Solid Earth, 5, 1151–1168, 2014
www.solid-earth.net/5/1151/2014/
doi:10.5194/se-5-1151-2014
© Author(s) 2014. CC Attribution 3.0 License.

Solid Earth Open Access

Wave-equation-based travel-time seismic tomography – Part 1: Method

P. Tong[1], D. Zhao[2], D. Yang[3], X. Yang[4], J. Chen[4], and Q. Liu[1]

[1]Department of Physics, University of Toronto, Toronto, M5S 1A7, Ontario, Canada
[2]Department of Geophysics, Tohoku University, Sendai, Japan
[3]Department of Mathematical Sciences, Tsinghua University, Beijing, China
[4]Department of Mathematics, University of California, Santa Barbara, California, USA

Correspondence to: P. Tong (tongping85@gmail.com)

Received: 10 August 2014 – Published in Solid Earth Discuss.: 25 August 2014
Revised: 22 October 2014 – Accepted: 24 October 2014 – Published: 26 November 2014

Abstract. In this paper, we propose a wave-equation-based travel-time seismic tomography method with a detailed description of its step-by-step process. First, a linear relationship between the travel-time residual $\Delta t = T^{\mathrm{obs}} - T^{\mathrm{syn}}$ and the relative velocity perturbation $\delta c(\boldsymbol{x})/c(\boldsymbol{x})$ connected by a finite-frequency travel-time sensitivity kernel $K(\boldsymbol{x})$ is theoretically derived using the adjoint method. To accurately calculate the travel-time residual Δt, two automatic arrival-time picking techniques including the envelop energy ratio method and the combined ray and cross-correlation method are then developed to compute the arrival times T^{syn} for synthetic seismograms. The arrival times T^{obs} of observed seismograms are usually determined by manual hand picking in real applications. Travel-time sensitivity kernel $K(\boldsymbol{x})$ is constructed by convolving a forward wavefield $u(t,\boldsymbol{x})$ with an adjoint wavefield $q(t,\boldsymbol{x})$. The calculations of synthetic seismograms and sensitivity kernels rely on forward modeling. To make it computationally feasible for tomographic problems involving a large number of seismic records, the forward problem is solved in the two-dimensional (2-D) vertical plane passing through the source and the receiver by a high-order central difference method. The final model is parameterized on 3-D regular grid (inversion) nodes with variable spacings, while model values on each 2-D forward modeling node are linearly interpolated by the values at its eight surrounding 3-D inversion grid nodes. Finally, the tomographic inverse problem is formulated as a regularized optimization problem, which can be iteratively solved by either the LSQR solver or a nonlinear conjugate-gradient method. To provide some insights into future 3-D tomographic inversions, Fréchet kernels for different seismic phases are also demonstrated in this study.

1 Introduction

Seismic tomography is one of the core methodologies for imaging the structural heterogeneity of the Earth's interior at a variety of scales. Ever since the pioneering works of Aki and Lee (1976) and Dziewonski et al. (1977), tomographic images have provided crucial information for the understanding of plate tectonics, volcanism, and geodynamics (e.g., Romanowicz, 1991; Liu and Gu, 2012; Zhao, 2012). Seismic tomography itself has also gone through significant development over the last 3 decades, including advances in both methodology and data usage.

In the first 2 decades of its history, seismic tomography was mainly based on the ray theory which assumes that seismic travel-time is determined by the structure along the infinitely thin ray path only. However, because of scattering, wave front healing, and other finite-frequency effects, seismic measurements (such as travel-time and amplitude), especially those made on broadband recordings, are sensitive to three-dimensional (3-D) structures off the ray path (e.g., Marquering et al., 1999; Dahlen et al., 2000; Tape et al., 2007). Ray theory is actually only valid when the scale length of the variation of material properties is much larger than the seismic wavelength (Rawlinson et al., 2010). To take into account the sensitivity to off-ray structures, finite-frequency tomography methods that construct 2-D or 3-D travel-time and

