates		Corresponding Epochs		Channels	
Station	Earliest	Latest	Start	End	(BH_)	Figures
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

is shown in Figures E.7–E.9 for station PER.CI for an earthquake on 2008.12.06. From stations MSJ.CI to PER.CI to RVR.CI, I sweep the azimuth in a clockwise manner. The records for MSJ and RVR are similar, but the station in between, BTP, has the polarity flipped for both the transverse (T) and radial (R) components.

Most of the stations with reported polarity problems are not exhibiting the problems at present. In other words, the problems are restricted to specific epochs of the stations, and may be restricted to particular components as well (Table E.1). Of course, it is most important that the stations are providing accurate waveforms at present time, in order to properly record future earthquakes. However, it is also important that the waveforms in the past are accurate as well, since these waveforms may be used to improve the current 3D structure model or to simulate past earthquakes. In most of the cases presented below, it would be a relatively simple matter of adjusting the sign within the station response files (i.e., dataless seed files), and then the waveforms would be usable.

I also observed a problem with station amplifications for three stations recording events prior to 2000 (Section E.3). Detecting systematic amplifications is more subtle than detecting the polarity problems. An example of the amplification is shown in Figures E.23–E.25: the relative-low amplitude on all three components at VCS (Figure E.24) is not observed at stations in azimuthal directions on either side of VCS (Figures E.23 and E.25).

E.2 Station–epochs with probable incorrect polarity (Figures E.1–E.22)

- CRP.CI: Figure E.1. From the 3D synthetics, it appears that the seismograms (all three components) are "good", but flipped upsidedown. In Figure E.2, I show the effect of simply flipping the sign of `CONSTANT` in the pole-zero file. A sign flip appears to solve the problem for this station. NOTE: CRP.CI is fine as of 2006.11.03 (10215753).

- HWB.AZ: Figures E.3–E.4. From the 3D synthetics, it appears that the seismograms (all three components) are "good", only flipped upsidedown.

  There is something peculiar about HWB.AZ records. By 2008.12.06 (14408052), HWB.AZ records look great, but the PZ file is the same as before. This suggests that HWB.AZ was changed, but the dataless seed file was not updated.

- BVDA2.AZ: Figures E.5–E.6. From the 3D synthetics, it appears that the seismograms (all three components) are "good", only flipped upsidedown.

- PER.CI: Figures E.7–E.9. From stations MSJ.CI to PER.CI to RVR.CI, I sweep the azimuth in a clockwise manner. The records for MSJ and RVR are similar, but the station in between, BTP, has the polarity flipped for both the transverse (T) and radial (R) components. The earthquake occurred near Hector Mine on 2008.12.06.

- BTP.CI: Figures E.10–E.12. From stations ALP.CI to BTP.CI to OSI.CI, I sweep the azimuth in a clockwise manner. The records for ALP and OSI are similar, but the station in between, BTP, has the polarity flipped for both the transverse (T) and radial (R) components.

- NSS2.CI: Figures E.13–E.15. From stations CTC.CI to NSS2.CI to THX.CI, I sweep the azimuth in a clockwise manner. The records for CTC and THX are similar, but the station in between, NSS2, has the polarity flipped for both the transverse (T) and radial (R) components.

- 109C.TA: Figures E.16–E.18. From stations SDR.CI to 109C.TA to SDG.CI, I sweep the azimuth in a clockwise manner. It appears that something is wrong with the horizontal components for 109C.TA, though it may not be a simply sign error or switch between the E and N components.

- OSI.CI. Figures E.19 and E.22. The pattern for OSI.CI suggests that only the east component has a polarity problem, or that there was some misalignment of the horizontal components. For earthquakes from an easterly direction, the problem is more apparent on the radial component (Figures E.19 and E.20). For earthquakes from a northerly direction, the problem is more apparent on the transverse component (Figures E.21 and E.22).

  There is something peculiar about the 1998 OSI.CI records. All of the problematic records occur in 1998, but the PZ file indicates the same epoch through 2002. This suggests that OSI.CI was changed, but the dataless seed file was not updated. Also, even *between* the earliest (1998-01-05) and latest (1998-10-27) identified problematic records, there are some *good* records: 9064093 (1998-08-16) and 9065468 (1998-08-20).

Detailed list of seismograms exhibiting polarity problems

Figures E.1 and E.2 show one example for one station (CRP.CI). I will now list all the paths for which the polarity on **all three components** appears to be flipped:

1. **CRP.CI**. 9968977 9983429 10059745 10097009 10100053 10148421 14073800 14077668 14095540 14095628 14096196 14116972 14138080 14151344 14155260 14165408 14169456 14178236 14178248 14186612 14236768

2. **HWB.AZ**. 9967901 10100053 10215753 13966396 14095628 14096196 14151344 14155260 14178184 14178188 14178212 14178236 14178248 14179292 14179736 14236768 14263712 14263716 14383980

3. **BVDA2.AZ**. 9967901 10215753 10230869 13966396 14095540 14095628 14169456 14178184 14178188 14178212 14178236 14178248 14179288 14179292 14179736 14186612 14233632 14236768 14263544 14263712 14263716

Figures E.7–E.9 shows one example for one station (PER.CI). I will now list all the paths for which the polarity on the **horizontal components only** appears to be a problem:

1. **BTP.CI**. 9854597 9882325 9882329 13935988 13936812 13938812 13945908

2. **PER.CI**. 9967901 9968977 9983429 10006857 10059745 10063349 10097009 10100053 10148421 10215753 10230869 10370141 14007388 14072464 14073800 14077668 14095540 14095628 14096196 14116972 14118096 14133048 14138080 14151344 14158696 14169456 14178184 14178188 14178212 14178236 14178248 14179288 14179292 14179736 14186612 14236768 14239184 14263544 14383980 14408052 10370141 14418600

3. **NSS2.CI**. 10059745 10097009 14095628 14138080 14151344 14155260 14178184 14178188 14178248 14179292 14179736

4. **109C.TA**. 10059745 10097009 10100053 10148421 14073800 14095628 14116972 14118096 14138080 14151344 14155260 14169456 14178184 14178188 14178212 14178236 14178248 14179288 14179292 14179736 14186612

5. **OSI.CI (BHE only?)**. 3298292 (BHR) 9038699 (BHR) 9069997 (BHR) 9044650 (BHT) 9045109 (BHT) 9045697 (BHT)

Next I list all the events above in order of increasing origin time, with each problematic station listed on the following line(s). Note that these records are only the ones that I have identified directly as having a problem. I expect that records at the same stations during the same epochs would exhibit the **polarity problem** as well.

```
 9038699 1998-01-05   18:14:06  Mw 3.9   -117.7178     33.9462  12.98 km
         OSI.CI-E
 9044650 1998-03-06   07:36:35  Mw 4.0   -117.6505     36.0737   7.93 km
         OSI.CI-E
 9045109 1998-03-07   00:36:46  Mw 4.5   -117.6200     36.0912   6.99 km
         OSI.CI-E
 9045697 1998-03-08   15:28:41  Mw 3.7   -117.6133     36.0827   4.81 km
         OSI.CI-E
 3298292 1998-03-11   12:18:51  Mw 4.2   -117.2222     34.0355  16.19 km
         OSI.CI-E
 9069997 1998-10-27   01:08:40  Mw 4.4   -116.8418     34.3208   6.02 km
         OSI.CI-E
 9854597 2002-10-29   14:16:54  Mw 4.4   -116.2650     34.8068   7.89 km
        BTP.CI-EN
 9882325 2003-01-25   09:11:02  Mw 3.9   -118.6632     35.3152   4.41 km
        BTP.CI-EN
 9882329 2003-01-25   09:16:10  Mw 4.2   -118.6585     35.3128   4.12 km
        BTP.CI-EN
13935988 2003-02-22   12:19:10  Mw 4.8   -116.8460     34.3103   4.55 km
        BTP.CI-EN
13936812 2003-02-22   19:33:45  Mw 4.2   -116.8482     34.3097   4.87 km
        BTP.CI-EN
13938812 2003-02-25   04:03:04  Mw 4.0   -116.8407     34.3137   3.84 km
        BTP.CI-EN
13945908 2003-03-11   19:28:17  Mw 4.2   -116.1303     34.3582   8.08 km
        BTP.CI-EN
13966396 2003-05-24   02:04:28  Mw 4.1   -115.5538     32.9475   8.72 km
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
14007388 2003-12-04   06:15:52  Mw 3.5   -117.5664     35.6352   2.13 km
        PER.CI-EN
 9967901 2003-12-23   18:17:11  Mw 4.5   -121.0428     35.6493   7.20 km
        PER.CI-EN
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
 9968977 2003-12-25   11:50:01  Mw 4.3   -120.8385     35.5487   8.18 km
        PER.CI-EN
       CRP.CI-ZEN
 9983429 2004-02-14   12:43:11  Mw 4.5   -119.1412     35.0118  11.81 km
        PER.CI-EN
       CRP.CI-ZEN
10006857 2004-05-09   08:57:17  Mw 4.2   -120.0142     34.4135  10.97 km
        PER.CI-EN
14072464 2004-07-09   04:43:45  Mw 3.7   -115.7441     32.5392  10.37 km
        PER.CI-EN
14073800 2004-07-14   00:53:52  Mw 3.8   -116.0520     33.7152  12.20 km
        PER.CI-EN
       109C.TA-EN
       CRP.CI-ZEN
14077668 2004-07-24   12:55:19  Mw 4.0   -119.4365     34.3885   8.66 km
        PER.CI-EN
       CRP.CI-ZEN
14095540 2004-09-29   17:10:04  Mw 4.8   -120.5134     35.9528  10.69 km
        PER.CI-EN
       CRP.CI-ZEN
     BVDA2.AZ-ZEN
14095628 2004-09-29   22:54:54  Mw 4.8   -118.6292     35.3852   7.66 km
       NSS2.CI-EN
        PER.CI-EN
       109C.TA-EN
       CRP.CI-ZEN
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
14096196 2004-09-30   18:54:29  Mw 4.6   -120.5403     35.9821   9.87 km
        PER.CI-EN
       CRP.CI-ZEN
       HWB.AZ-ZEN
10059745 2004-11-13   17:39:16  Mw 3.8   -116.8413     34.3533  10.31 km
       NSS2.CI-EN
        PER.CI-EN
       109C.TA-EN
       CRP.CI-ZEN
10063349 2004-11-29   01:54:14  Mw 4.0   -120.4963     35.9437  10.19 km
        PER.CI-EN
14116972 2005-01-06   14:35:27  Mw 4.1   -117.4438     34.1272   5.04 km
        PER.CI-EN
       109C.TA-EN
       CRP.CI-ZEN
14118096 2005-01-12   08:10:46  Mw 3.9   -116.3912     33.9578   8.51 km
        PER.CI-EN
       109C.TA-EN
```

```
14133048 2005-03-22   08:55:05  Mw 3.6   -116.2515     33.2884   4.74 km
       PER.CI-EN
14138080 2005-04-16   19:18:13  Mw 4.6   -119.1940     34.9987  10.16 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
10097009 2005-05-06   02:29:09  Mw 4.0   -119.1958     35.0023  13.01 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
10100053 2005-05-16   07:24:37  Mw 4.2   -120.4792     35.9269   9.15 km
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
      HWB.AZ-ZEN
14151344 2005-06-12   15:41:46  Mw 5.1   -116.5675     33.5380  13.91 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
      HWB.AZ-ZEN
14155260 2005-06-16   20:53:25  Mw 4.8   -117.0072     34.0612  14.19 km
      NSS2.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
      HWB.AZ-ZEN
14158696 2005-06-27   22:17:33  Mw 3.6   -117.0232     34.0615  13.63 km
       PER.CI-EN
14165408 2005-07-24   12:59:42  Mw 3.8   -119.7527     33.6853   3.85 km
      CRP.CI-ZEN
14169456 2005-08-06   05:40:33  Mw 3.9   -118.0652     36.1488   3.67 km
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
    BVDA2.AZ-ZEN
14178184 2005-08-31   22:47:45  Mw 4.7   -115.6207     33.1544   4.50 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14178188 2005-08-31   22:50:24  Mw 4.4   -115.6098     33.1639   1.59 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14178212 2005-08-31   23:07:16  Mw 4.3   -115.6157     33.1548   5.01 km
       PER.CI-EN
      109C.TA-EN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14178236 2005-08-31   23:27:32  Mw 4.1   -115.5924     33.1748   3.95 km
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14178248 2005-08-31   23:32:11  Mw 4.3   -115.5969     33.1712   5.05 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14179288 2005-09-01   13:48:25  Mw 3.8   -115.6168     33.1538   4.74 km
       PER.CI-EN
      109C.TA-EN
    BVDA2.AZ-ZEN
14179292 2005-09-01   13:50:20  Mw 4.4   -115.6064     33.1643   2.63 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14179736 2005-09-02   01:27:20  Mw 5.0   -115.6295     33.1479   4.90 km
      NSS2.CI-EN
       PER.CI-EN
      109C.TA-EN
      HWB.AZ-ZEN
    BVDA2.AZ-ZEN
14186612 2005-09-22   20:24:48  Mw 4.4   -119.0247     35.0178  10.24 km
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
    BVDA2.AZ-ZEN
10148421 2005-10-18   07:31:03  Mw 4.1   -116.7715     34.0182  18.32 km
       PER.CI-EN
      109C.TA-EN
      CRP.CI-ZEN
```

```
14236768 2006-06-30   00:28:06  Mw 4.1   -116.0220     33.2450   3.84 km
        PER.CI-EN
       CRP.CI-EN
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
14239184 2006-07-10   02:54:43  Mw 3.7   -117.1103     33.8567  16.77 km
        PER.CI-EN
10215753 2006-11-03   15:56:43  Mw 4.1   -116.0520     32.7165  14.76 km
        PER.CI-EN
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
14263544 2006-11-29   12:17:35  Mw 3.7   -115.9628     32.8423   3.37 km
        PER.CI-EN
     BVDA2.AZ-ZEN
14263712 2006-11-29   21:10:55  Mw 4.0   -115.9672     32.8385   7.16 km
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
14263716 2006-11-29   21:12:52  Mw 3.6   -115.9672     32.8377   3.43 km
       HWB.AZ-ZEN
     BVDA2.AZ-ZEN
10230869 2007-02-09   03:33:43  Mw 3.9   -116.1357     33.2220  20.99 km
        PER.CI-EN
     BVDA2.AZ-ZEN
14383980 2008-07-29   18:42:16  Mw 5.4   -117.7610     33.9530  14.23 km
        PER.CI-EN
       HWB.AZ-ZEN
14408052 2008-12-06   04:18:43  Mw 5.0   -116.4190     34.8130   6.10 km
        PER.CI-EN
10370141 2009-01-09   03:49:46  Mw 4.4   -117.3040     34.1070  14.20 km
        PER.CI-EN
14418600 2009-01-31   21:09:22  Mw 3.9   -117.7860     35.4130   8.50 km
        PER.CI-EN
```

E.3 Station–epochs with probable incorrect amplification (Figures E.23–E.29)

- VCS.CI: Figures E.23–E.25. The relative-low amplitude on all three components at VCS (Figure E.24) is not observed at stations in azimuthal directions on either side of VCS (Figures E.23 and E.25).
- SMTC.AZ: Figures E.26 and E.27. The relative-low amplitude for BHZ and BHT at SMTC.AZ (Figure E.24) is not observed at adjacent station SWS.CI (Figure E.26).
- BAR.CI: Figures E.28 and E.29. It is possible that for BAR.CI the amplification problem is only with the east component. This inference is based on the fact that the amplification is observed primarily on the transverse component for north-south paths (Figures E.28 and E.29), and it is observed primarily on the radial component for east-west paths (9075803, 9154092).

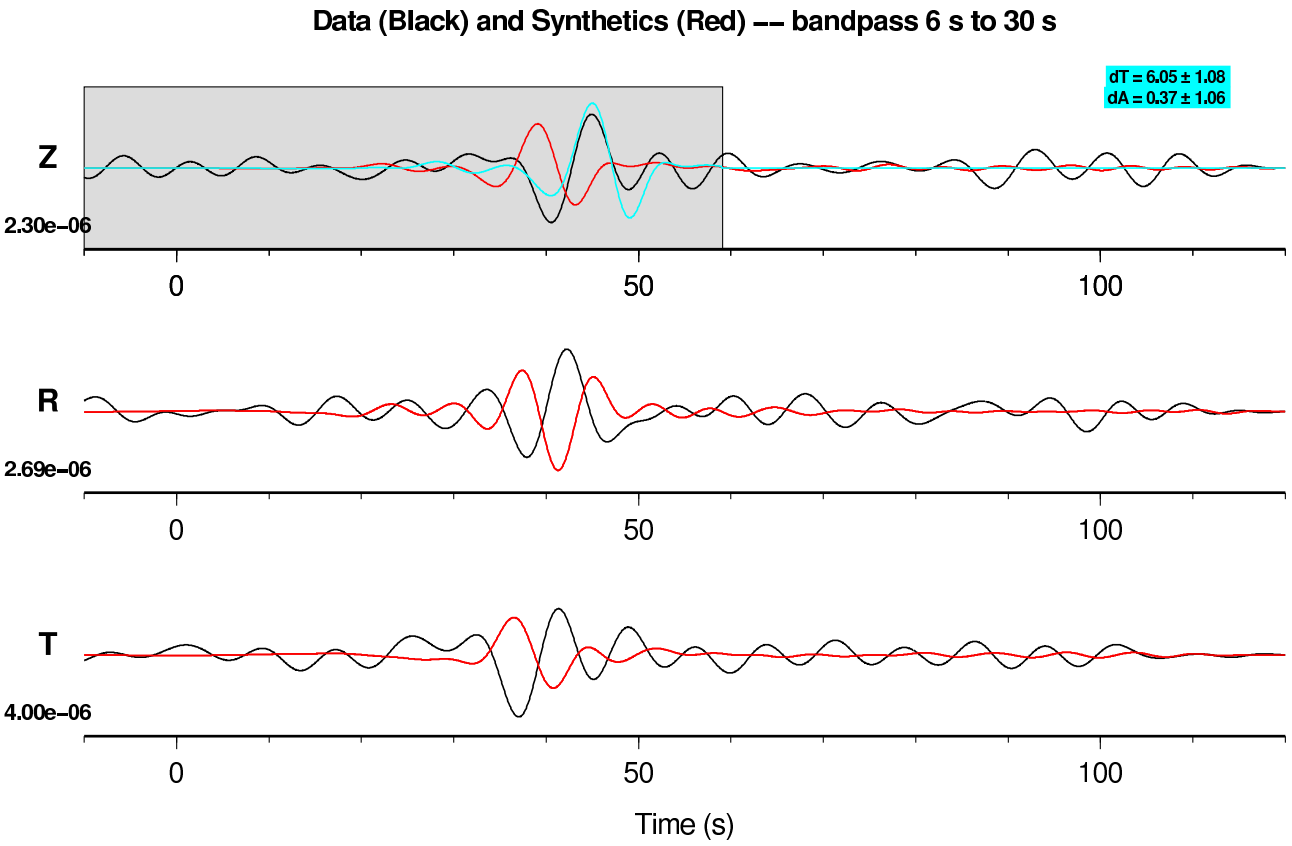
Detailed list of seismograms exhibiting amplification problems

- VCS.CI: 7112721 9044494 9044650 9045109 9064093 9064568 9069997 9070083
- SMTC.AZ: 3317364 9075803 9086693

- BAR.CI: 9075803 (R) 9154092 (R) 3320736 (T) 3321426 (T) 3321590 (T) 9085734 (T) 9109287 (T) 9109442 (T) 9109636 (T) 9110685 (T) 9112735 (T) 9114763 (T) 9114812 (T) 9114858 (T) 9117942 (T) 9119414 (T) 9140050 (T)

Next I list all the events above in order of increasing origin time, with each problematic station listed on the following line(s). Note that these records are only the ones that I have identified directly as having a problem. I expect that records at the same stations during the same epochs would exhibit the **amplification problem** as well.

```
9044494 1998-03-06   05:47:40  Mw 4.9   -117.6405     36.0778   7.17 km
      VCS.CI-ZEN
9044650 1998-03-06   07:36:35  Mw 4.0   -117.6505     36.0737   7.93 km
      VCS.CI-ZEN
9045109 1998-03-07   00:36:46  Mw 4.5   -117.6200     36.0912   6.99 km
      VCS.CI-ZEN
9064093 1998-08-16   13:34:40  Mw 4.4   -116.9232     34.1245   5.98 km
      VCS.CI-ZEN
9064568 1998-08-20   23:49:58  Mw 4.1   -117.6502     34.3737   9.51 km
      VCS.CI-ZEN
7112721 1998-10-01   18:18:15  Mw 4.2   -116.9158     34.1155   5.40 km
      VCS.CI-ZEN
9069997 1998-10-27   01:08:40  Mw 4.4   -116.8418     34.3208   6.02 km
      VCS.CI-ZEN
9070083 1998-10-27   15:40:16  Mw 3.8   -116.8455     34.3202   5.48 km
      VCS.CI-ZEN
9075803 1999-01-13   13:20:56  Mw 4.2   -115.9248     32.7190   8.00 km
      SMT.AZ-ZEN
        BAR.CI-E
9085734 1999-05-05   02:17:46  Mw 3.6   -116.3697     34.0725   2.58 km
        BAR.CI-E
9086693 1999-05-14   08:22:07  Mw 3.9   -116.3623     34.0375   3.98 km
      SMT.AZ-ZEN
3317364 1999-05-14   10:52:35  Mw 4.1   -116.3582     34.0378   4.01 km
      SMT.AZ-ZEN
3320736 1999-10-16   12:57:21  Mw 5.3   -116.2465     34.4368   7.96 km
        BAR.CI-E
9109287 1999-10-16   18:01:57  Mw 4.0   -116.3013     34.7087   6.65 km
        BAR.CI-E
9109442 1999-10-16   20:13:37  Mw 4.3   -116.2793     34.6940   3.18 km
        BAR.CI-E
9109636 1999-10-16   22:53:41  Mw 4.1   -116.3570     34.7097   9.52 km
        BAR.CI-E
9110685 1999-10-17   16:22:48  Mw 4.1   -116.1375     34.3465   3.96 km
        BAR.CI-E
9112735 1999-10-19   12:20:44  Mw 4.0   -116.3442     34.7110   9.34 km
        BAR.CI-E
3321590 1999-10-21   01:54:34  Mw 4.8   -116.3955     34.8735   3.33 km
        BAR.CI-E
9114763 1999-10-22   12:40:52  Mw 3.7   -116.2085     34.3300  11.53 km
        BAR.CI-E
9114812 1999-10-22   16:08:48  Mw 4.8   -116.4060     34.8620   3.02 km
        BAR.CI-E
9114858 1999-10-22   16:48:22  Mw 3.8   -116.3820     34.8292   5.15 km
        BAR.CI-E
9117942 1999-10-29   12:36:37  Mw 4.0   -116.2707     34.5200   2.90 km
        BAR.CI-E
3321426 1999-11-03   02:55:05  Mw 3.6   -116.2888     34.8031   6.06 km
        BAR.CI-E
9119414 1999-11-03   03:27:57  Mw 3.9   -116.3570     34.8470   5.90 km
        BAR.CI-E
9140050 2000-02-21   13:49:43  Mw 4.1   -117.2432     34.0588  16.34 km
        BAR.CI-E
9154092 2000-06-14   19:00:20  Mw 4.3   -115.5035     32.8898   8.73 km
        BAR.CI-E
```



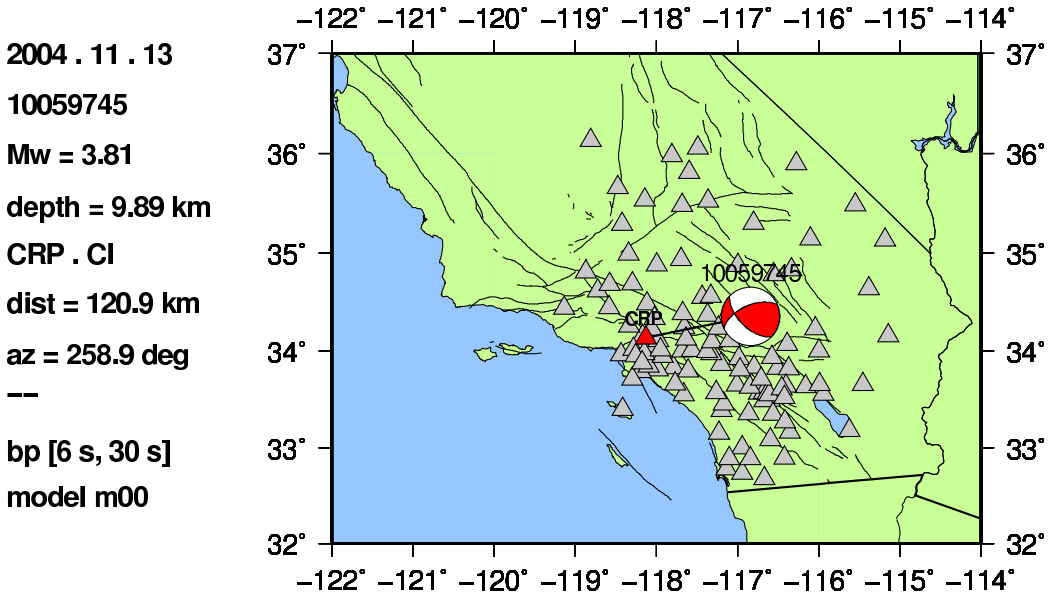


Figure E.1: Data (black) and synthetics (red), 6–30 s, from 10059745 to CRP.CI. The measurement algorithm selects a large time shift for the Rayleigh wave, but this is due to the station error, not to the source or structure. Compare with Figure E.2.