amplitude sensitivity kernels are proposed, including those based on the paraxial approximation and dynamic ray tracing (e.g., Marquering et al., 1999; Dahlen et al., 2000; Tian et al., 2007; Tong et al., 2011) and those based on the normal mode theory (e.g., Zhao et al., 2000; Zhao and Jordan, 2006; To and Romanowicz, 2009). Tomographic models with improved resolutions were reported by recent finite-frequency tomographic studies (e.g., Montelli et al., 2004; Hung et al., 2004, 2011; Gautier et al., 2008), although comparison to ray-based tomography remains controversial (de Hoop and van der Hilst, 2005a; Dahlen and Nolet, 2005; de Hoop and van der Hilst, 2005b). The underlying problem of the finite-frequency tomography based on paraxial approximation and dynamic ray tracing is that its kernel computation still relies on the ray theory, although it was devised to account for non-geometrical finite-frequency phenomena. In the last decade or so, rapid advances in high-performance computing and forward modeling techniques have made it feasible to solve the seismic wave equations in realistic Earth models by fully numerical methods (e.g., Komatitsch and Tromp, 2002a, b; Komatitsch et al., 2004; Operto et al., 2007). This opens the way to compute sensitivity kernels based on numerical simulation of the full seismic wavefield, avoiding the use of approximate theories (e.g., Liu and Tromp, 2006, 2008; Fichtner et al., 2009). It also made the conceptual wave-equation-based seismic inversion methods such as the one presented by Tarantola (1984) feasible in realistic applications (Tape et al., 2009; Fichtner and Trampert, 2011; Zhu et al., 2012). To our best knowledge, adjoint tomography (Tromp et al., 2005; Fichtner et al., 2006), scattering integral methods (L. Zhao et al., 2005; Chen et al., 2007b), and full waveform inversion (FWI) in the frequency domain (Pratt and Shipp, 1999; Operto et al., 2006) are among the most popular tomographic techniques based upon solving full wave equations. FWI in the frequency domain has been mainly used in exploration problems (e.g., Virieux and Operto, 2009; Lee et al., 2010). Adjoint tomography and scattering integral tomography are closely related to each other, and a detailed comparison between adjoint tomography and scattering integral tomography can be found in Chen et al. (2007a). For brevity, we restrict our following discussions to adjoint tomography (Liu and Gu, 2012).

Adjoint tomography is currently one of the most popular and promising tomographic methods for resolving strongly varying structures. It takes advantage of full 3-D numerical simulations in forward modeling and sensitivity kernel calculation, often iteratively improving models through optimization techniques (Tromp et al., 2005; Tape et al., 2007). The use of full numerical simulations allows for the freedom of choosing either 1-D or 3-D reference models and accurate calculations of seismograms (Tong et al., 2014a, c) and sensitivity kernels for complex models (Liu and Tromp, 2006, 2008). Using this approach, Tape et al. (2009, 2010) obtained a 3-D velocity model of the southern California crust that captures strong local heterogeneity up to $\pm 30\,\%$. Similarly, Zhu et al. (2012) generated a tomographic model of the European upper mantle based on adjoint tomography that reveals nice correlations between structural features and regional tectonics and dynamics. Similarly, Rickers et al. (2013) presented a 3-D *S* wave velocity model of the North Atlantic region, revealing structural features in unprecedented detail down to a depth of 1300 km. These successful applications reveal the promising future of next-generation seismic tomographic models based on full numerical simulations. However, the expensive computation cost associated with adjoint-type of wave-equation-based tomographic methods, especially for 3-D problems, is still a major stumbling block to its wider application. For example, for a moderate number of three-component seismograms, 0.8 million and 2.3 million central processing unit hours were used to generate the tomographic models of the southern California crust and the European upper mantle, respectively (Tape et al., 2009; Zhu et al., 2012). The severity of the cost issue may be remedied when simulations are ported to the graphic processing unit (GPU) hardware (e.g., Komatitsch et al., 2010; Michéa and Komatitsch, 2010). However, ray-based tomographic methods remains the most popular and accessible techniques for mapping the heterogeneous structures of the Earth's interior (e.g., Li et al., 2008; Hung et al., 2011; Tong et al., 2012; Zhao et al., 2012).