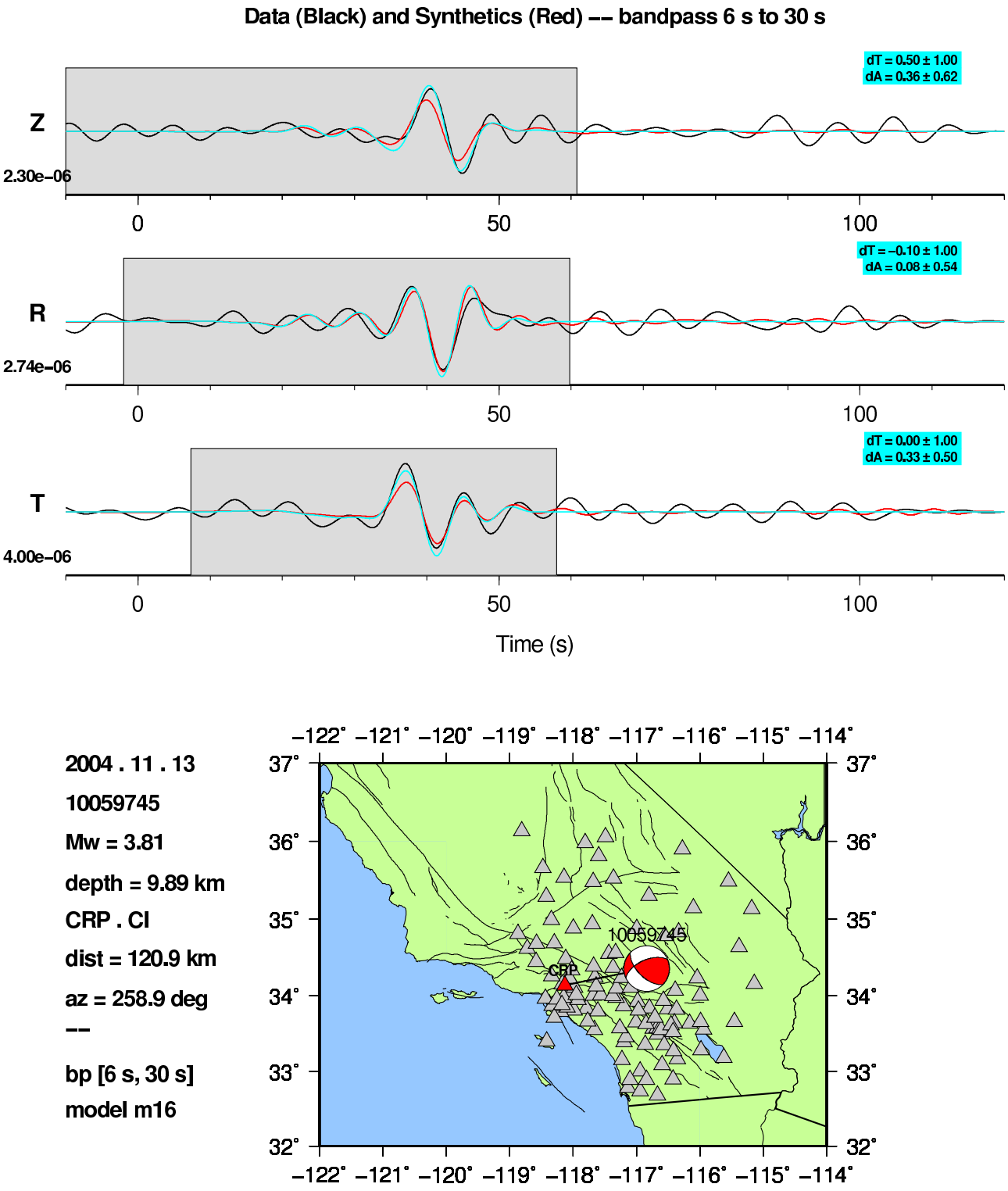
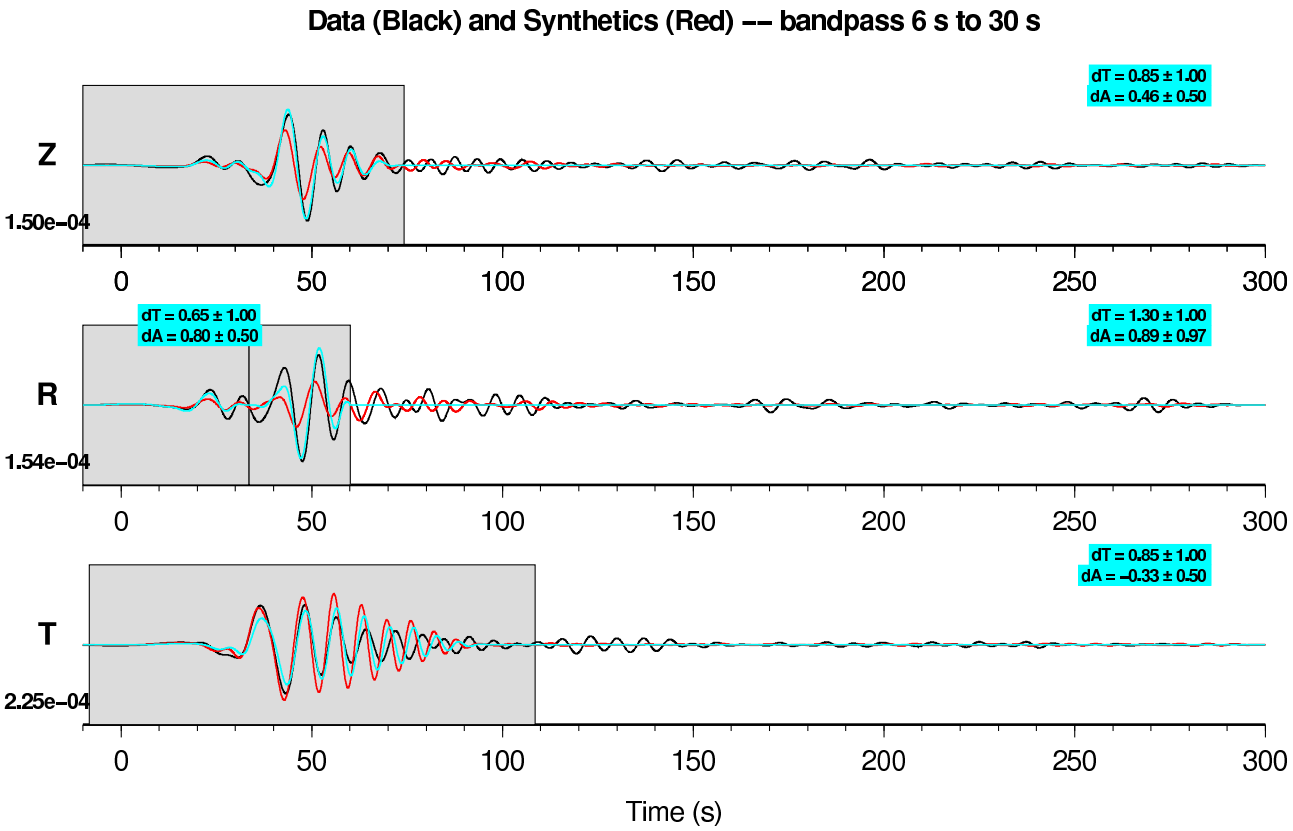


Figure E.2: Data (black) and synthetics (red), 6–30 s, from 10059745 to CRP.CI. In this case, I have flipped the sign of the constant value in the pole-zero file. Compare with Figure E.1.



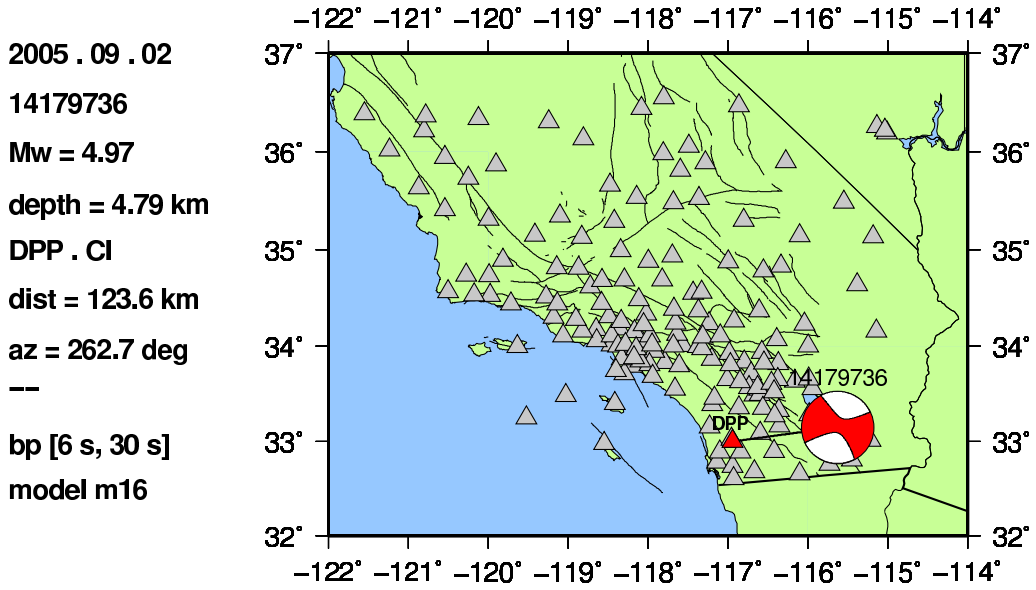
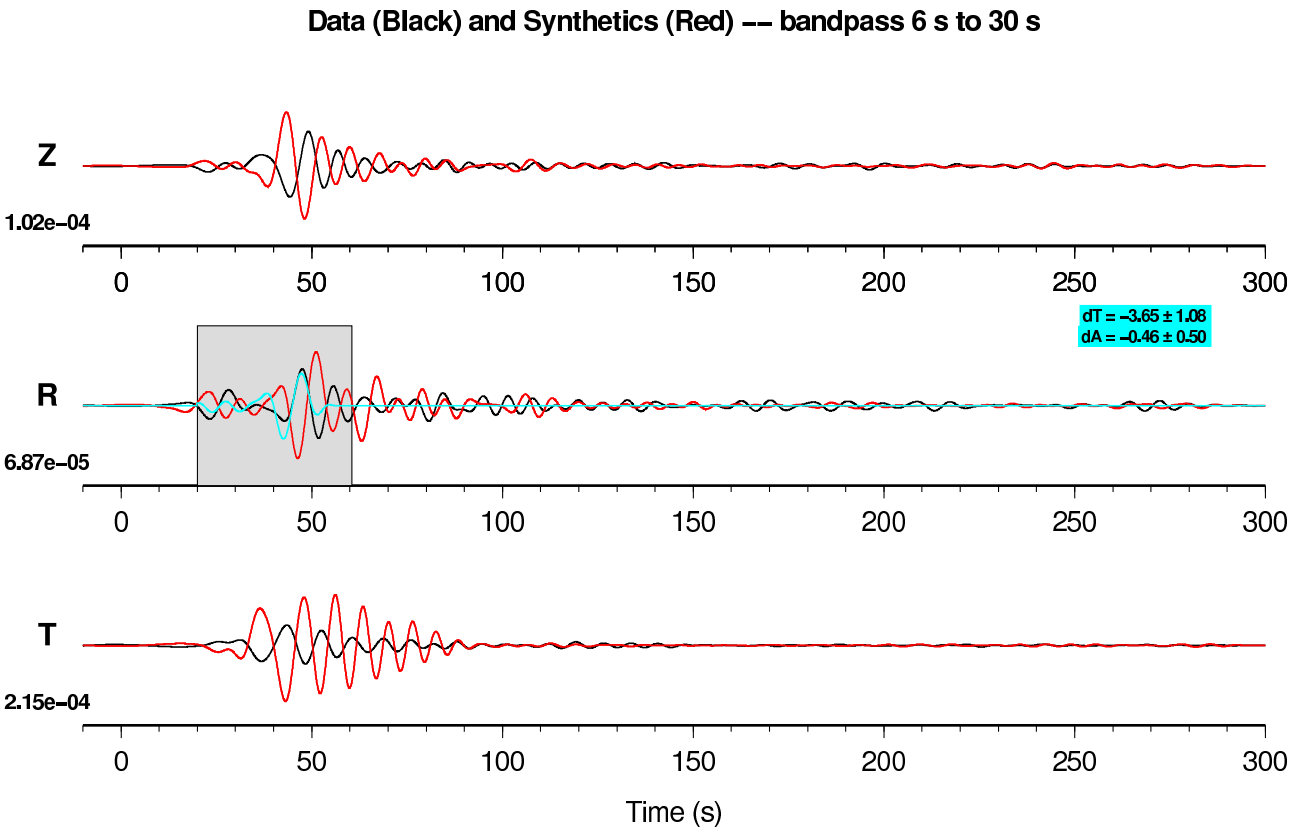


Figure E.3: Data (black) and synthetics (red), 6–30 s, from 14179736 to DPP.CI. Compare with Figure E.4.



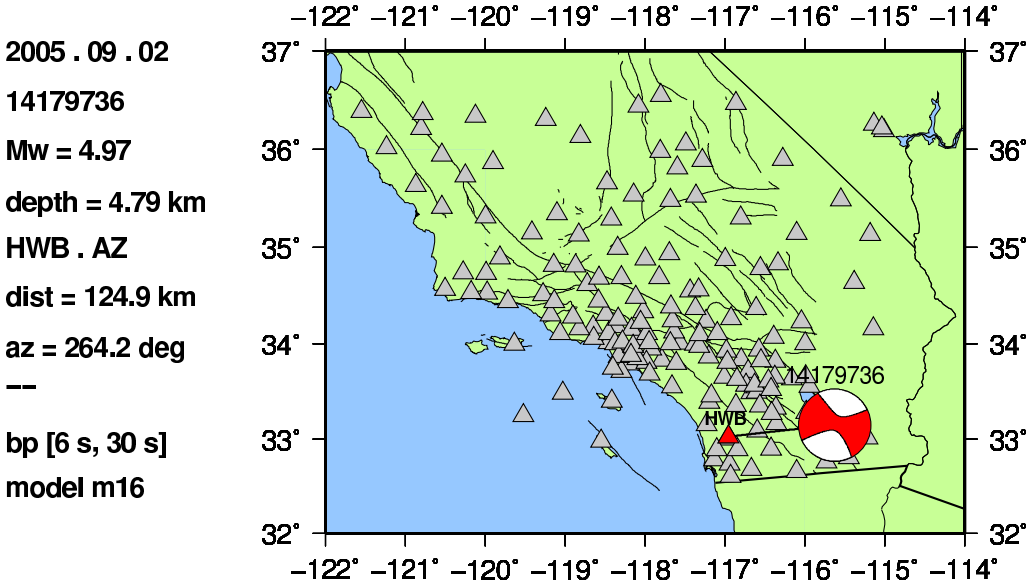
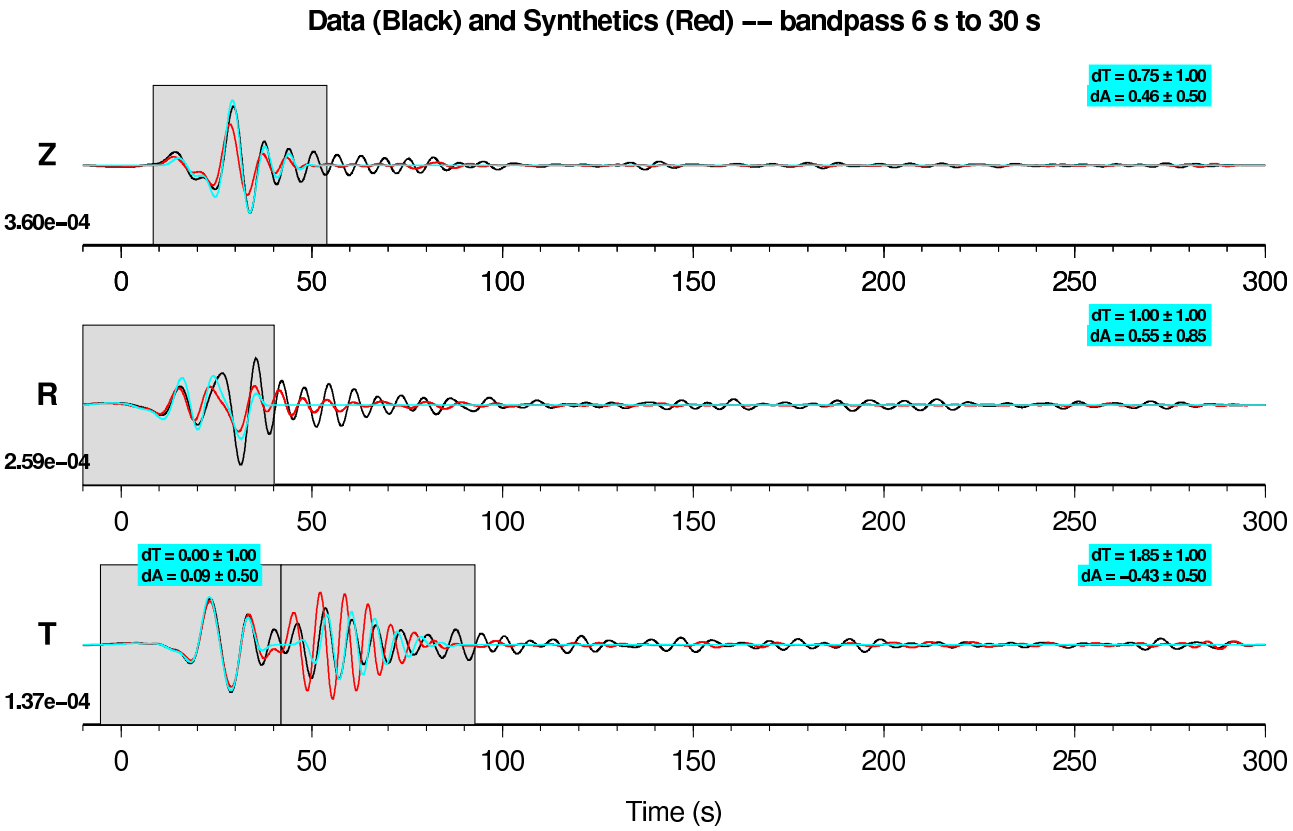


Figure E.4: Data (black) and synthetics (red), 6–30 s, from 14179736 to HWB.AZ. The measurement algorithm selects a large time shift for the radial-component Rayleigh wave, but this is due to the station error, not to the source or structure. Compare with Figure E.3.



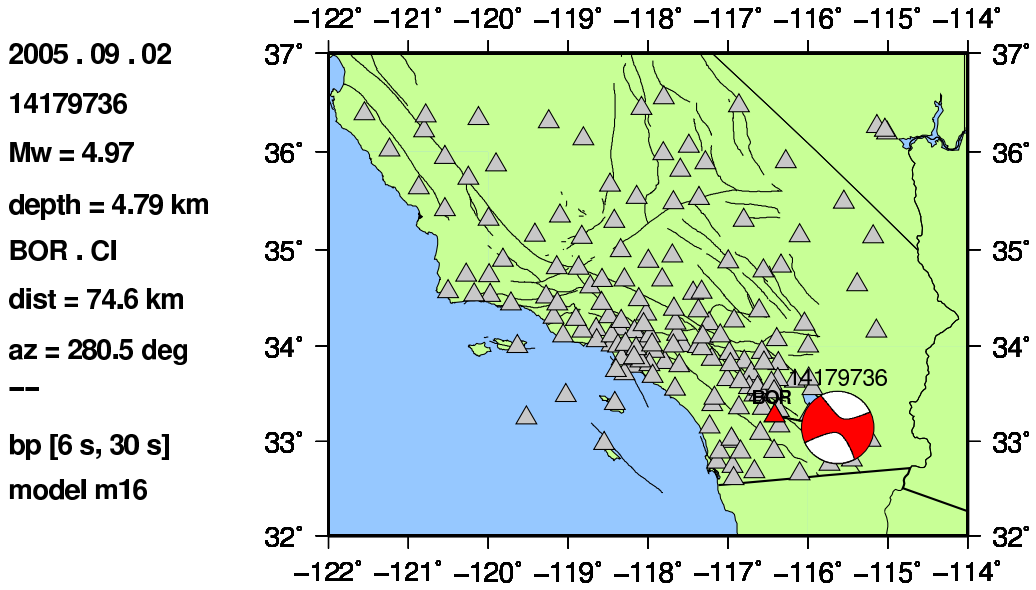
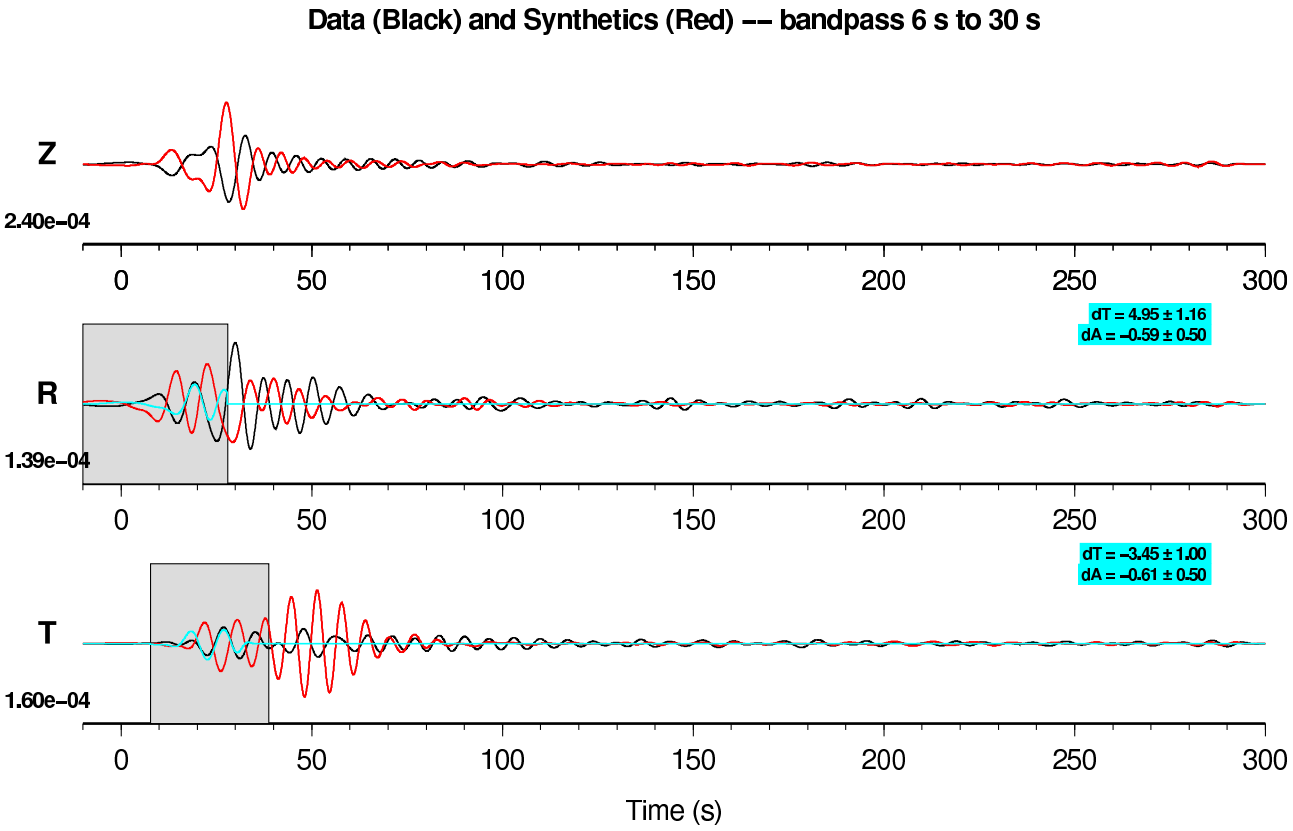


Figure E.5: Data (black) and synthetics (red), 6–30 s, from 14179736 to BOR.CI. Compare with Figure E.6.