As mentioned above, full 3-D numerical simulations in forward modeling and sensitivity kernel calculations guarantee the accuracy of synthetic seismograms and sensitivity kernels for 3-D complex models. However, they also make adjoint tomography computationally demanding and even unaffordable. To strike a balance between the computational efficiency and accuracy of full wave-equation-based tomographic methods, we propose conducting the forward modeling and sensitivity kernel calculation in the 2-D source-receiver vertical plane by a high-order finite-difference scheme. As we will show, if only travel-time measurements are considered, this 2-D approximation offers acceptable accuracy. Meanwhile, by numerically solving 2-D wave equations, finite-frequency effects such as wavefront healing are naturally taken into account, and the accuracy of sensitivity kernels in complex heterogeneous models is also improved. Although forward modelings are restricted to 2-D planes, we still plan to invert for 3-D tomographic models on a 3-D inversion grid. The 2-D forwarding modeling and the 3-D tomographic inversion are linked by expressing the model parameters (such as velocity perturbation) at each 2-D forward modeling grid node as a linear interpolation of the model parameters at its surrounding 3-D inversion grid nodes. We name the resultant 2-D–3-D tomographic method the wave-equation-based travel-time seismic tomography (WETST). Compared with the 3-D–3-D adjoint tomography based on the spectral element method (Tromp et al., 2005; Fichtner et al., 2006), this 2-D–3-D WETST based upon a 2-D finite-difference scheme is generally more computationally affordable. This also entails that WETST

can be applied to tomographic inversions involving significant amounts of data based on even moderate computational resources.

Arrival time picking is another important issue for travel-time seismic tomography. Since the early era of ray-based seismic tomography, researchers have mainly relied on manually picked arrival times to map subsurface structures (e.g., Aki and Lee, 1976; Zhao et al., 1992). Arrival times are usually picked within time windows centered at the predicted travel times (Kennett and Engdah, 1991; Maggi et al., 2009). In recent years, increasing numbers of deployed broadband seismic arrays have resulted in the proliferation of seismic data. To increase efficiency and reduce the amount of manual labor and human errors in seismic data processing, fast and automatic travel-time picking algorithms with high accuracy are highly demanded in order to process vast amount of seismic recordings. Indeed, various techniques have been presented for the automatic/semi-automatic detecting and picking of the arrivals of different seismic phases, and the most widely used of which is the short-term-average (STA) to long-term-average (LTA) ratio method and its variations (e.g., Coppens, 1985; Baer and Kradolfer, 1987; Saari, 1991; Earle and Shearer, 1994; Han et al., 2010). Zhang et al. (2003) developed an automatic P wave arrival detection and picking algorithm based on the wavelet transform and Akaike information criteria. The cross-correlation method is another routinely used technique to obtain the travel-time anomalies of broadband pulses, which is especially favored by finite-frequency tomographic applications (e.g., Luo and Schuster, 1991; Dahlen et al., 2000; Tape et al., 2007). However, the quality of picked arrivals by these methods may vary in accuracy for data sets of different signal-to-noise ratio (SNR), and often only arrivals on low-noise seismograms can be effectively picked (Akram, 2011). Specifically, the validity of the correlation-based methods requires that the synthetic seismograms be reasonably similar to the observed seismograms. Less restrictive automatic arrival picking algorithms need to be further developed. In this study, we propose two different automatic arrival-time determination methods (Sect. 3) that form an integral part of our wave-equation-based travel-time seismic tomography method.

When arrival-time data and sensitivity kernels are determined or computed, wave-equation-based travel-time seismic tomography is cast as an optimization problem. Model parameterization, regularization, and methods solving the optimization problem are discussed in Sects. 4 and 5. Finally, examples of sensitivity kernels for different seismic waves are shown in Sect. 6, which provide the basis for future tomographic inversions with various seismic phases. This paper focuses on theoretical derivation of the wave-equation-based travel-time seismic tomography. An application of the WETST method is presented in the second paper (Tong et al., 2014b).

2 Tomographic equation

In this section, we set up a linear relationship between the perturbation of arrival time and velocity perturbation in a reference model.

2.1 Travel-time residual

Travel-time seismic tomography generally inverts travel-time residuals of some seismic phases to map internal Earth structures. A travel-time residual Δt corresponding to the event occurred at $\boldsymbol{x}_\mathrm{s}$, and the seismic station located at $\boldsymbol{x}_\mathrm{r}$ is written as,

$$\Delta t = T^{\mathrm{obs}} - T^{\mathrm{syn}}, \tag{1}$$

where the observed travel-time T^{obs} is automatically or manually picked on recorded seismogram $d(t)$, and the synthetic arrival time T^{syn} is predicted based on a reference model. In geometrical ray theory, T^{syn} is usually computed by integrating the slowness along a travelling path.