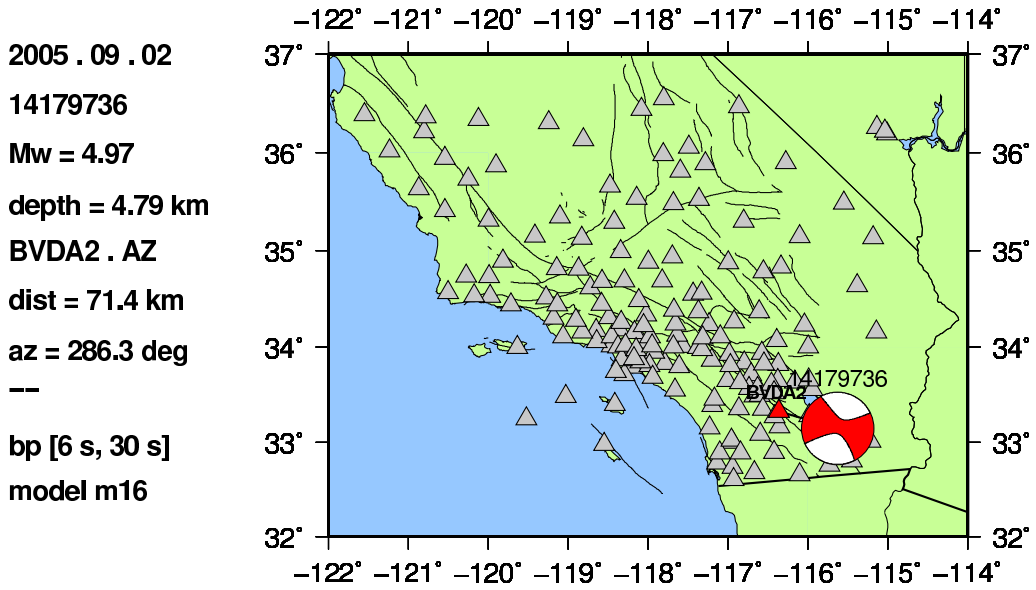
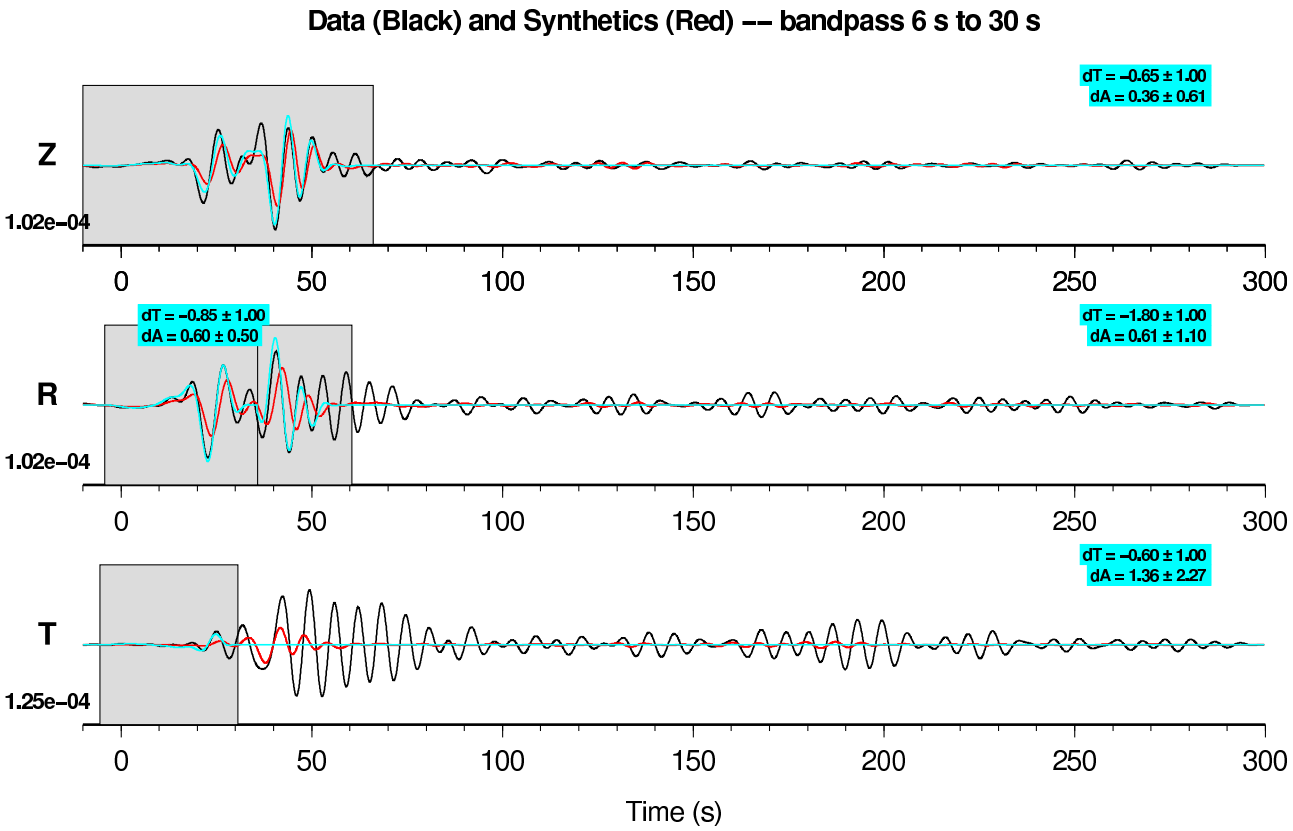


Figure E.6: Data (black) and synthetics (red), 6–30 s, from 14179736 to BVDA2.AZ. The measurement algorithm mistakenly selects the large time shifts for the Rayleigh wave, but this is due to the station error, not to the source or structure. Compare with Figure E.5.



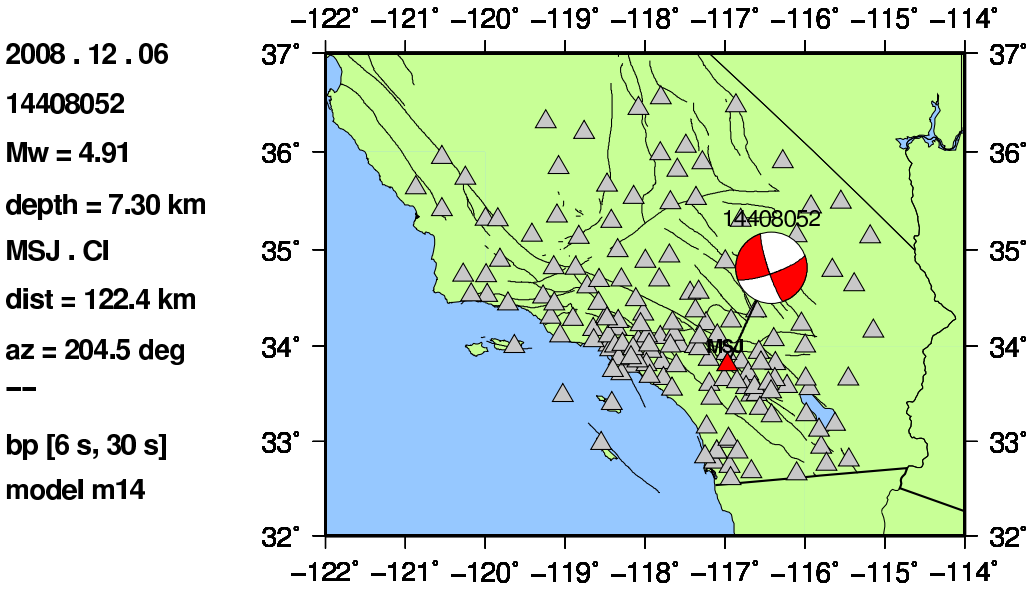
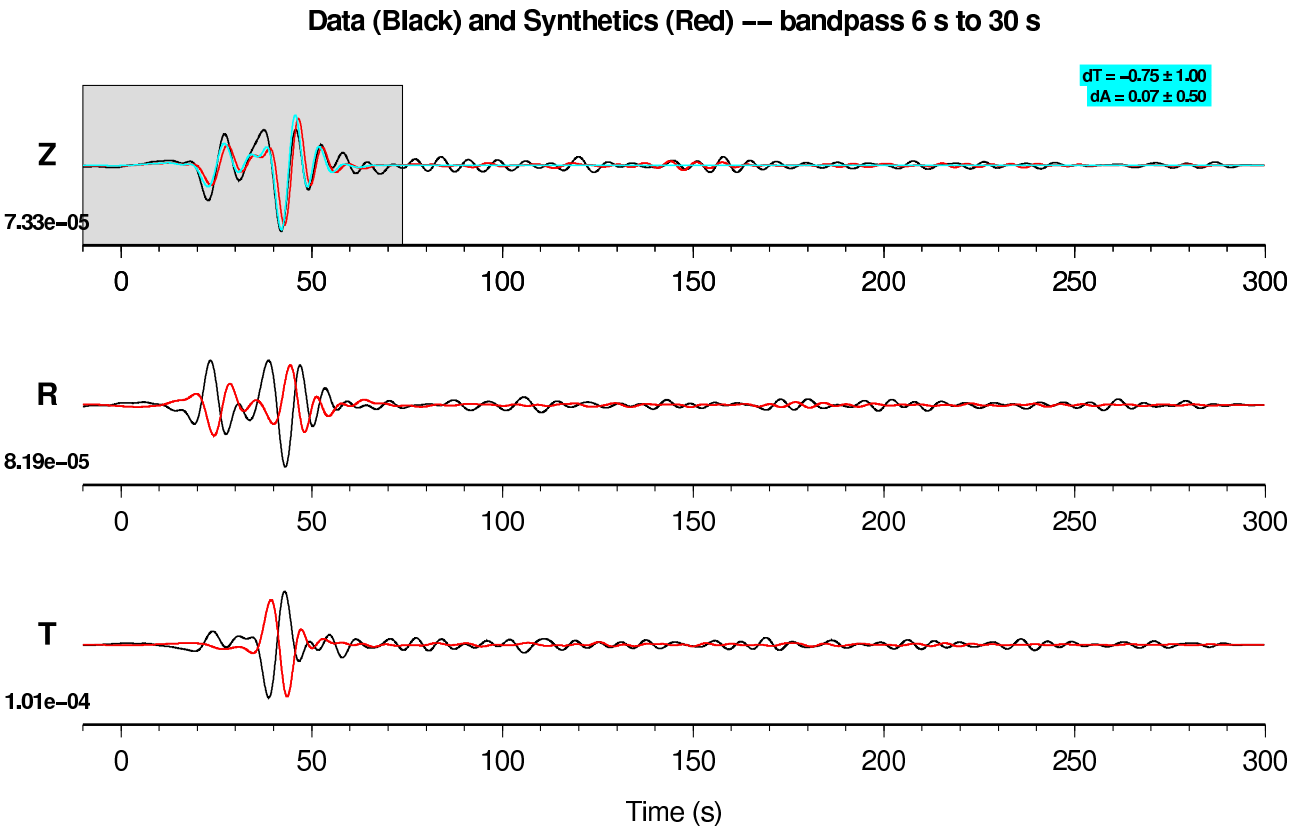


Figure E.7: Data (black) and synthetics (red), 6–30 s, from 14408052 to MSJ.CI. Compare with Figure E.8.



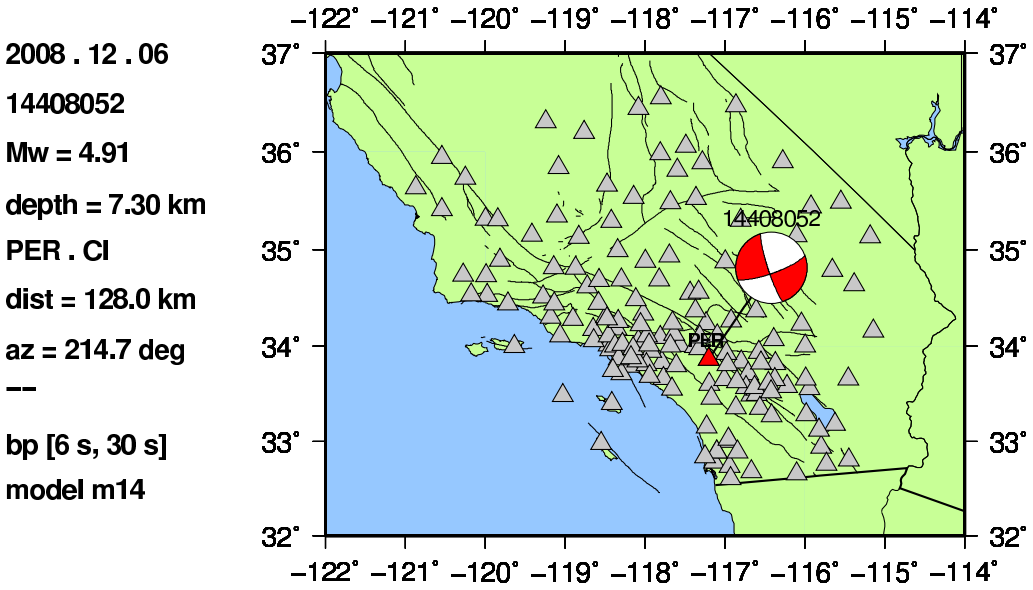
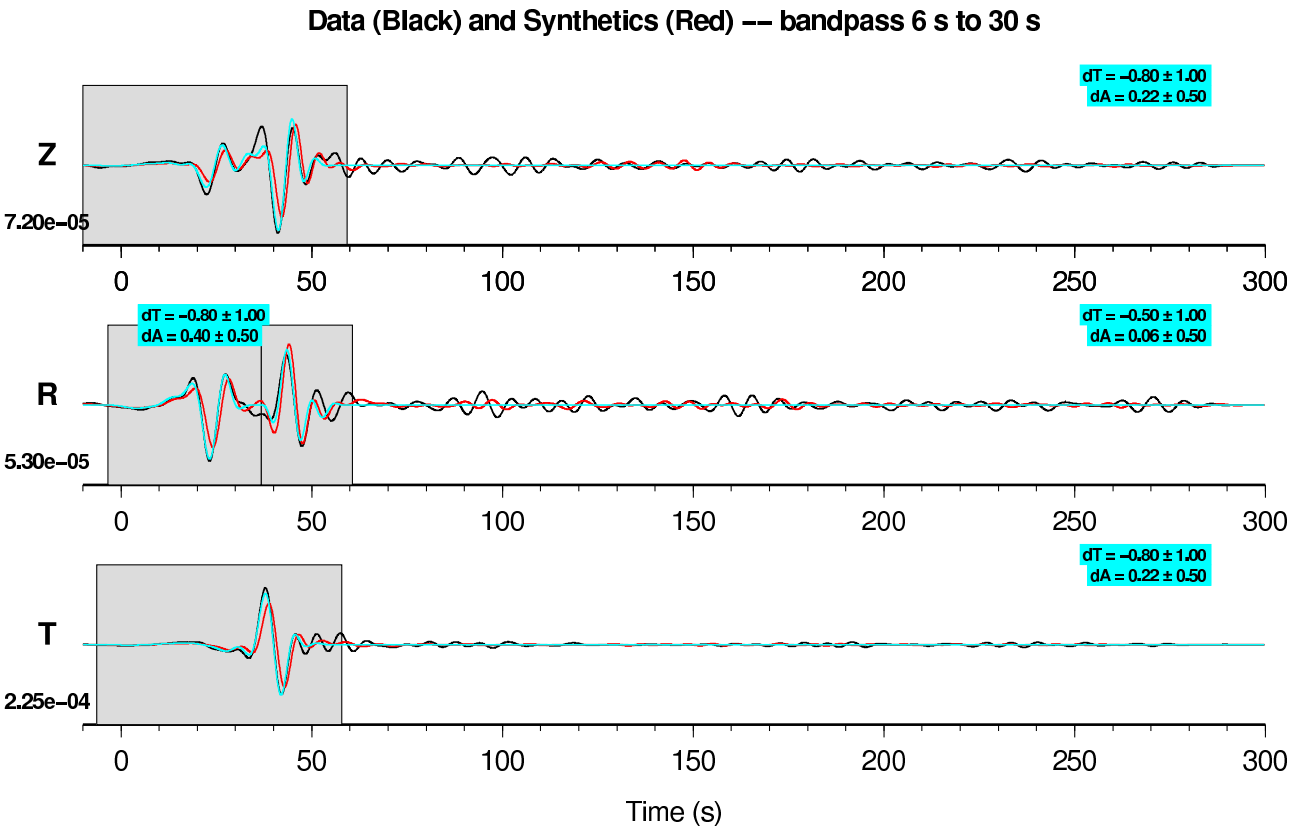


Figure E.8: Data (black) and synthetics (red), 6–30 s, from 14408052 to PER.CI. Compare with Figures E.7 and E.9.



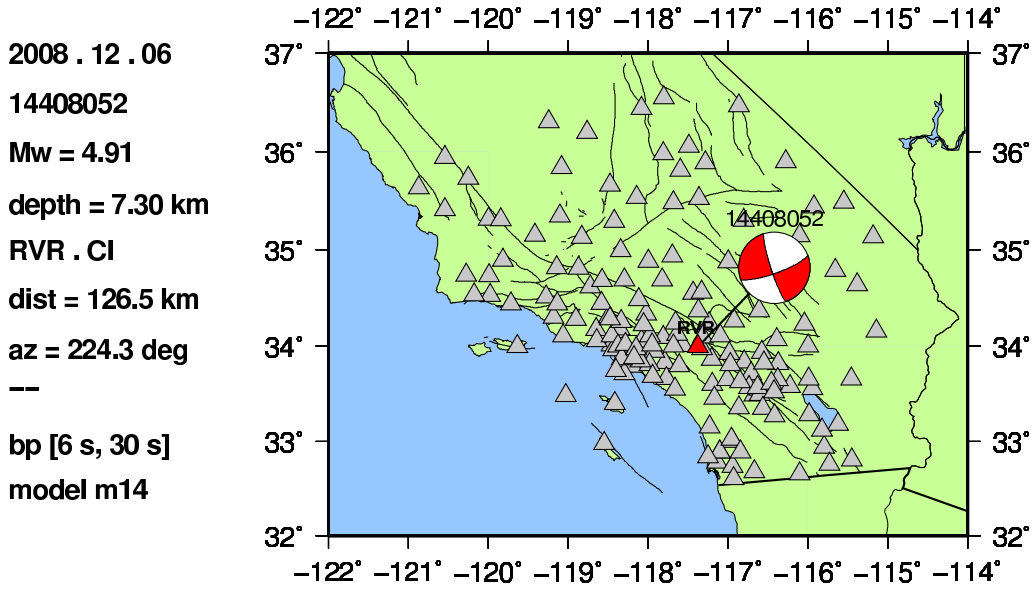
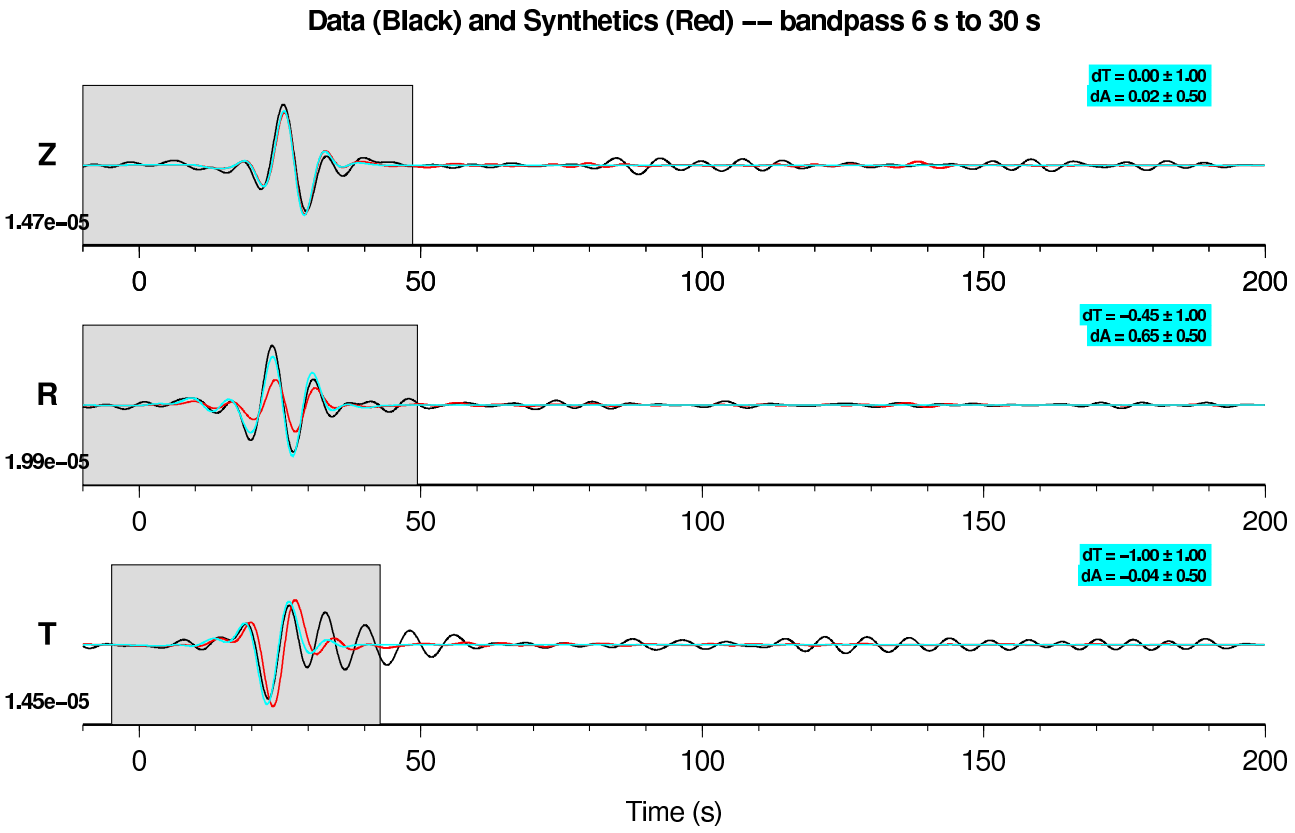


Figure E.9: Data (black) and synthetics (red), 6–30 s, from 14408052 to RVR.CI. Compare with Figure E.8.



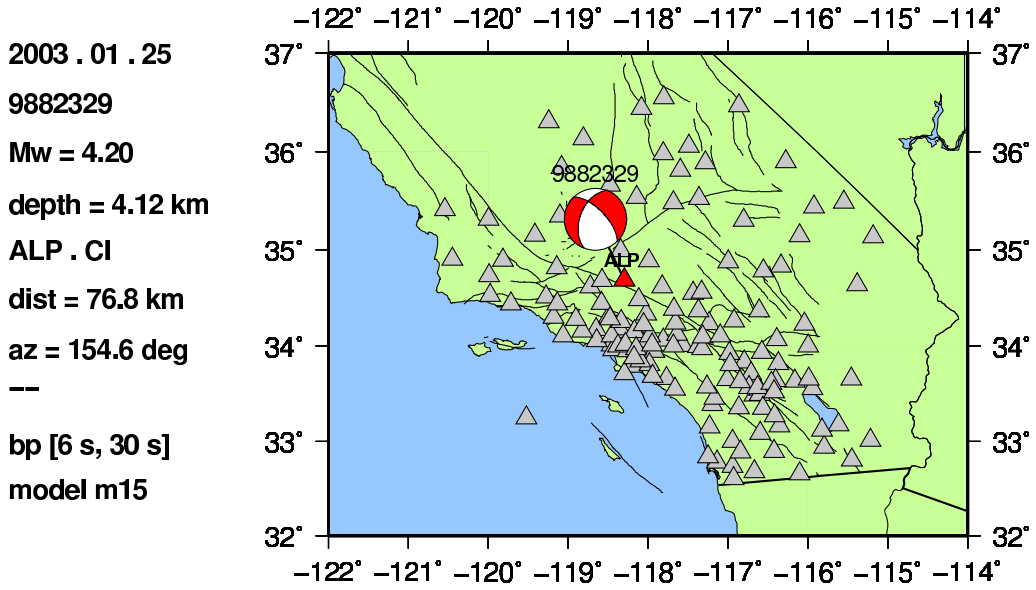
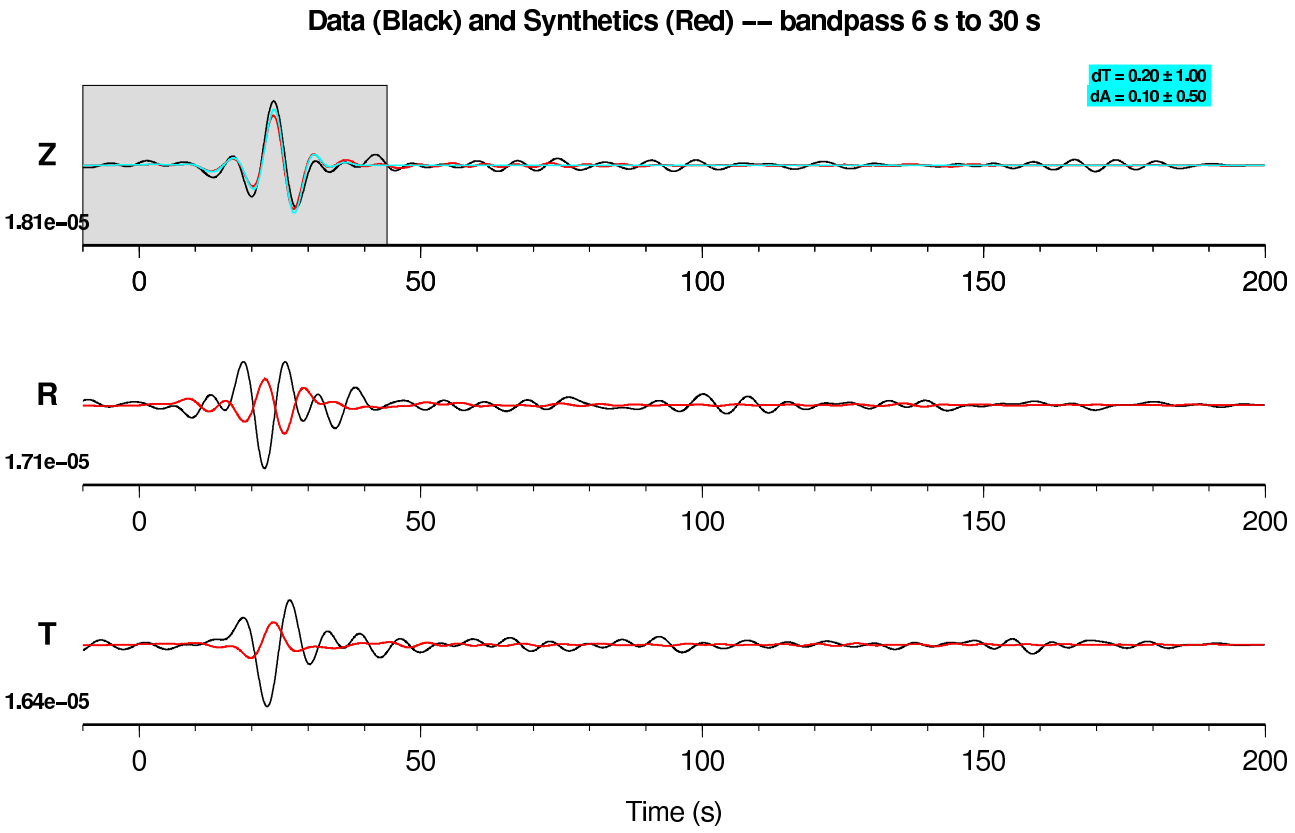


Figure E.10: Data (black) and synthetics (red), 6–30 s, from 14138080 to ALP.CI. Compare with Figure E.11.



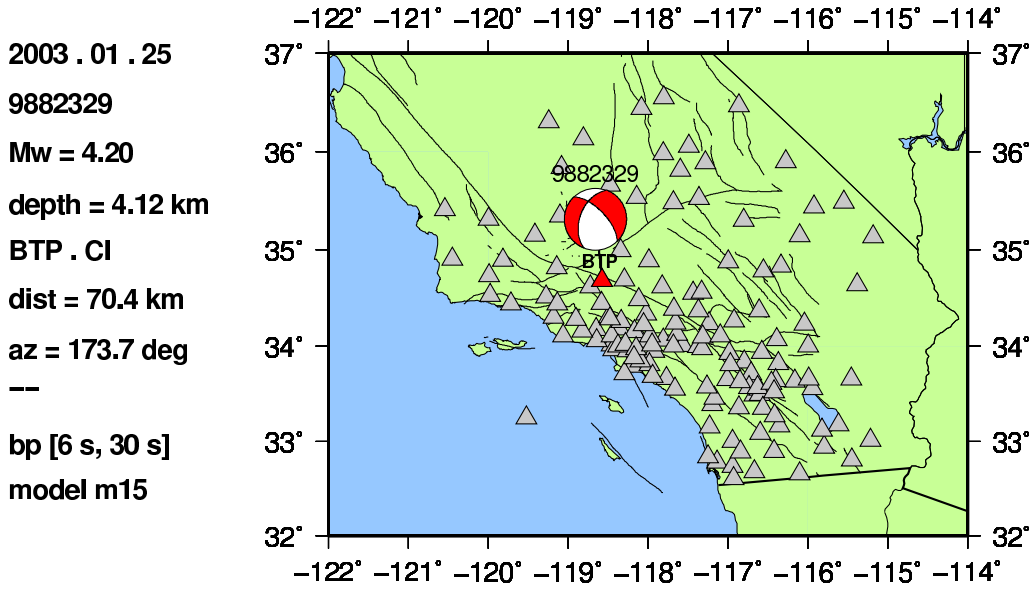
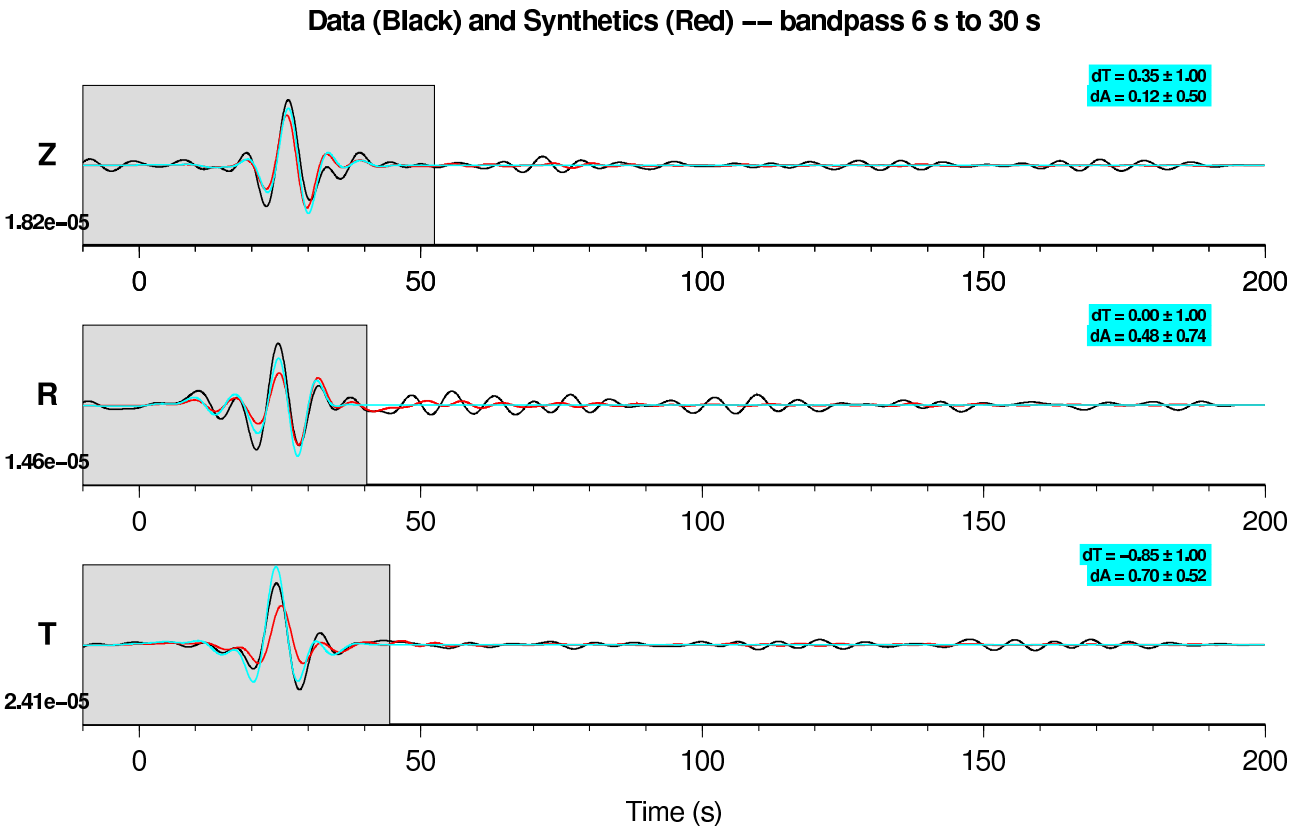


Figure E.11: Data (black) and synthetics (red), 6–30 s, from 14138080 to BTP.CI. Compare with Figures E.10 and E.12.



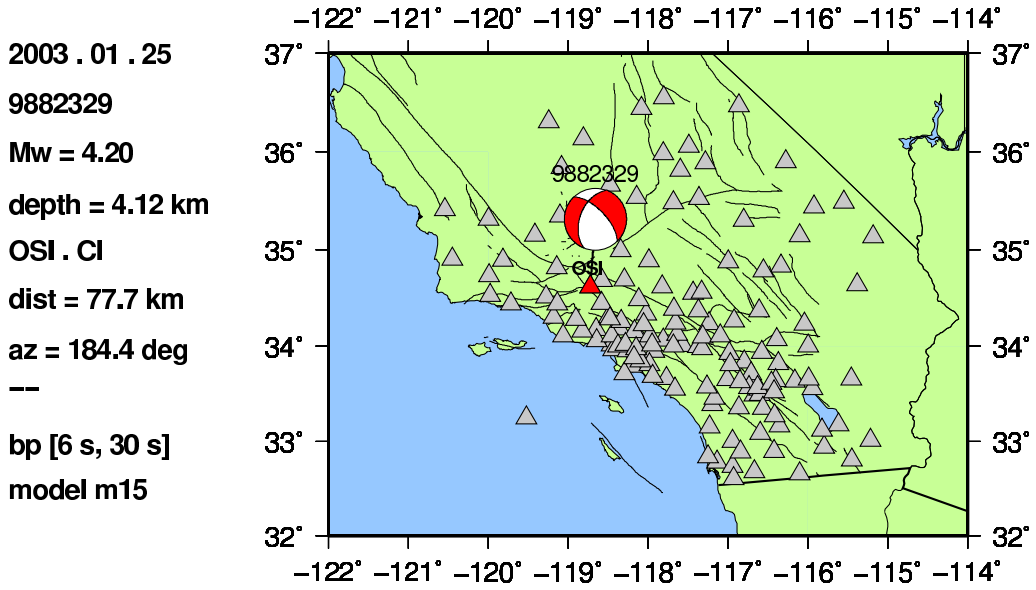
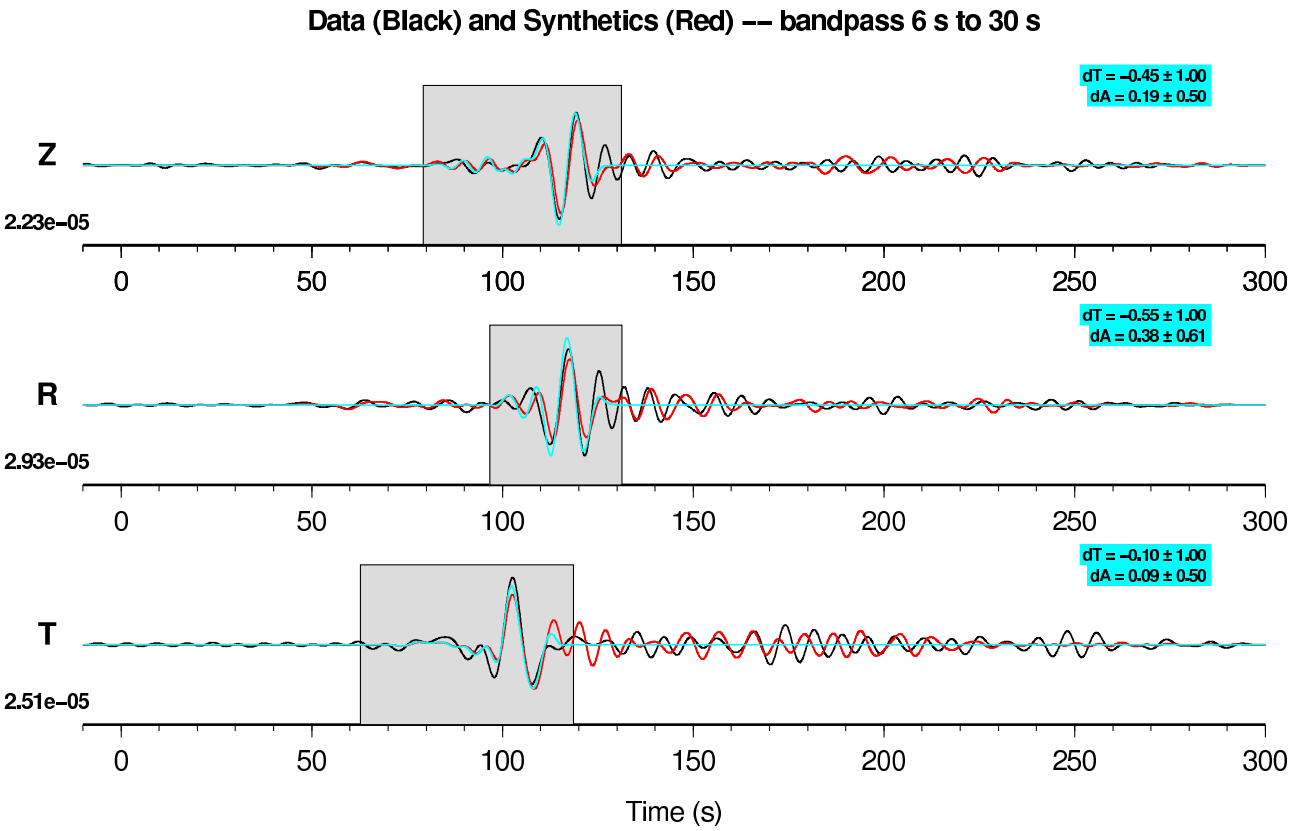


Figure E.12: Data (black) and synthetics (red), 6–30 s, from 14138080 to OSI.CI. Compare with Figure E.11.



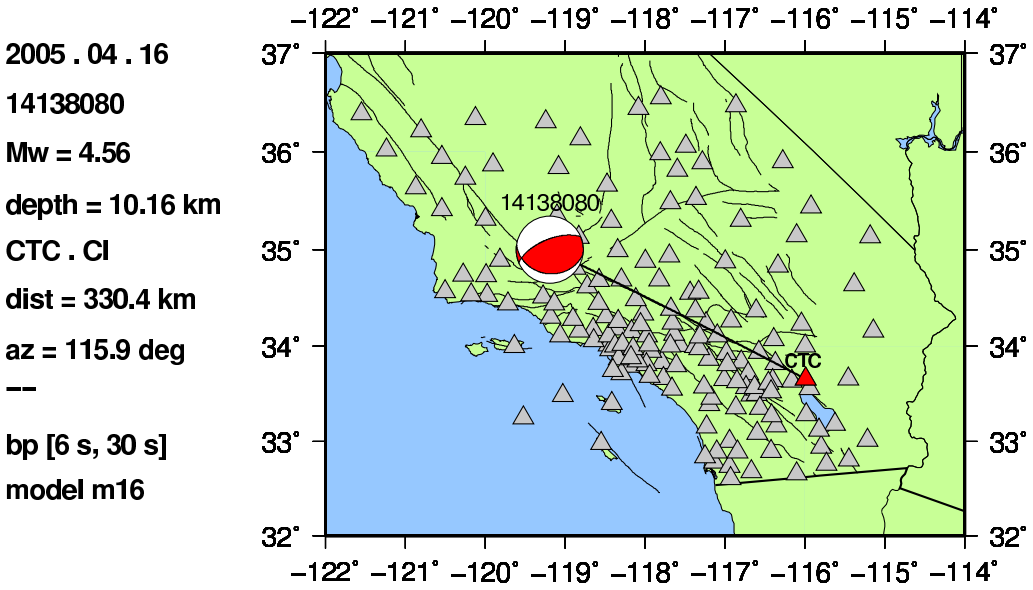
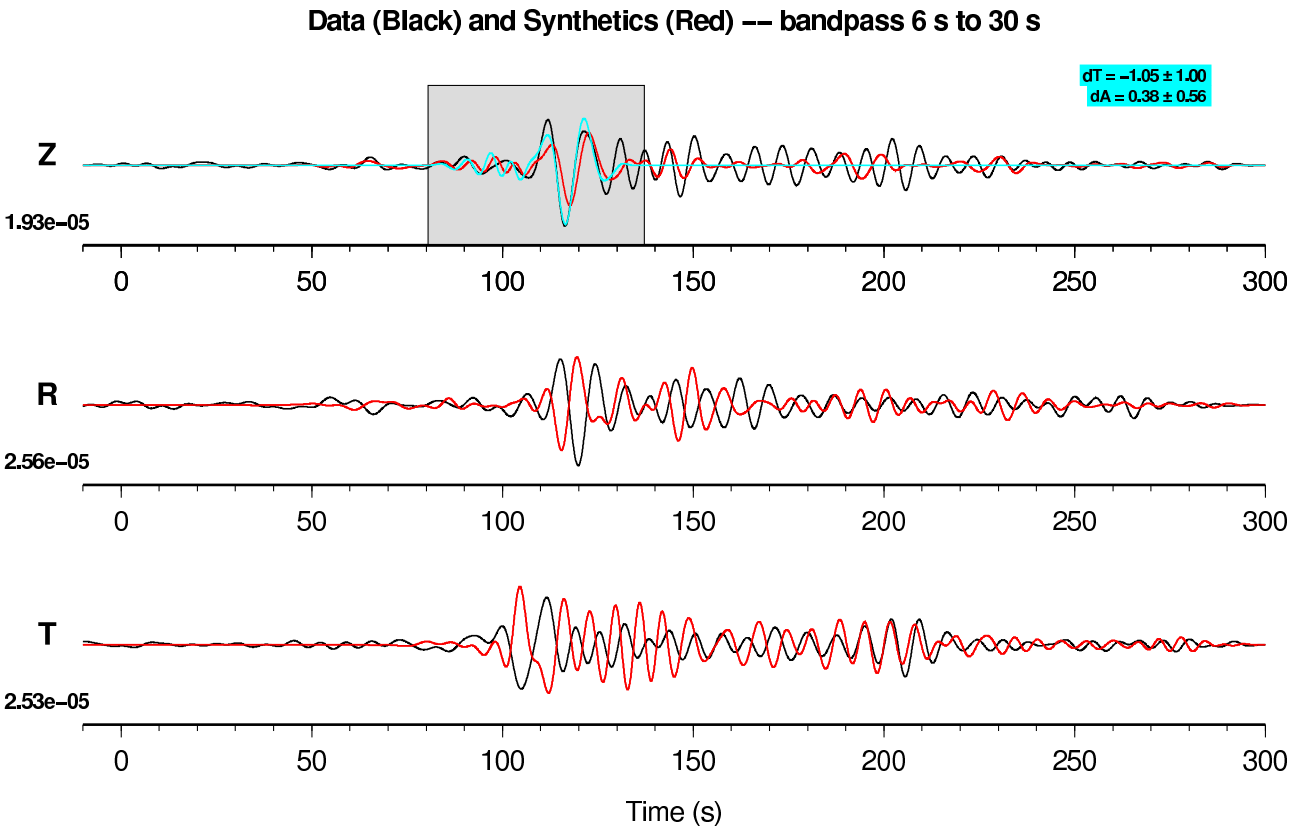


Figure E.13: Data (black) and synthetics (red), 6–30 s, from 14138080 to CTC.CI. Compare with Figure E.14.



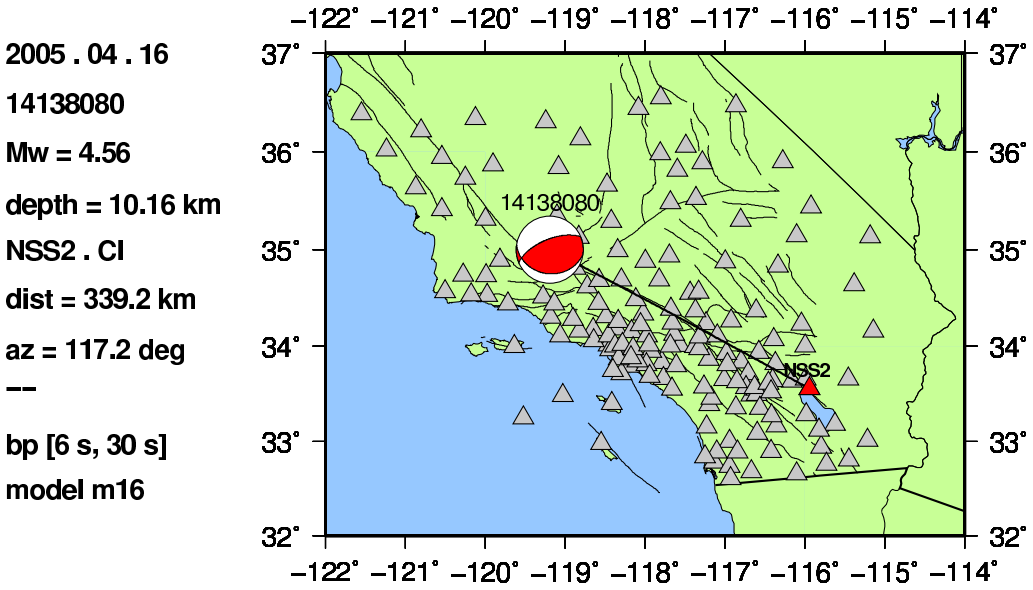
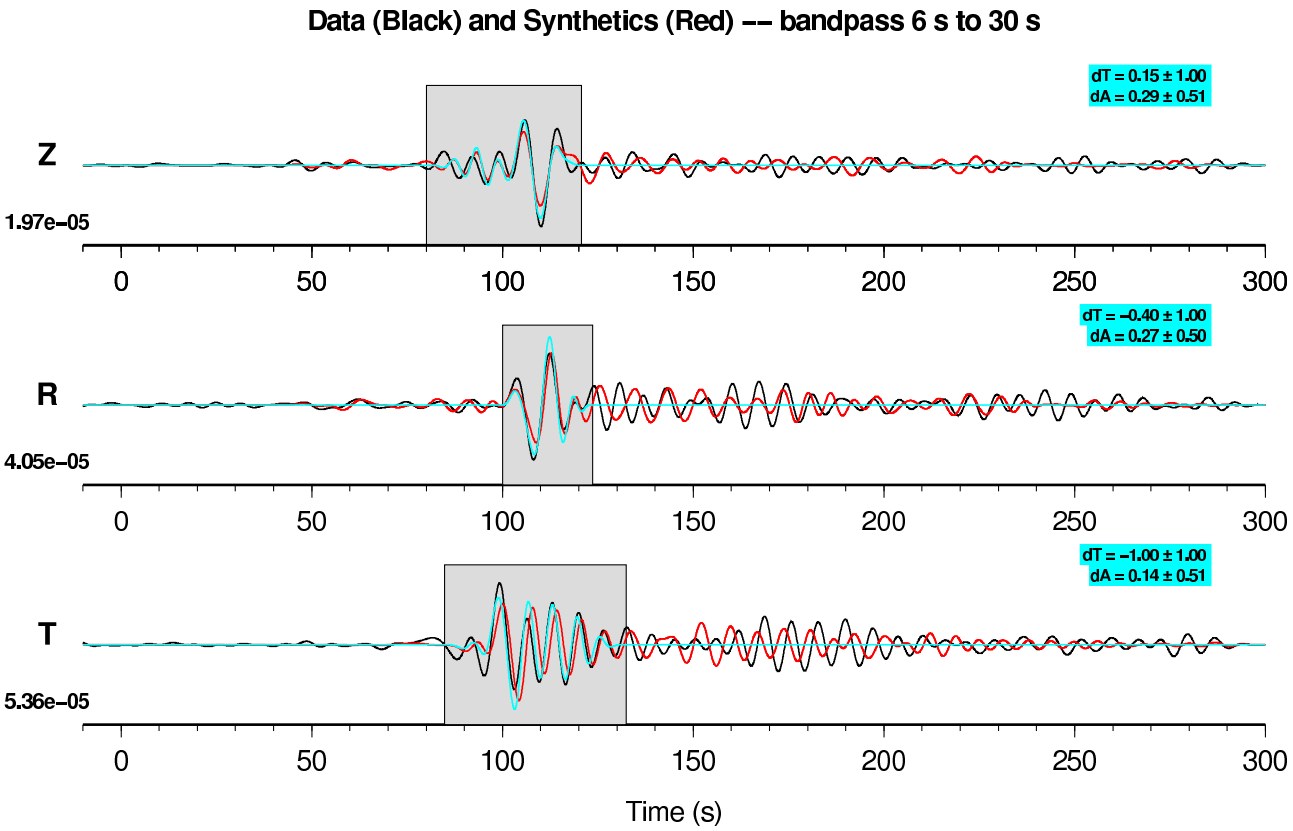


Figure E.14: Data (black) and synthetics (red), 6–30 s, from 14138080 to NSS2.CI. Compare with Figures E.13 and E.15.



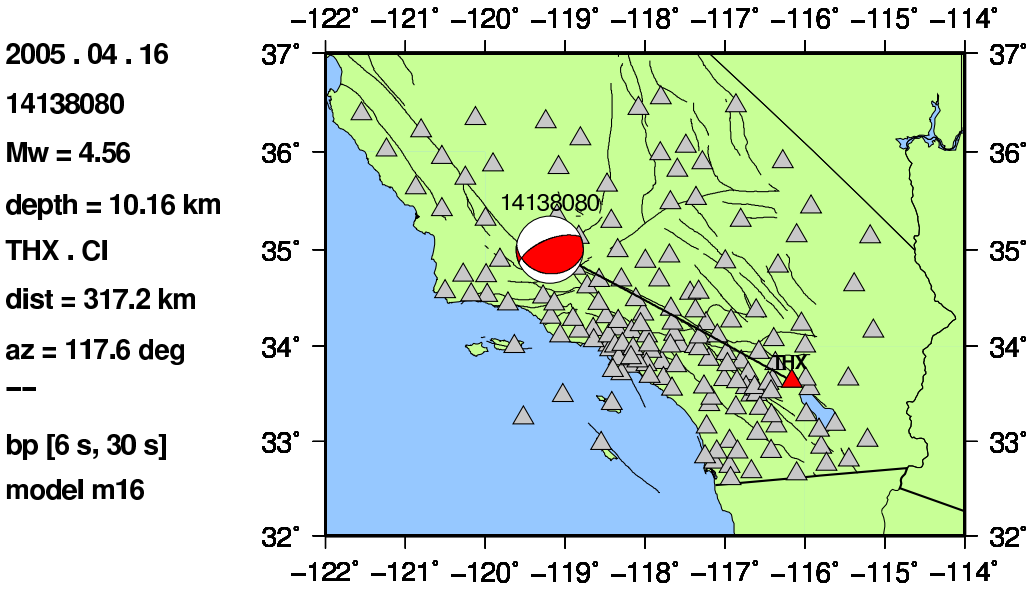
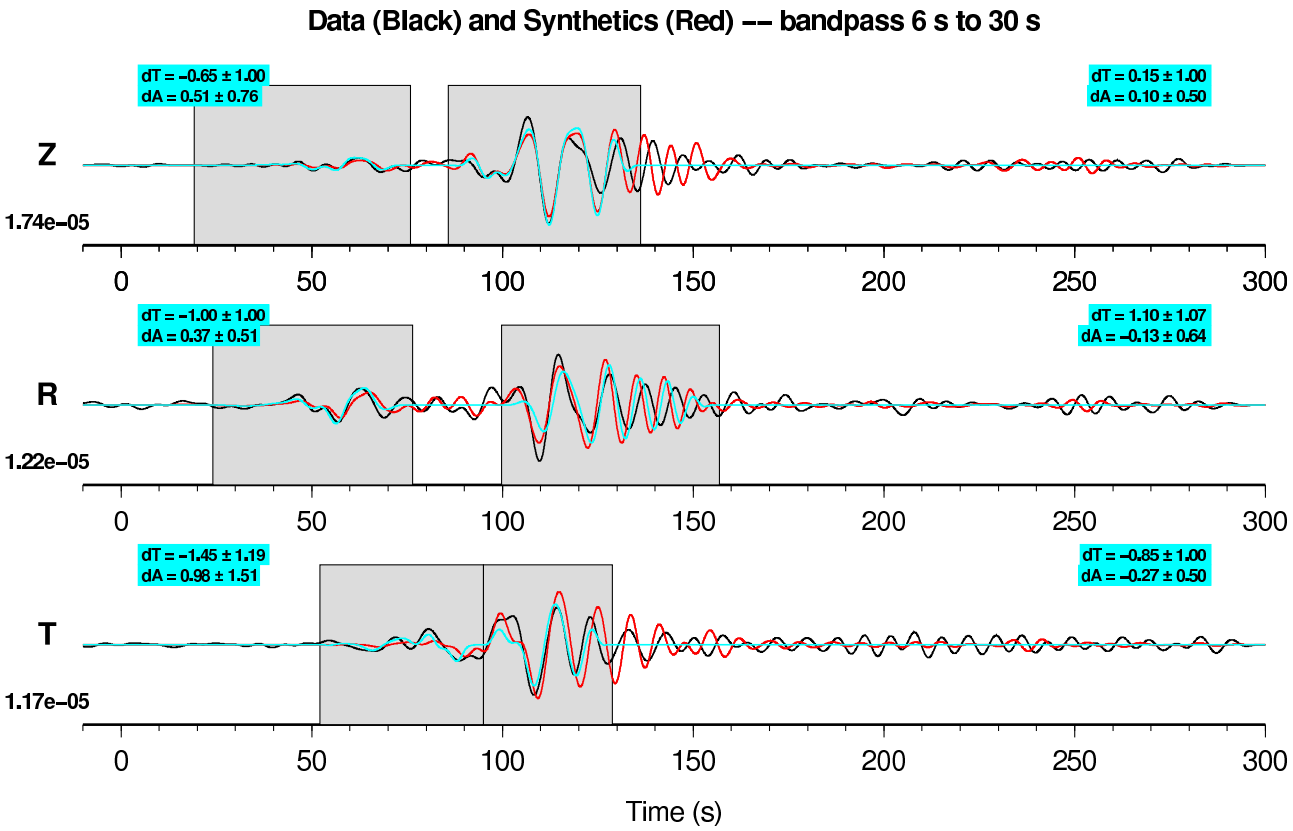


Figure E.15: Data (black) and synthetics (red), 6–30 s, from 14138080 to THX.CI. Compare with Figure E.14.



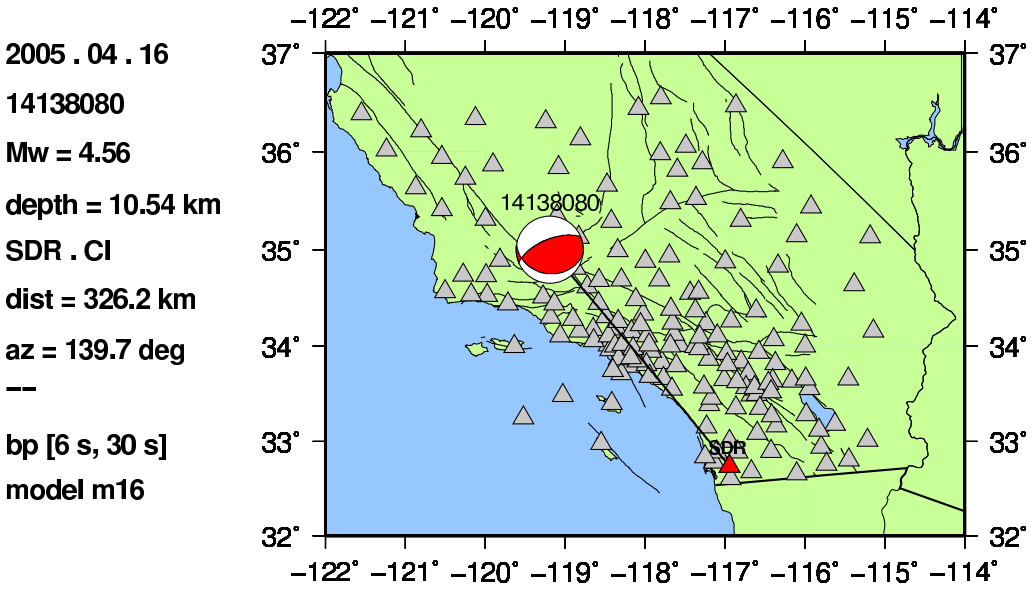
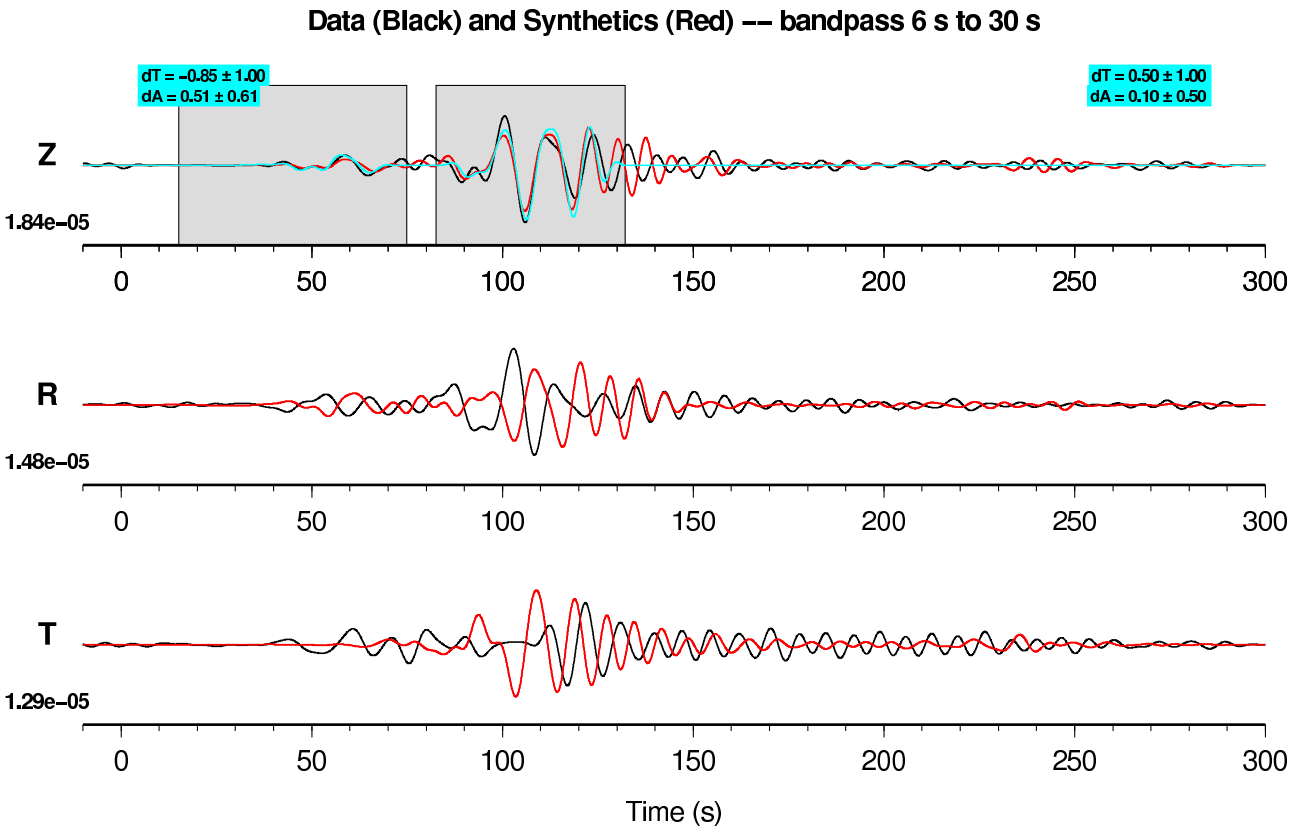


Figure E.16: Data (black) and synthetics (red), 6–30 s, from 14138080 to SDR.CI. Compare with Figure E.17.



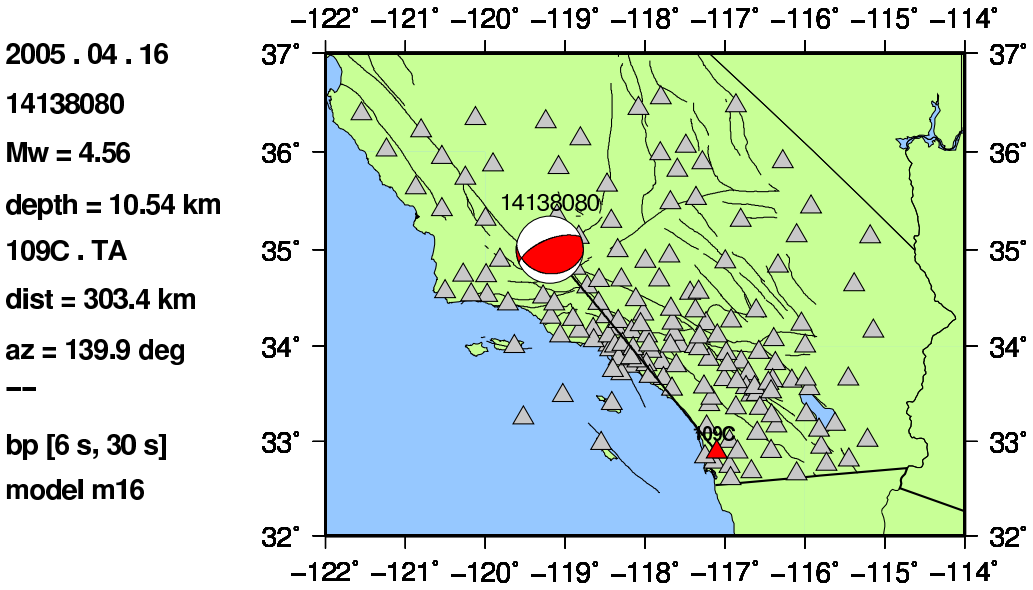
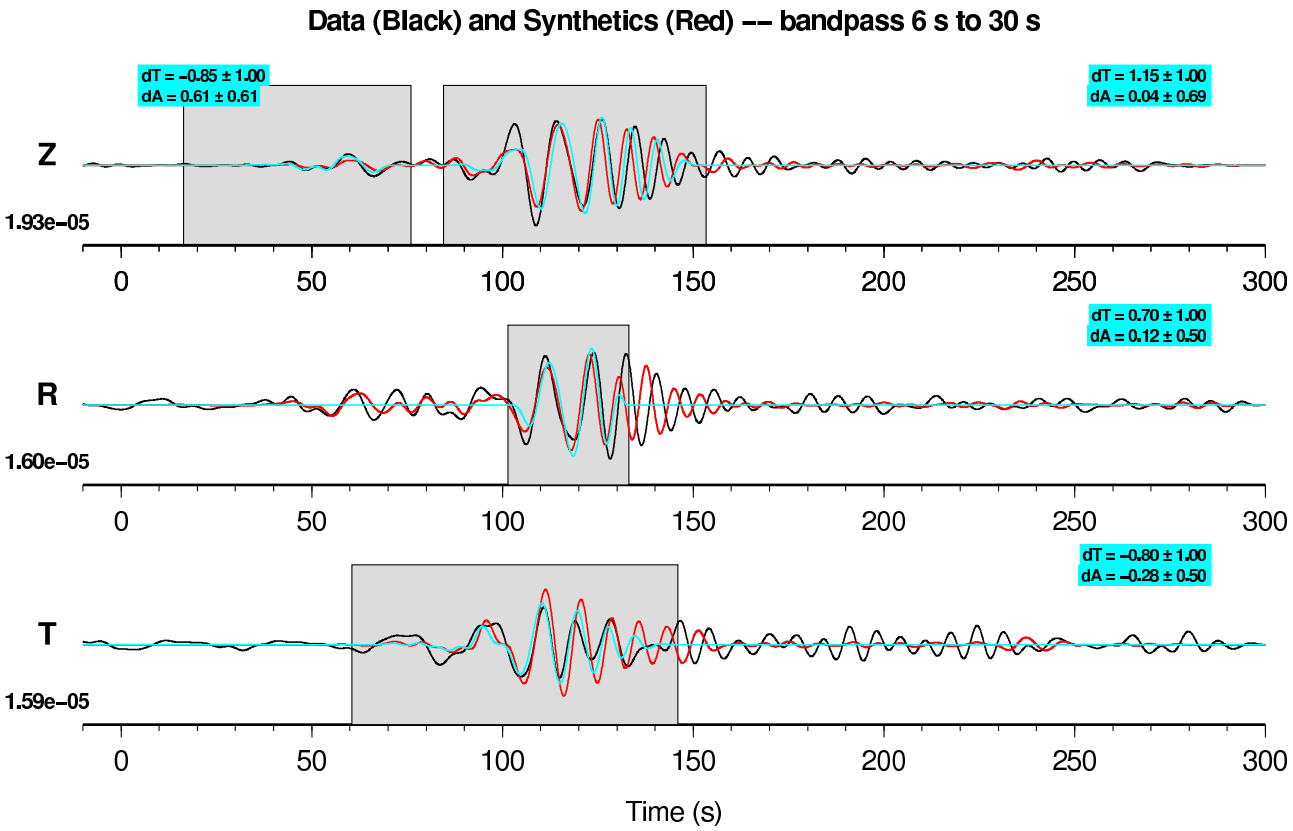


Figure E.17: Data (black) and synthetics (red), 6–30 s, from 14138080 to 109C.TA. Compare with Figures E.16 and E.18.



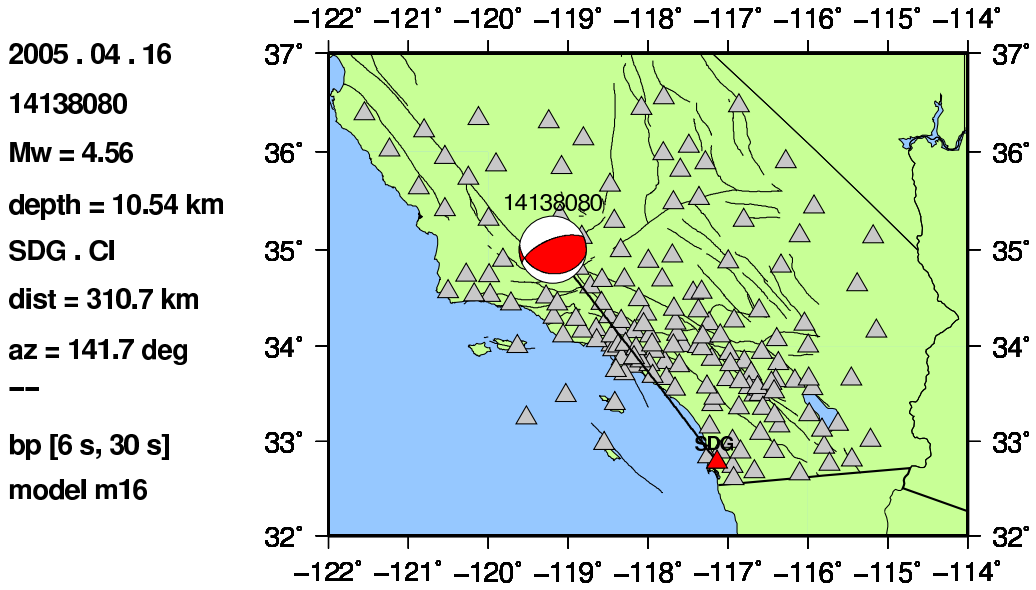
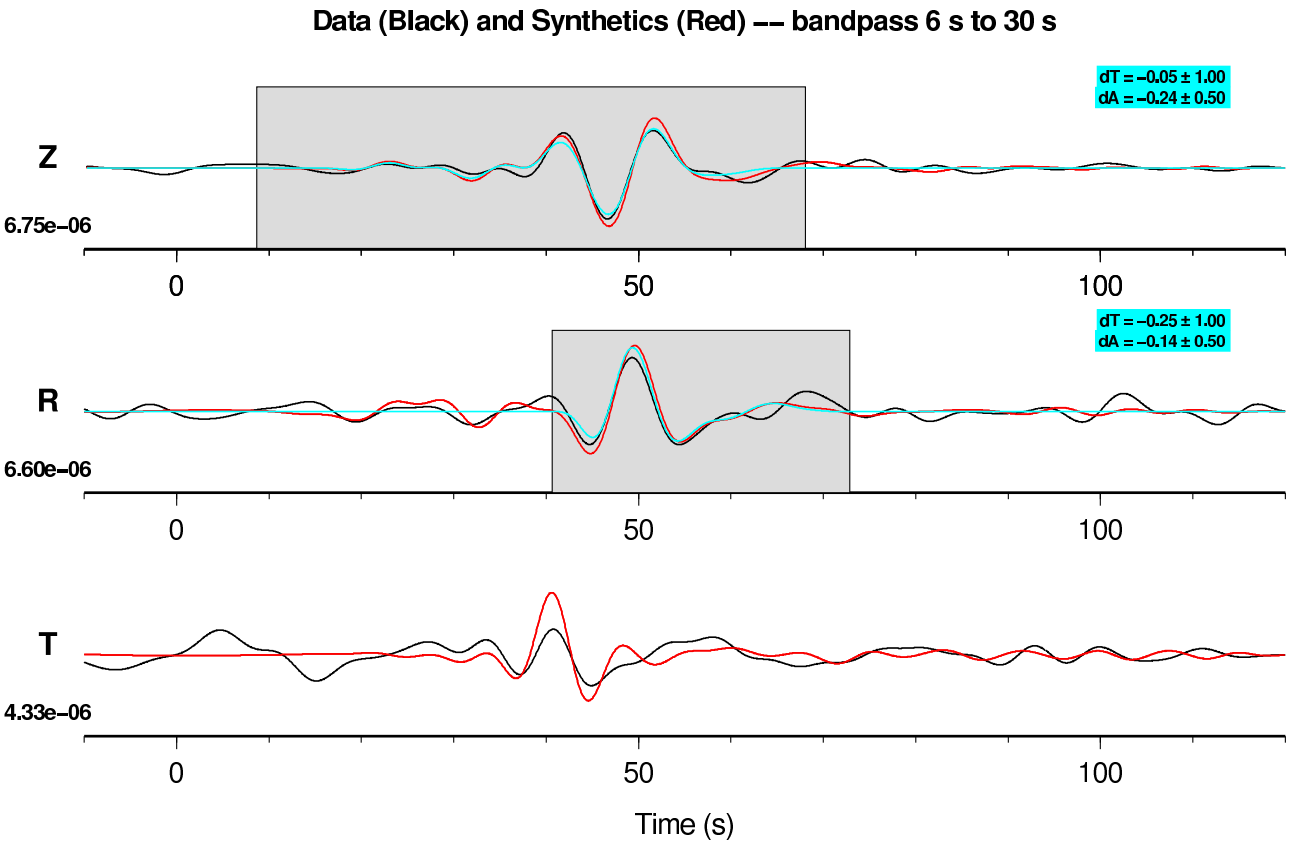


Figure E.18: Data (black) and synthetics (red), 6–30 s, from 14138080 to SDG.CI. Compare with Figure E.17.



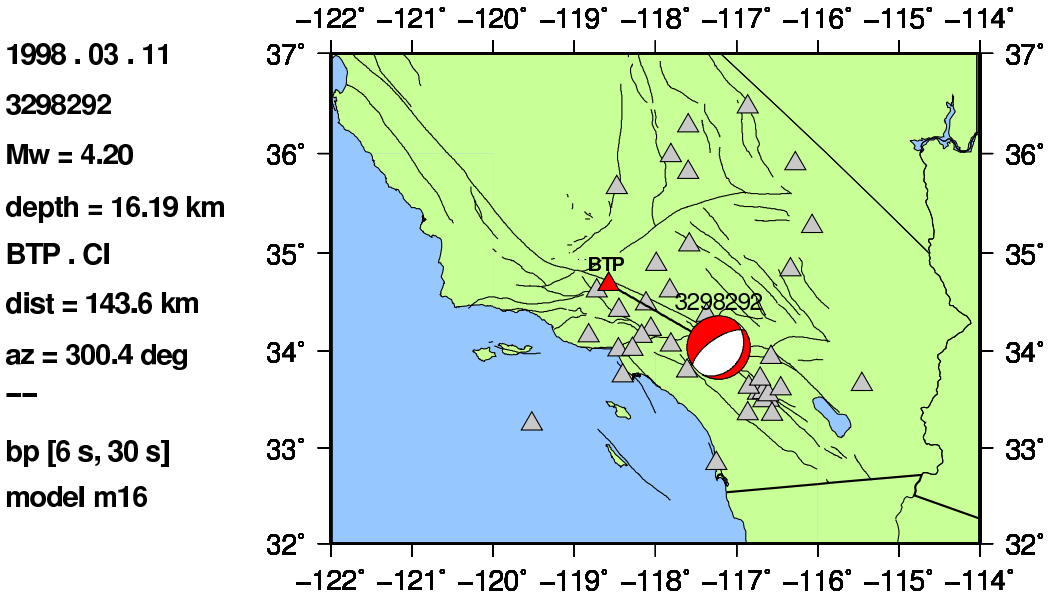
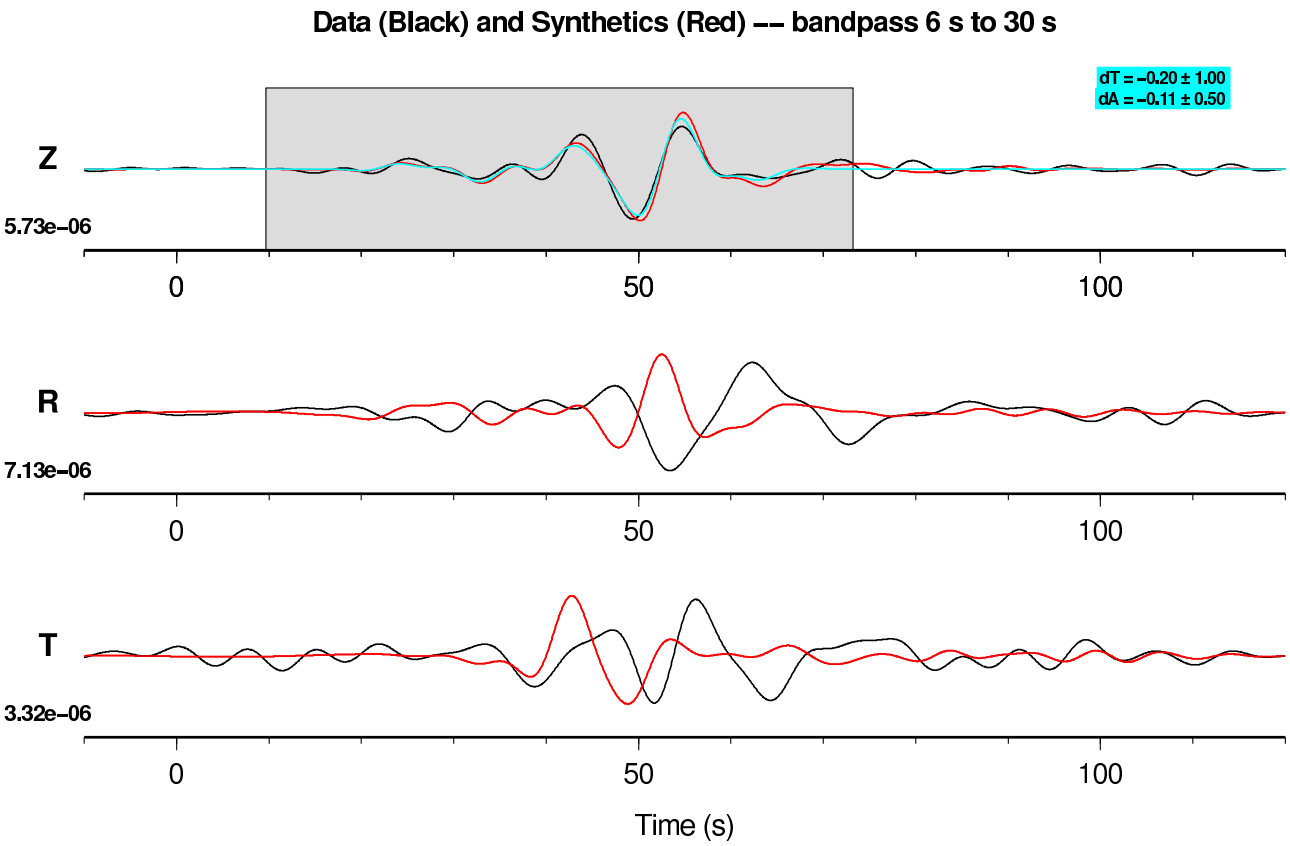


Figure E.19: Data (black) and synthetics (red), 6–30 s, from 3298292 to BTP.CI. Note that the polarity problem on this east-west-oriented path is most apparent on the radial component. Compare with Figure E.20.



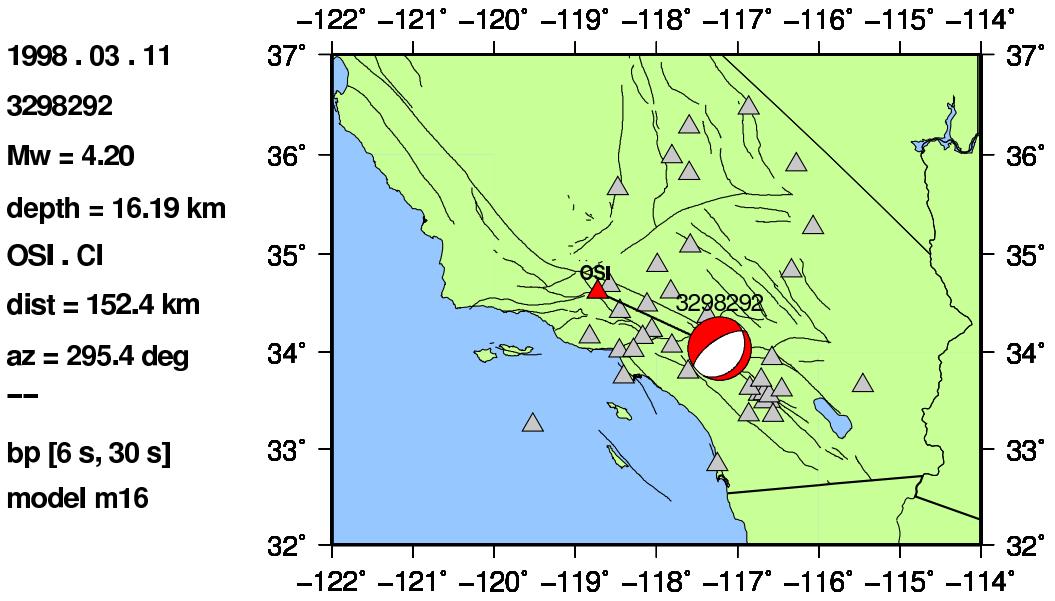
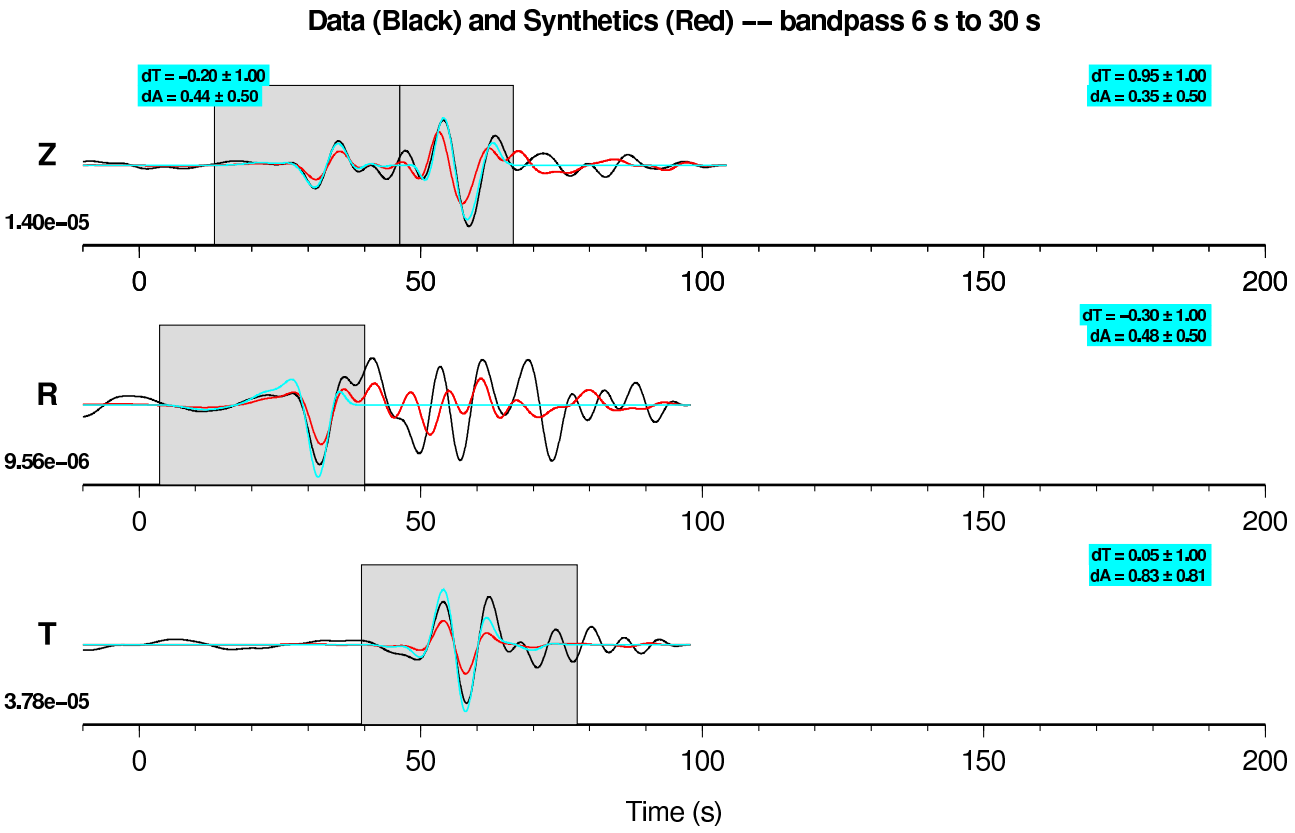


Figure E.20: Data (black) and synthetics (red), 6–30 s, from 3298292 to OSI.CI. Compare with Figure E.12.



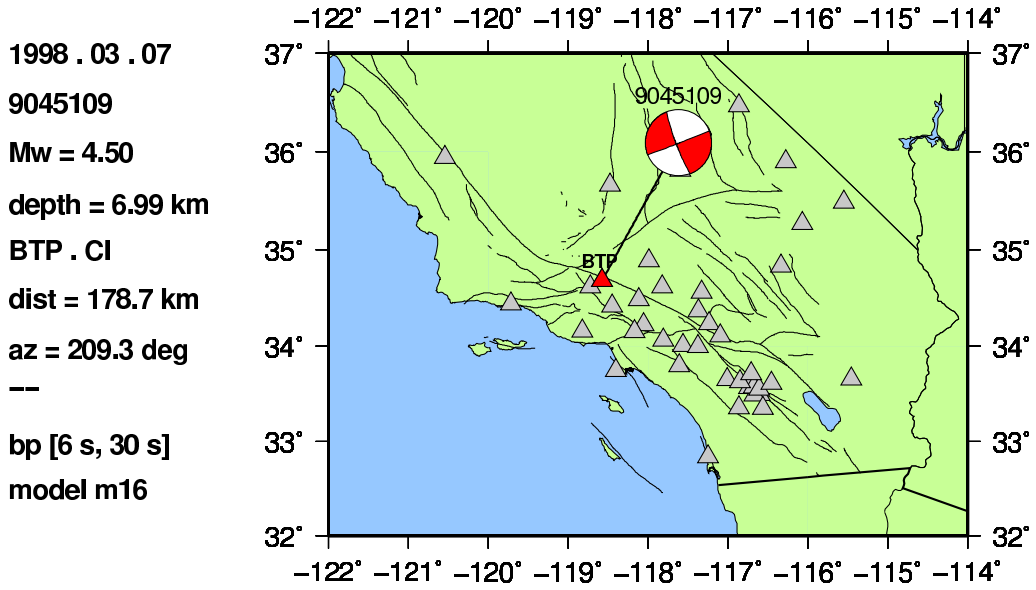
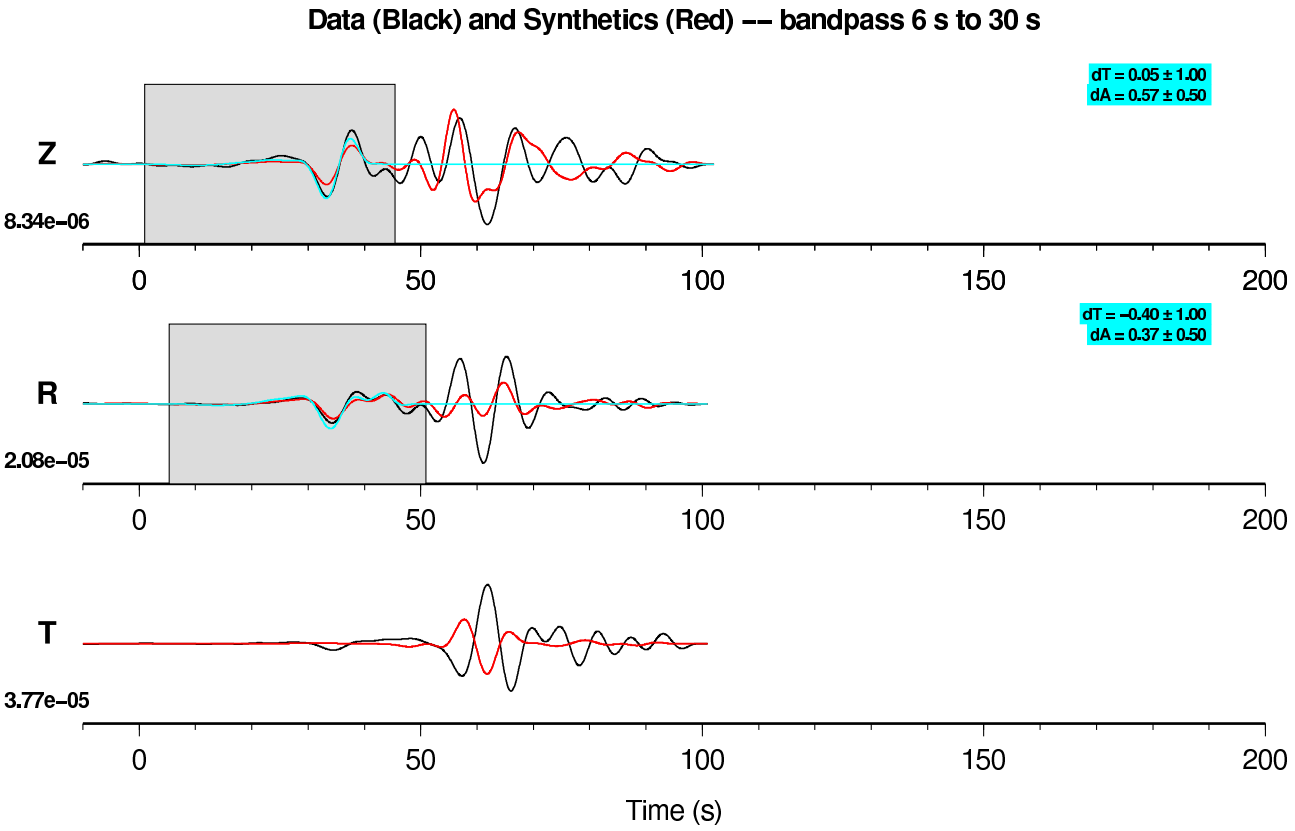


Figure E.21: Data (black) and synthetics (red), 6–30 s, from 9045109 to BTP.CI. Compare with Figure E.22.



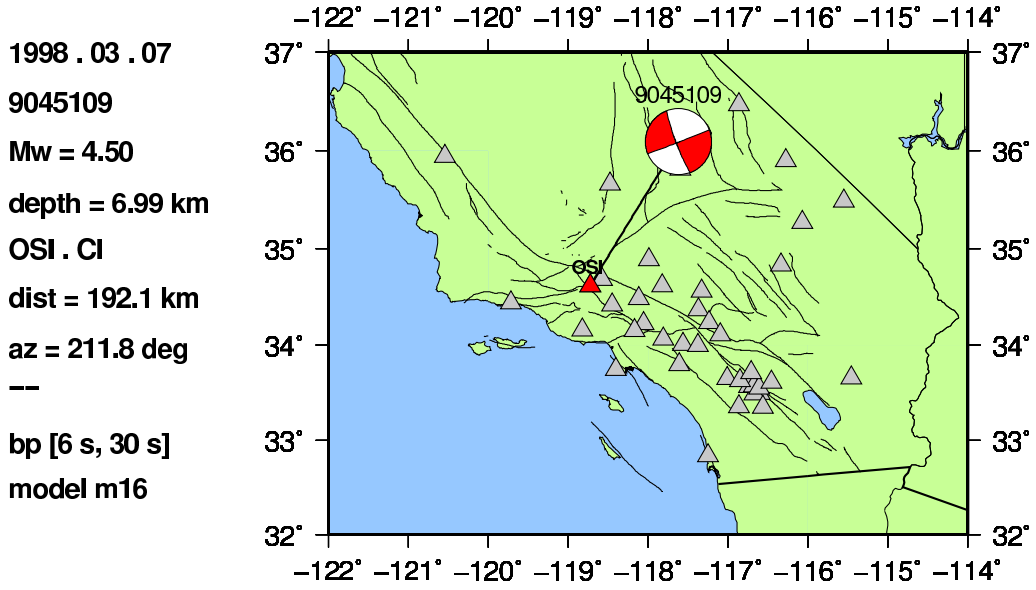
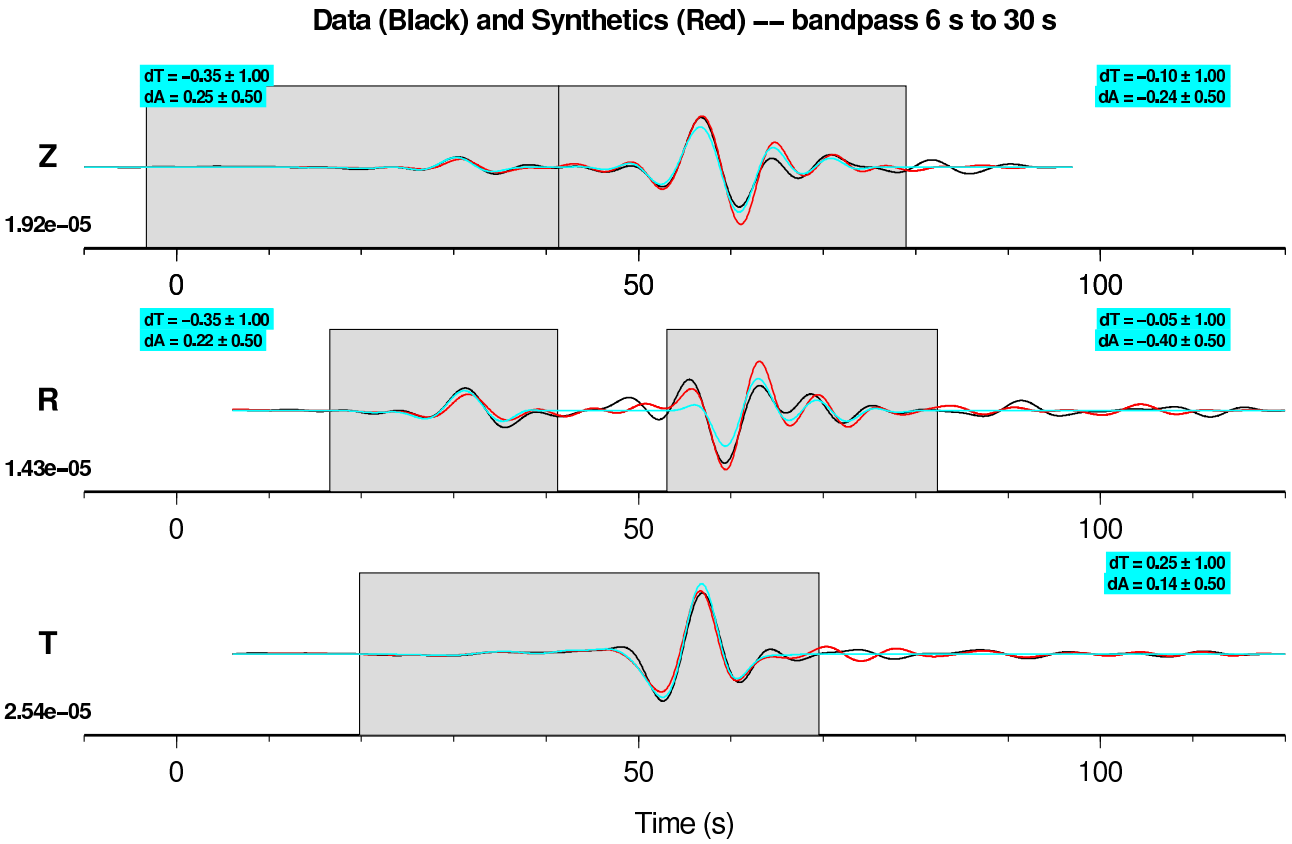


Figure E.22: Data (black) and synthetics (red), 6–30 s, from 9045109 to OSI.CI. Note that the polarity problem on this north-south-oriented path is most apparent on the transverse component. Compare with Figure E.21.



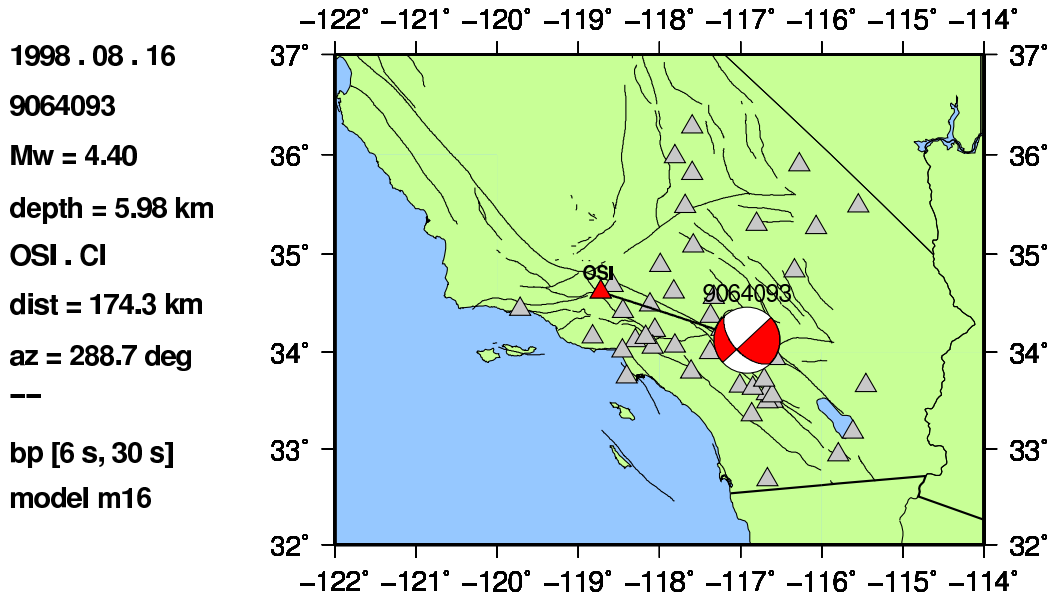
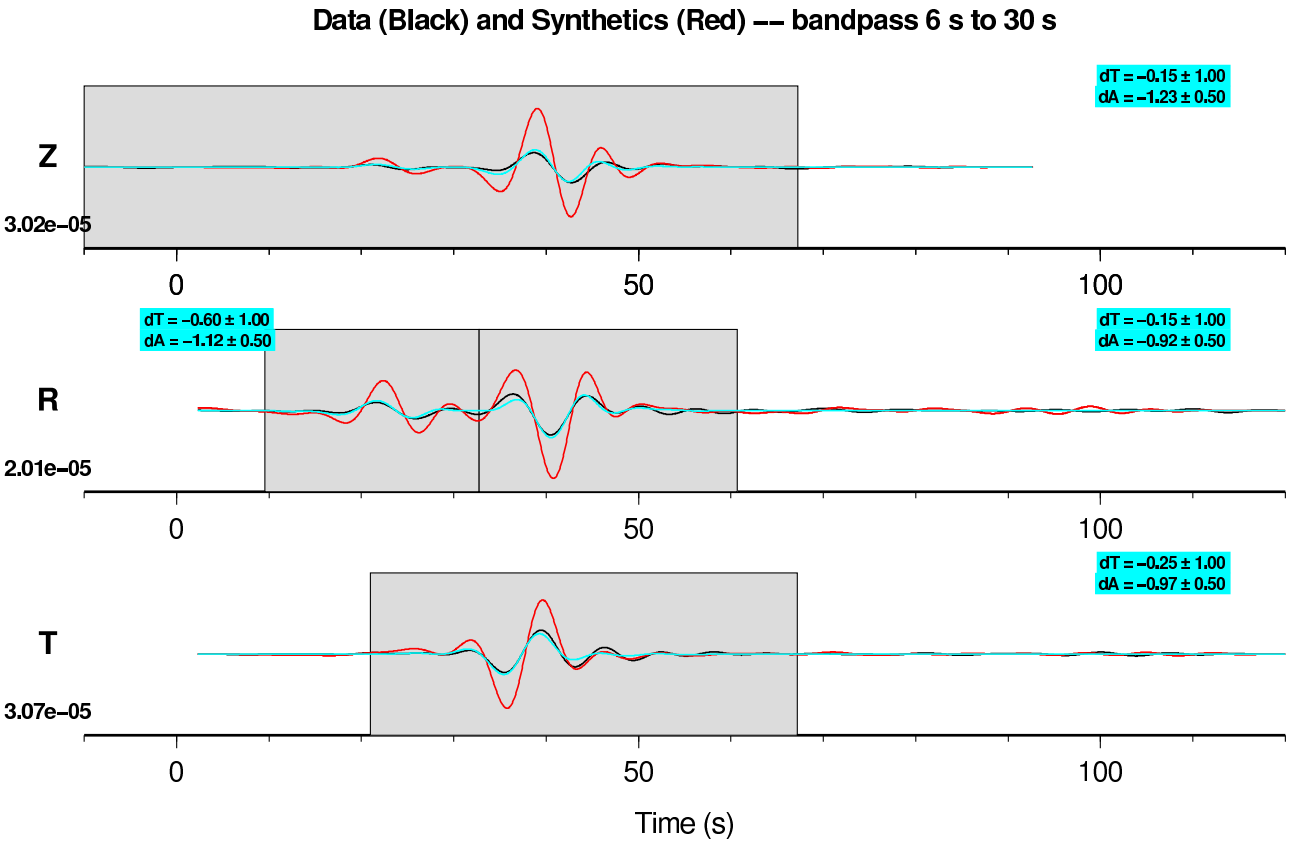


Figure E.23: Data (black) and synthetics (red), 6–30 s, from 9064093 to OSI.CI. Compare with Figure E.24.



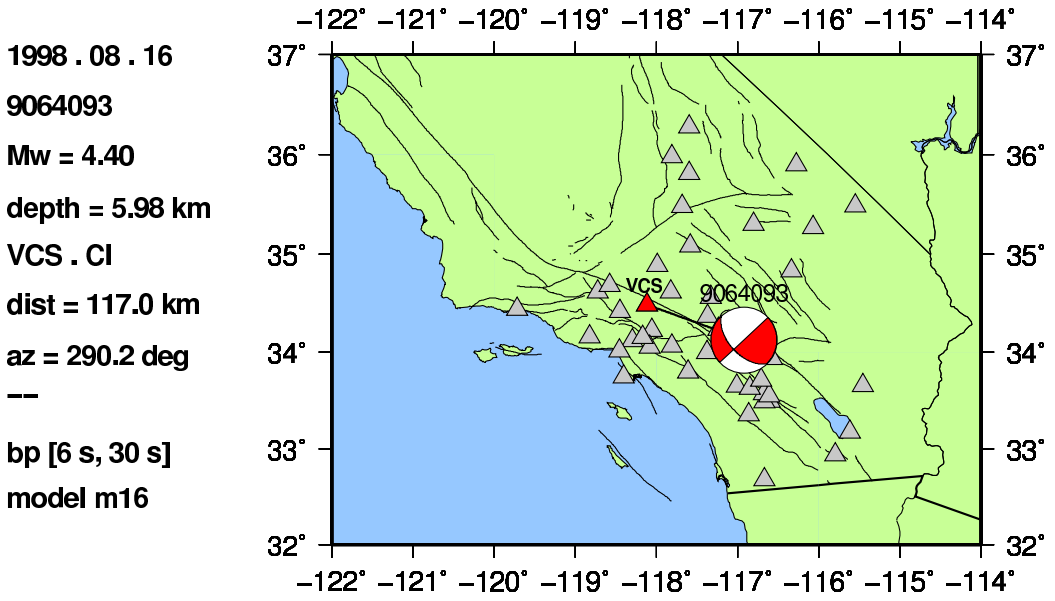
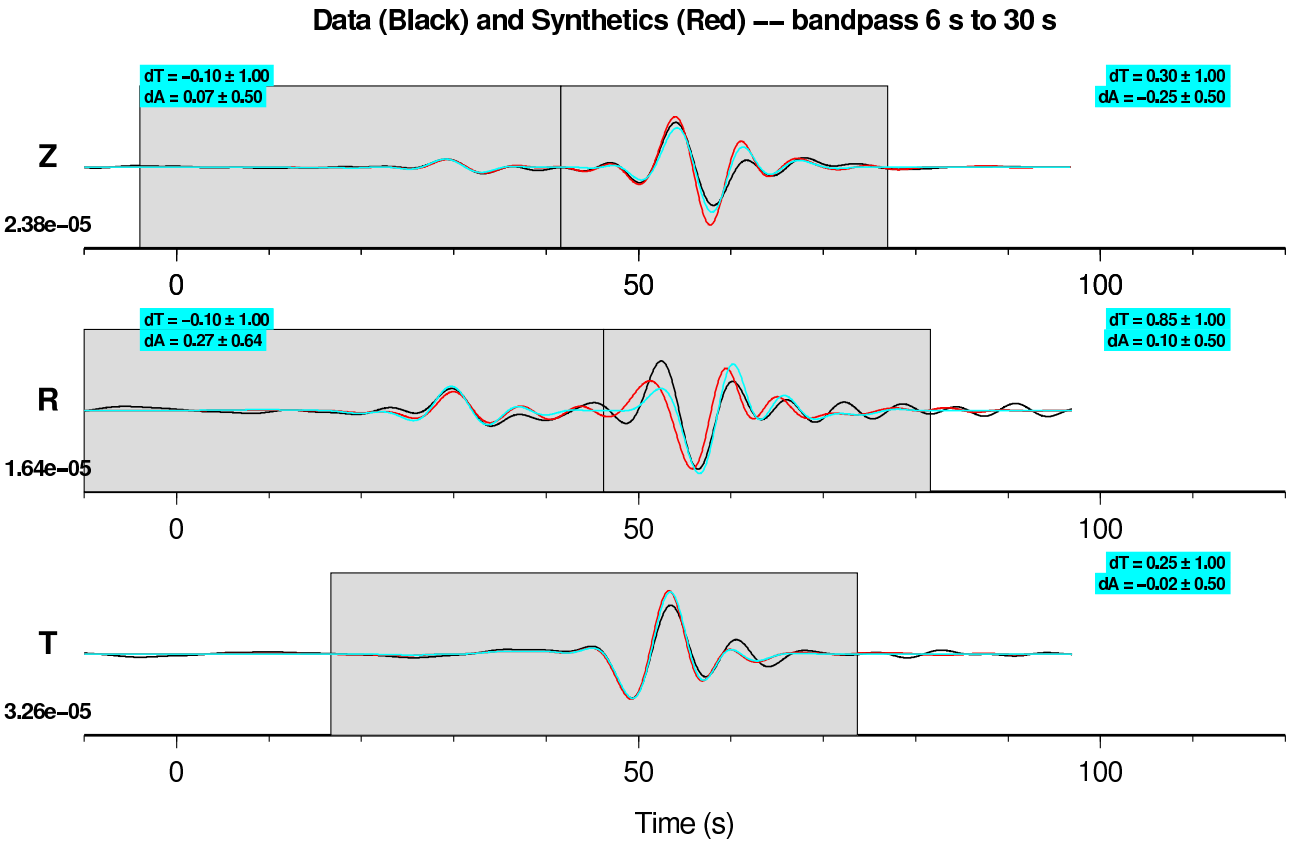


Figure E.24: Data (black) and synthetics (red), 6–30 s, from 9064093 to VCS.CI. The amplification problem is on all three components. Compare with Figures E.25 and E.23.



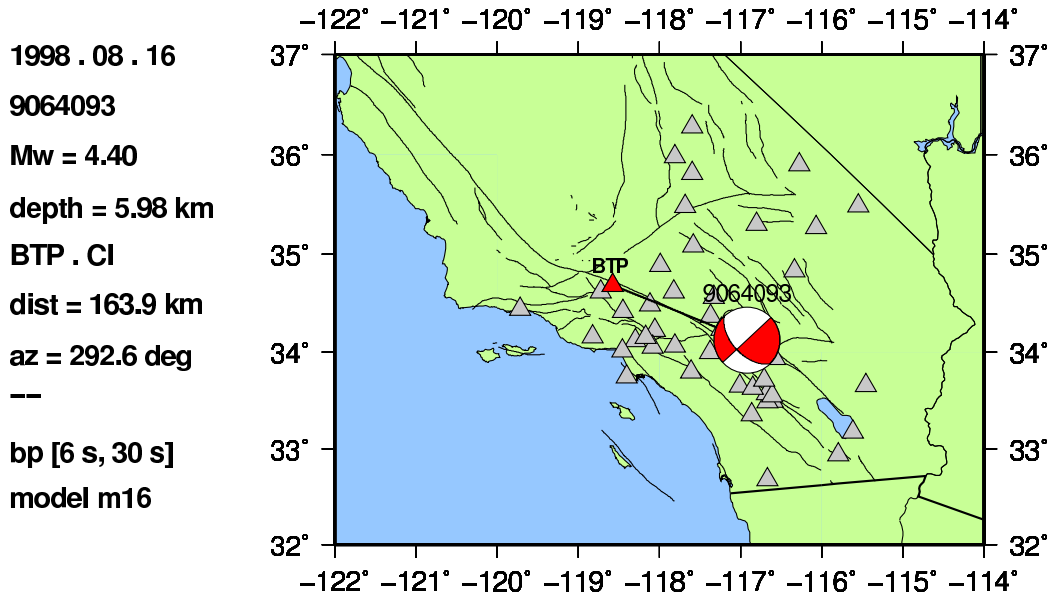
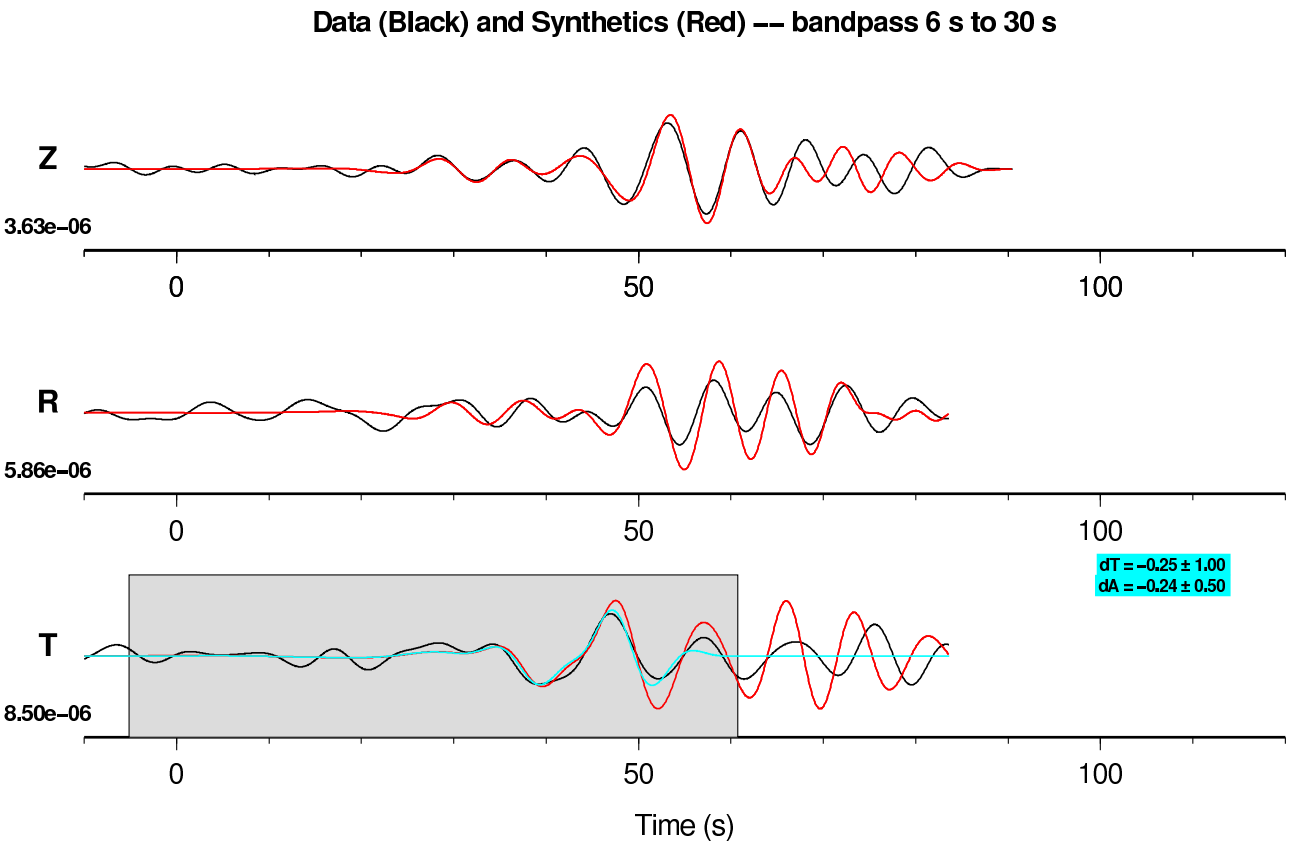


Figure E.25: Data (black) and synthetics (red), 6–30 s, from 9064093 to BTP.CI. Compare with Figure E.24.



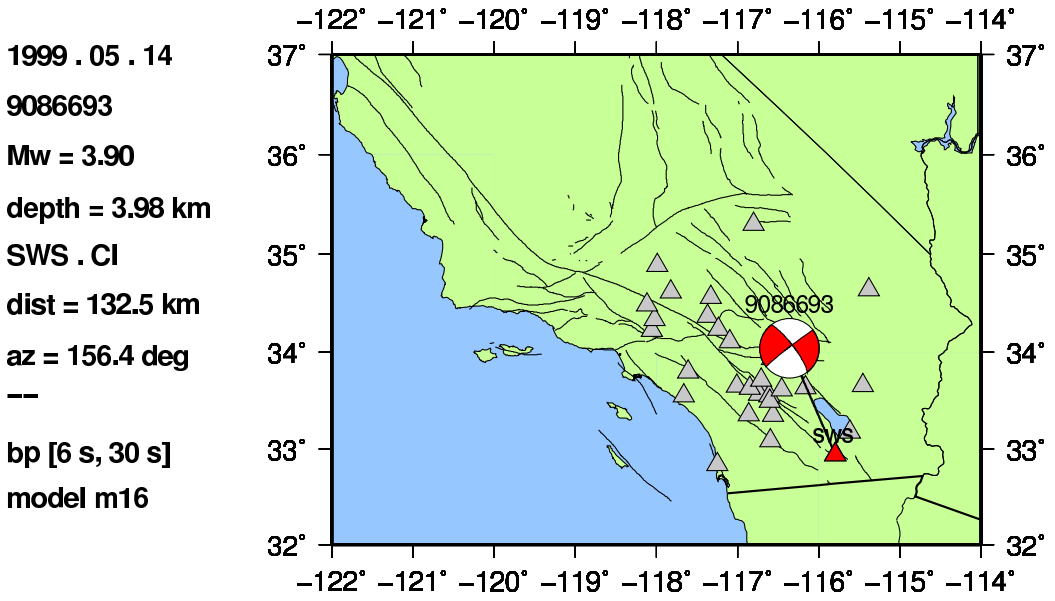
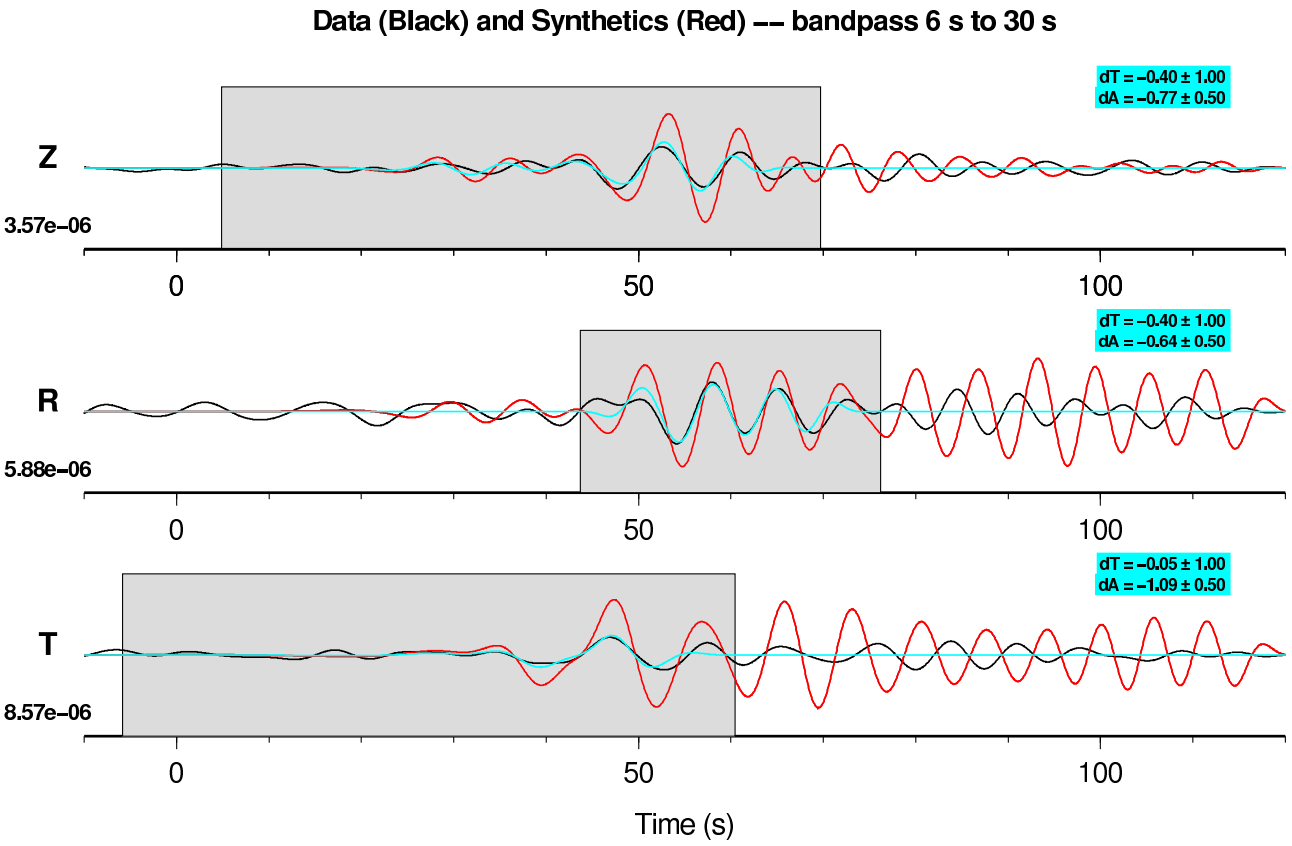


Figure E.26: Data (black) and synthetics (red), 6–30 s, from 9086693 to SWS.CI. Compare with Figure E.27.



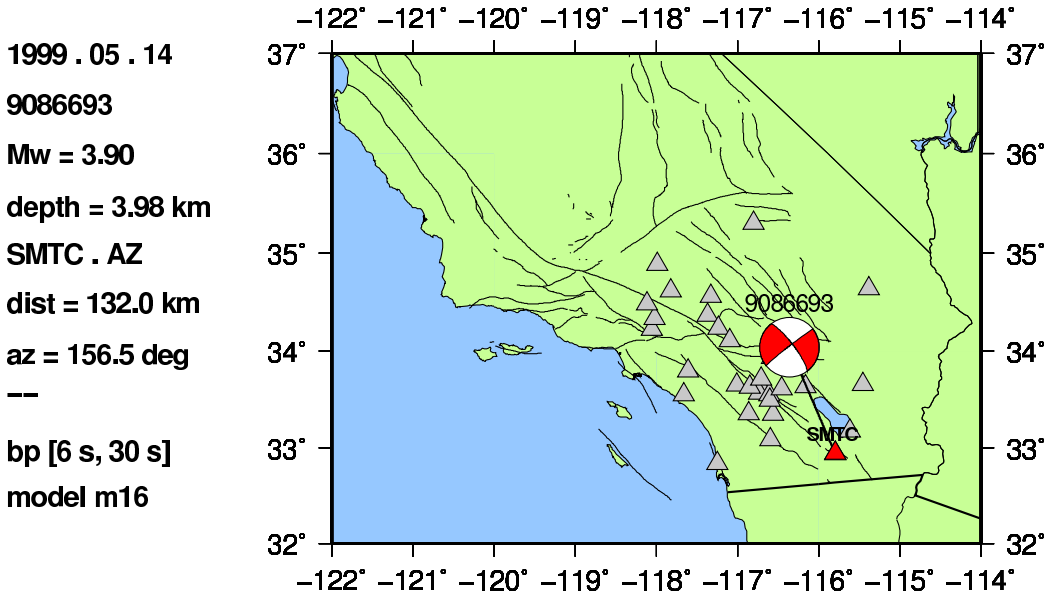
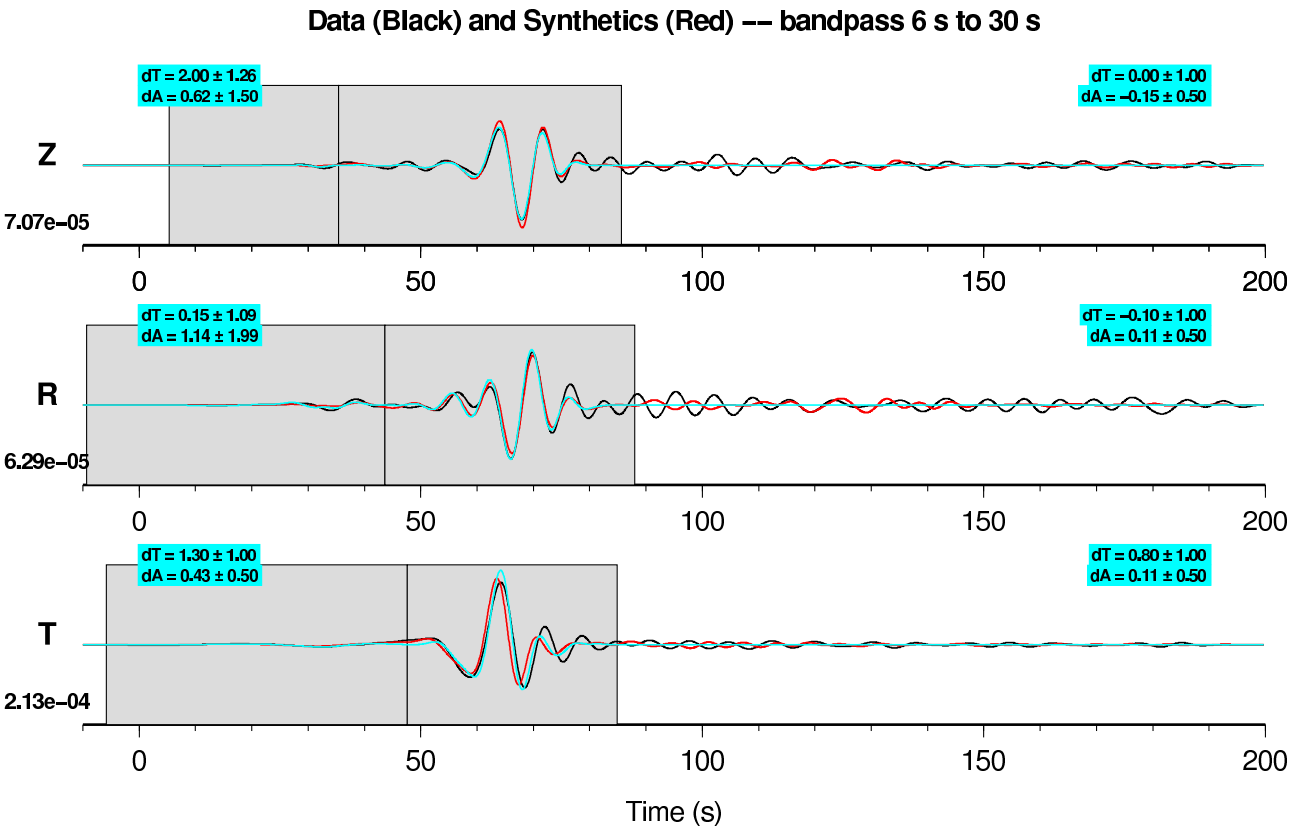


Figure E.27: Data (black) and synthetics (red), 6–30 s, from 9086693 to SMTC.AZ. The amplification problem is most apparent on the vertical and transverse components for this north-south path. Compare with Figure E.26.



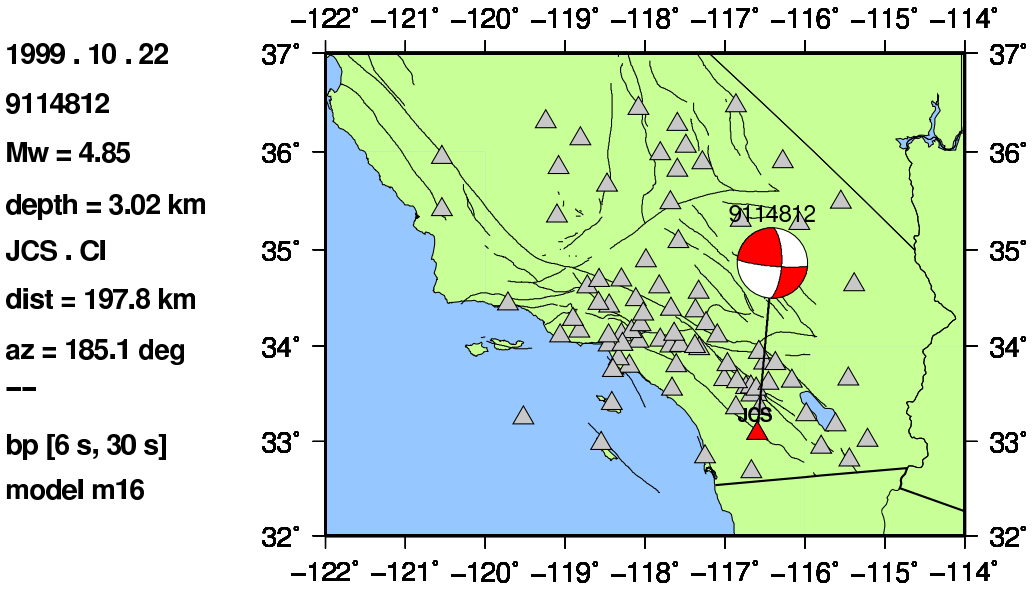
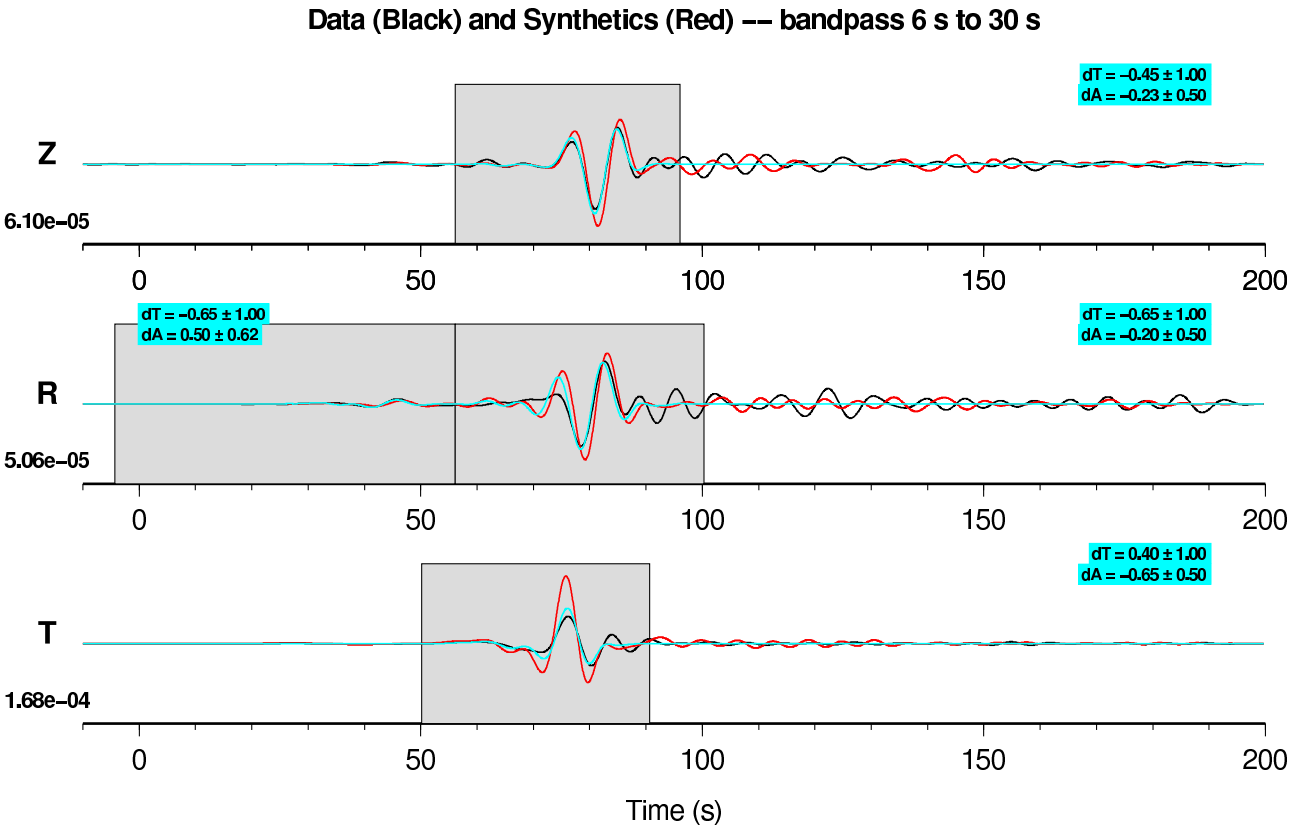


Figure E.28: Data (black) and synthetics (red), 6–30 s, from 9114812 to JCS.CI. Compare with Figure E.29.



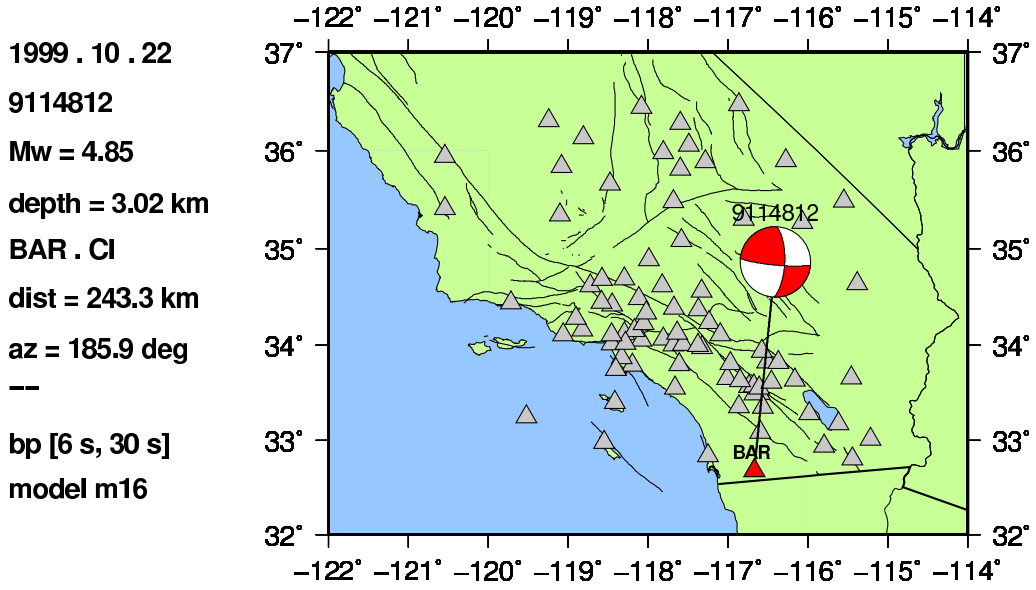


Figure E.29: Data (black) and synthetics (red), 6–30 s, from 9114812 to BAR.CI. Note that the amplification problem on this north-south-oriented path is only on the transverse component. Compare with Figure E.28.

Bibliography

Aagaard, B. T., T. M. Brocher, D. Dolenc, D. Dreger, R. W. Graves, S. Harmsen, S. Hartzell, S. Larsen, and M. L. Zoback, Ground-motion modeling of the 1906 San Francisco earthquake, part I: Validation using the 1989 Loma Prieta earthquake, *Bull. Seismol. Soc. Am.*, *98*, 989–1011, 2008a.

Aagaard, B. T., et al., Ground-motion modeling of the 1906 San Francisco earthquake, part II: Gound-motion estimates for the 1906 earthquake and scenario events, *Bull. Seismol. Soc. Am.*, *98*, 1012–1046, 2008b.

Akçelik, V., G. Biros, and O. Ghattas, Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation, *Proc. ACM/IEEE Supercomputing SC'2002 conference*, 2002, published on CD-ROM and at www.sc-conference.org/sc2002.

Akçelik, V., et al., High resolution forward and inverse earthquake modeling on terascale computers, *Proc. ACM/IEEE Supercomputing SC'2003 conference*, 2003, published on CD-ROM and at www.sc-conference.org/sc2003.

Allen, R. V., Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.*, *68*, 1521–1532, 1978.

Amante, C., and B. W. Eakins, ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis, National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce, Boulder, CO, 2008.

Bijwaard, H., and W. Spakman, Nonlinear global P-wave tomography by iterated linearized inversion, *Geophys. J. Int.*, *141*, 71–82, 2000.

Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis, Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging, *J. Geophys. Res.*, *112*, 2007.

Bonner, J. L., D. D. Blackwell, and E. T. Herrin, Thermal constraints on earthquake depths in California, *Bull. Seismol. Soc. Am.*, *93*, 2333–2354, 2003.

Boschi, L., G. Ekström, and B. Kustowski, Multiple resolution surface wave tomography: the Mediterranean basin, *Geophys. J. Int.*, *157*, 293–304, 2004.

Brocher, T. M., Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. Seismol. Soc. Am.*, *95*, 2081–2092, 2005.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, Multiscale seismic waveform inversion, *Geophysics*, *60*, 1457–1473, 1995.

Capdeville, Y., E. Chaljub, J. P. Vilotte, and J. P. Montagner, Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophys. J. Int.*, *152*, 34–67, 2003.

Capdeville, Y., Y. Gung, and B. Romanowicz, Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.*, *162*, 541–554, 2005.

Chen, P., L. Zhao, and T. H. Jordan, Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. Seismol. Soc. Am.*, *97*, 1094–1120, 2007.

Christensen, N. I., Poisson's ratio and crustal seismology, *J. Geophys. Res.*, *101*, 3139–3156, 1996.

Christensen, N. I., and W. D. Mooney, Seismic velocity structure and composition of the continental crust: A global view, *J. Geophys. Res.*, *100*, 9761–9788, 1995.

Claerbout, J. F., Toward a unified theory of reflector mapping, *Geophysics*, *36*, 467–481, 1971.

Clinton, J. F., E. Hauksson, and K. Solanki, An evaluation of the SCSN moment tensor solutions: Robustness of the $M_{\rm w}$ magnitude scale, style of faulting, and automation of the method, *Bull. Seismol. Soc. Am.*, *96*, 1689–1705, 2006.

Courtier, P., and O. Talagrand, Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results, *Q. J. R. Meteorol. Soc.*, *113*, 1329–1347, 1987.

Custódio, S., P. Liu, and R. J. Archuleta, The 2004 $M_{\rm w}$6.0 Parkfield, California, earthquake: Inversion of near-source ground motion using multiple data sets, *Geophys. Res. Lett.*, *32*, 2005.

Dahlen, F. A., Finite-frequency sensitivity kernels for boundary topography perturbations, *Geophys. J. Int.*, *162*, 525–540, 2005.

Dahlen, F. A., and A. M. Baig, Fréchet kernels for body-wave amplitudes, *Geophys. J. Int.*, *150*, 440–446, 2002.

Dahlen, F. A., and J. Tromp, *Theoretical Global Seismology*, Princeton U. Press, New Jersey, 1998.

Dahlen, F. A., S.-H. Hung, and G. Nolet, Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.*, *141*, 157–174, 2000.

Dreger, D., R. Uhrhammer, M. Pasyanos, J. Franck, and B. Romanowicz, Regional and far-regional earthquake locations and source parameters using sparse broadband networks: A test on the Ridgecrest sequence, *Bull. Seismol. Soc. Am.*, *88*, 1353–1362, 1998.

Dreger, D. S., and D. V. Helmberger, Broadband modeling of local earthquakes, *Bull. Seismol. Soc. Am.*, *80*, 1162–1179, 1990.

Dziewonski, A., and D. Anderson, Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356, 1981.

Ekström, G., J. Tromp, and E. W. F. Larson, Measurements and global models of surface wave propagation, *J. Geophys. Res.*, *102*, 8137–8157, 1997.

Fichtner, A., H.-P. Bunge, and H. Igel, The adjoint method in seismology—I. Theory, *Phys. Earth Planet. Inter.*, *157*, 86–104, 2006.

Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge, Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.(in review)*, 2009.

Fink, M., Time reversal of ultrasonic fields—Part I: Basic principles, *IEEE Trans. Ultras. Ferroelec. Freq. Contr.*, *39*, 555–566, 1992.

Fink, M., Time reversed acoustics, *Phys. Today*, *50*, 34–40, 1997.

Fink, M., C. Prada, F. Wu, and D. Cassereau, Self focusing in inhomogeneous media with "time reversal" acoustic mirrors, in *Proc. IEEE Ultrasonic. Symp. 1989*, vol. 2, pp. 681–686, 1989.

Fletcher, R., and C. M. Reeves, Function minimization by conjugate gradients, *Comp. J.*, *7*, 149–154, 1964.

Fuis, G. S., and W. M. Kohler, Crustal structure and tectonics of the Imperial Valley region, California, in *The Imperial Basin—Tectonics, Sedimentation, and Thermal Aspects*, edited by C. A. Rigsby, pp. 1–13, Pacific Section S.E.P.M, Los Angeles, Calif., USA, 1984.

Fuis, G. S., et al., Fault systems of the 1971 San Fernando and 1994 Northridge earthquakes, southern California: Relocated aftershocks and seismic images from LARSE II, *Geology*, *31*, 171–174, 2003.

Gauthier, O., J. Virieux, and A. Tarantola, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results, *Geophysics*, *51*, 1387–1403, 1986.

Goodman, E. D., and P. E. Malin, Evolution of the southern San Joaquin basin and mid-Tertiary "transitional" tectonics, central California, *Tectonics*, *11*, 478–498, 1992.

Grand, S. P., R. D. van der Hilst, and S. Widiyantoro, Global seismic tomography: A snapshot of convection in the Earth, *GSA Today*, *7*, 1–7, 1997.

Hansen, P. C., *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, Penn., 1998.

Hardebeck, J. L., and P. M. Shearer, Using S/P amplitude ratios to constrain the focal mechanisms of small earthquakes, *Bull. Seismol. Soc. Am.*, *93*, 2434–2444, 2003.

Hauksson, E., Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California, *J. Geophys. Res.*, *105*, 13,875–13,903, 2000.

Hauksson, E., and J. Unruh, Regional tectonics of the Coso geothermal area along the intracontinental plate boundary in central eastern California: Three-dimensional V_p and V_p/V_s models, spatial-temporal seismicity patterns, and seismogenic deformation, *J. Geophys. Res.*, *112*, 2007.

Humphreys, E. D., and R. W. Clayton, Tomographic image of the southern California mantle, *J. Geophys. Res.*, *95*, 19,725–19,746, 1990.

Hung, S.-H., F. A. Dahlen, and G. Nolet, Fréchet kernels for finite-frequency traveltimes—II. Examples, *Geophys. J. Int.*, *141*, 175–203, 2000.

Jarrard, R. D., Relations among subduction parameters, *Rev. Geophys.*, *24*, 217–284, 1986.

Jennings, C. W., Fault activity map of California and adjacent areas, with locations and ages of recent volcanic eruptions, Calif. Div. Mines and Geology, 1994, Geologic Data Map No. 6, map scale 1:750,000.

Kanamori, H., and D. Hadley, Crustal structure and temporal velocity change in southern California, *Pure App. Geophys.*, *113*, 257–280, 1975.

Kennett, B. L. N., M. S. Sambridge, and P. R. Williamson, Subspace methods for large inverse problems with multiple parameter classes, *Geophys. J. Int.*, *94*, 237–247, 1988.

Kiyashchenko, D., R.-E. Plessix, B. Kashtan, and V. Troyan, A modified imaging principle for true-amplitude wave-equation migration, *Geophys. J. Int.*, *168*, 1093–1104, 2007.

Komatitsch, D., and J. Tromp, Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.*, *139*, 806–822, 1999.

Komatitsch, D., and J. Tromp, Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.*, *149*, 390–412, 2002a.

Komatitsch, D., and J. Tromp, Spectral-element simulations of global seismic wave propagation—II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys. J. Int.*, *150*, 308–318, 2002b.

Komatitsch, D., and J.-P. Vilotte, The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. Seismol. Soc. Am.*, *88*, 368–392, 1998.

Komatitsch, D., J. Ritsema, and J. Tromp, The spectral-element method, Beowulf computing, and global seismology, *Science*, *298*, 1737–1742, 2002.

Komatitsch, D., Q. Liu, J. Tromp, P. Süss, C. Stidham, and J. H. Shaw, Simulations of ground motion in the Los Angeles basin based upon the spectral-element method, *Bull. Seismol. Soc. Am.*, *94*, 187–206, 2004.

Laske, G., and G. Masters, Constraints on global phase velocity maps from long-period polarization data, *J. Geophys. Res.*, *101*, 16,059–16,075, 1996.

Lee, S.-J., H.-W. Chen, Q. Liu, D. Komatitsch, B.-S. Huang, and J. Tromp, Three-dimensional simulations of seismic-wave propagation in the Taipei basin with realistic topography based upon the spectral-element method, *Bull. Seismol. Soc. Am.*, *98*, 253–264, 2008.

Lin, G., P. Shearer, and Y. Fialko, Obtaining absolute locations for quarry seismicity using remote sensing data, *Bull. Seismol. Soc. Am.*, *96*, 722–728, 2006.

Lin, G., P. M. Shearer, and E. Hauksson, Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005, *J. Geophys. Res.*, *112*, 2007a.

Lin, G., P. M. Shearer, E. Hauksson, and C. H. Thurber, A three-dimensional crustal seismic velocity model for southern California from a composite event method, *J. Geophys. Res.*, *112*, 2007b.

Liu, Q., and J. Tromp, Finite-frequency kernels based on adjoint methods, *Bull. Seismol. Soc. Am.*, *96*, 2383–2397, 2006.

Liu, Q., and J. Tromp, Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, *174*, 265–286, 2008.

Liu, Q., J. Polet, D. Komatitsch, and J. Tromp, Spectral-element moment tensor inversions for earthquakes in southern California, *Bull. Seismol. Soc. Am.*, *94*, 1748–1761, 2004.

Lohman, R. B., and J. J. McGuire, Earthquake swarms driven by aseismic creep in the Salton Trough, California, *J. Geophys. Res.*, *112*, 2007, B04405, doi:10.1029/2006JB004596.

Lovely, P., J. H. Shaw, Q. Liu, and J. Tromp, A structural V_P model of the Salton Trough, California, and its implications for seismic hazard, *Bull. Seismol. Soc. Am.*, *96*, 1882–1896, 2006.

Luffi, P., J. B. Saleeby, C.-T. A. Lee, and M. N. Ducea, Lithospheric mantle duplex beneath the central Mojave Desert revealed by xenoliths from Dish Hill, California, *J. Geophys. Res.*, *114*, 2009.

Luo, Y., and G. Schuster, Parsimonious staggered grid finite-differencing of the wave equation, *Geophys. Res. Lett.*, *17*, 155–158, 1990.

Luyendyk, B. P., M. K. Kamerling, and R. Terres, Geometric model for Neogene crustal rotations in southern California, *Geol. Soc. Am. Bull.*, *91*, 211–217, 1980.

Ma, S., R. J. Archuleta, and M. T. Page, Effects of large-scale surface topography on ground motions, as demonstrated by a study of the San Gabriel Mountains, Los Angeles, California, *Bull. Seismol. Soc. Am.*, *97*, 2066–2079, 2007.

Maggi, A., C. Tape, M. Chen, D. Chao, and J. Tromp, An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, 2009, The automated algorithm FLEXWIN is available for download at www.geodynamics.org.

Magistrale, H., S. Day, R. W. Clayton, and R. Graves, The SCEC Southern California reference three-dimensional velocity model Version 2, *Bull. Seismol. Soc. Am.*, *90*, S65–S76, 2000.

Marquering, H., F. A. Dahlen, and G. Nolet, Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, *137*, 805–815, 1999.

McLaren, M. K., J. L. Hardebeck, N. van der Elst, J. Unruh, G. W. Bawden, and J. L. Blai, Complex faulting associated with the 22 December 2003 M_W 6.5 San Simeon, California, earthquake, aftershocks, and postseismic surface deformation, *Bull. Seismol. Soc. Am.*, *98*, 1659–1680, 2008.

McMechan, G. A., Determination of source parameters by wavefield extrapolation, *Geophys. J. R. Astron. Soc.*, *71*, 613–628, 1982.

Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung, Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, *303*, 338–343, 2004.

Mora, P., Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, *52*, 1211–1228, 1987.

Mori, J., and D. Helmberger, Large-amplitude Moho reflections (SmS) from Landers aftershocks, southern California, *Bull. Seismol. Soc. Am.*, *86*, 1845–1852, 1996.

Nadin, E. S., and J. B. Saleeby, Quaternary reactivation of the Kern Canyon fault system, southern Sierra Nevada, California, USA (in review), *Geol. Soc. Am. Bull.*, 2009.

Ni, S., E. Tan, M. Gurnis, and D. Helmberger, Sharp sides to the African superplume, *Science*, *296*, 1850–1852, 2002.

Nicholson, C., C. C. Sorlien, T. Atwater, J. C. Crowell, and B. P. Luyendyk, Microplate capture, rotation of the western Transverse Ranges, and initiation of the San Andreas transform as a low-angle fault system, *Geology*, *22*, 491–495, 1994.

Nolet, G., Waveform tomography, in *Seismic Tomography: With Applications in Global Seismology and Exploration Geophysics*, edited by G. Nolet, pp. 301–322, Reidel Publishing, Dordrecht, The Netherlands, 1987.

Olsen, K. B., S. M. Day, J. B. Minster, Y. Cui, A. Chourasia, M. Faerman, R. Moore, P. Maechling, and T. Jordan, Strong shaking in Los Angeles expected from southern San Andreas earthquake, *Geophys. Res. Lett.*, *33*, 2006.

Pasyanos, M. E., D. S. Dreger, and B. Romanowicz, Toward real-time estimation of regional moment tensors, *Bull. Seismol. Soc. Am.*, *86*, 1255–1269, 1996.

Paulssen, H., Crustal anisotropy in southern California from local earthquake data, *Geophys. Res. Lett.*, *31*, 2004.

Percival, D., and A. Walden, *Spectral Analysis for Physical Applications*, Cambridge U. Press, 1993.

Peter, D., C. Tape, L. Boschi, and J. H. Woodhouse, Surface wave tomography: global membrane waves and adjoint methods, *Geophys. J. Int.*, *171*, 1098–1117, 2007.

Pratt, R. G., Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics*, *64*, 888–901, 1999.

Pratt, R. G., C. S. Shin, and G. J. Hicks, Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion, *Geophys. J. Int.*, *133*, 341–362, 1998.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge U. Press, Cambridge, 1994.

Rawlinson, N., G. A. Houseman, and C. D. N. Collins, Inversion of seismic refraction and wide-angle reflection traveltimes for three-dimensional layered crustal structure, *Geophys. J. Int.*, *145*, 381–400, 2001.

Ritsema, J., H. J. van Heijst, and J. H. Woodhouse, Complex shear velocity structure imaged beneath Africa and Iceland, *Science*, *286*, 1925–1928, 1999.

Rodgers, A., A. Petersson, S. Nilsson, B. Sjögreen, and K. McCandless, Broadband waveform modeling of moderate earthquakes in the San Francisco Bay area and preliminary assessment of the USGS 3D seismic velocity model, *Bull. Seismol. Soc. Am.*, *98*, 969–988, 2008.

Romanowicz, B., Global mantle tomography: Progress status in the past 10 years, *Annu. Rev. Earth Planet. Sci.*, *31*, 303–328, 2003.

Ruff, L., and H. Kanamori, Seismicity and the subduction process, *Phys. Earth Planet. Inter.*, *32*, 240–252, 1980.

Sambridge, M. S., Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D, *Geophys. J. Int.*, *102*, 653–677, 1990.

Sambridge, M. S., A. Tarantola, and B. L. N. Kennett, An alternative strategy for non-linear inversion of seismic waveforms, *Geophys. Prosp.*, *39*, 723–736, 1991.

Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller, High-resolution surface-wave tomography from ambient seismic noise, *Science*, *307*, 1615–1618, 2005.

Sieminski, A., Q. Liu, J. Trampert, and J. Tromp, Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, *168*, 1153–1174, 2007.

Slepian, D., Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V: The discrete case, *Bell Syst. Tech.*, *57*, 1371–1430, 1978.

Stich, D., P. Danecek, A. Morelli, and J. Tromp, Imaging lateral heterogeneity in the northern Apennines from time reversal of reflected surface waves, *Geophys. J. Int.*, 2009.

Süss, M. P., and J. H. Shaw, P-wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *J. Geophys. Res.*, *108*, 2003.

Takeuchi, N., and M. Kobayashi, Improvement of seismological earth models by using data weighting in waveform inversion, *Geophys. J. Int.*, *158*, 681–694, 2004.

Talagrand, O., and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328, 1987.

Tan, Y., Broadband Waveform Modeling Over a Dense Seismic Network, Ph.D. thesis, California Institute of Technology, 2006.

Tanimoto, T., Modelling curved surface wave paths: membrane surface wave synthetics, *Geophys. J. Int.*, *102*, 89–100, 1990.

Tape, C., Q. Liu, and J. Tromp, Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves, *Geophys. J. Int.*, *168*, 1105–1129, 2007.

Tape, C., P. Musé, M. Simons, D. Dong, and F. Webb, Multiscale estimation of GPS velocity fields, *Geophys. J. Int.(in review)*, 2009.

Tarantola, A., Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, *49*, 1259–1266, 1984.

Tarantola, A., A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, *51*, 1893–1903, 1986.

Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, Penn., 2005.

Thomson, D. J., Spectrum estimation and harmonic analysis, *IEEE Proc.*, *70*, 1055–1096, 1982.

Thurber, C., H. Zhang, F. Waldhauser, J. Hardebec, A. Michael, and D. Eberhart-Phillips, Three-dimensional compressional wavespeed model, earthquake relocations, and focal mechanisms for the Parkfield, California, region, *Bull. Seismol. Soc. Am.*, *96*, S38–S49, 2006.

Tromp, J., C. Tape, and Q. Liu, Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels, *Geophys. J. Int.*, *160*, 195–216, 2005.

Wald, L. A., L. K. Hutton, and D. D. Given, The Southern California Network Bulletin: 1990–1993 summary, *Seis. Res. Lett.*, *66*, 9–19, 1995.

Waldhauser, F., and W. L. Ellsworth, A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California, *Bull. Seismol. Soc. Am.*, *90*, 1353–1368, 2000.

Wang, Z., and F. A. Dahlen, Spherical-spline parameterization of three-dimensional Earth, *Geophys. Res. Lett.*, *22*, 3099–3102, 1995.

Wang, Z., J. Tromp, and G. Ekström, Global and regional surface-wave inversion: A spherical-spline parameterization, *Geophys. Res. Lett.*, *25*, 207–210, 1998.

Woodhouse, J. H., and A. M. Dziewonski, Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, *89*, 5953–5986, 1984.

Yan, Z., and R. W. Clayton, Regional mapping of the crustal structure in southern California from receiver functions, *J. Geophys. Res.*, *112*, 2007, B05311, doi:10.1029/2006JB004622.

Zhao, L., and T. H. Jordan, Structural sensitivities of finite-frequency seismic waves: a full-wave approach, *Geophys. J. Int.*, *165*, 981–990, 2006.

Zhao, L., T. H. Jordan, and C. H. Chapman, Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.*, *141*, 558–576, 2000.

Zhao, L., T. H. Jordan, K. B. Olsen, and P. Chen, Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Bull. Seismol. Soc. Am.*, *95*, 2066–2080, 2005.

Zhao, L.-S., and D. V. Helmberger, Source estimation from broadband regional seismograms, *Bull. Seismol. Soc. Am.*, *84*, 91–104, 1994.

Zhou, Y., F. A. Dahlen, and G. Nolet, Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, *158*, 142–168, 2004.

Zhou, Y., F. A. Dahlen, G. Nolet, and G. Laske, Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.*, *163*, 1087–1111, 2005.

Zhu, L., and D. Helmberger, Advancement in source estimation techniques using broadband regional seismograms, *Bull. Seismol. Soc. Am.*, *86*, 1634–1641, 1996.

Zhu, L., and H. Kanamori, Moho depth variation in southern California from teleseismic receiver functions, *J. Geophys. Res.*, *105*, 2969–2980, 2000.