If the corresponding synthetic seismogram $u(t)$ in the reference model is available, the travel-time residual Δt can be approximated by the cross-correlation technique (Dahlen et al., 2000):

$$\Delta t \approx \frac{1}{N_\mathrm{r}} \int_0^T w(t)\dot{u}(t)\left[\mathrm{d}(t) - u(t)\right]\mathrm{d}t, \tag{2}$$

where

$$N_\mathrm{r} = \int_0^T w(t)u(t)\ddot{u}(t)\mathrm{d}t,$$

and $w(t)$ is a weight function over the time interval $[0, T]$ that can be used to isolate particular seismic phases (Tromp et al., 2005). The accuracy of this approximation improves as data and synthetic pulse becomes more similar, i.e., waveform perturbation $d(t) - u(t)$ in Eq. (2) becomes tiny. Assuming infinitesimal perturbations, Eq. (2) becomes

$$\delta t = \frac{1}{N_\mathrm{r}} \int_0^T w(t)\dot{u}(t)\delta u(t)\,\mathrm{d}t, \tag{3}$$

which is used further to set up the relationship between travel-time residual and
velocity perturbation.

2.2 Relationship between travel-time residual and velocity perturbation

We consider seismic wave propagation in a two-dimensional (2-D) vertical plane which contains the source $\boldsymbol{x}_\mathrm{s}$ and the receiver $\boldsymbol{x}_\mathrm{r}$. Within this plane, the seismic wavefield of a particular phase (without mode conversion) could be assumed

to satisfy the 2-D acoustic wave equation with initial and boundary conditions,

$$\begin{cases} \frac{\partial^2}{\partial t^2}u(t,\boldsymbol{x}) = \nabla\cdot\left[c^2(\boldsymbol{x})\nabla u(t,\boldsymbol{x})\right] + f(t)\delta(\boldsymbol{x}-\boldsymbol{x}_\mathrm{s}), \quad \boldsymbol{x}\in S \\ u(0,\boldsymbol{x}) = \frac{\partial u(0,\boldsymbol{x})}{\partial t} = 0, \quad \boldsymbol{x}\in S, \\ \hat{\boldsymbol{n}}\cdot\left[c^2(\boldsymbol{x})\nabla u(t,\boldsymbol{x})\right] = 0, \quad \boldsymbol{x}\in\partial S, \end{cases} \tag{4}$$

where $u(t,\boldsymbol{x})$ is the displacement field, $c(\boldsymbol{x})$ is either the P or S wave velocity model, $f(t)$ is the source time function for the point source at $\boldsymbol{x}_\mathrm{s}$, and $\hat{\boldsymbol{n}}$ is the normal direction of the boundary ∂S. For a perturbation $\delta c(\boldsymbol{x})$ of the velocity model $c(\boldsymbol{x})$, a consequent perturbed displacement wavefield $\delta u(t,\boldsymbol{x})$ will be generated. In the framework of first-order or Born approximation (e.g., Aki and Richards, 2002; Tromp et al., 2005; Tong et al., 2011), the perturbed wavefield $\delta u(t,\boldsymbol{x})$ is the solution to the following wave equation with subsidiary conditions:

$$\begin{cases} \frac{\partial^2}{\partial t^2}\delta u(t,\boldsymbol{x}) = \nabla\cdot\left[c^2(\boldsymbol{x})\nabla\delta u(t,\boldsymbol{x}) + 2c(\boldsymbol{x})\delta c(\boldsymbol{x})\nabla u(t,\boldsymbol{x})\right], \quad \boldsymbol{x}\in S, \\ \delta u(0,\boldsymbol{x}) = \frac{\partial \delta u(0,\boldsymbol{x})}{\partial t} = 0, \quad \boldsymbol{x}\in S, \\ \hat{\boldsymbol{n}}\cdot\left[c^2(\boldsymbol{x})\nabla\delta u(t,\boldsymbol{x}) + 2c(\boldsymbol{x})\delta c(\boldsymbol{x})\nabla u(t,\boldsymbol{x})\right] = 0, \quad \boldsymbol{x}\in\partial S. \end{cases} \tag{5}$$

Multiplying an arbitrary test function $q(t,\boldsymbol{x})$ on both sides of the first equation in Eq. (5) and then integrating in the surface S and the time interval $[0,T]$ gives us

$$\begin{aligned} &\int_0^T \mathrm{d}t \int_S q(t,\boldsymbol{x})\frac{\partial^2}{\partial t^2}\delta u(t,\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &= \int_0^T \mathrm{d}t \int_S q(t,\boldsymbol{x})\nabla \\ &\cdot\left[c^2(\boldsymbol{x})\nabla\delta u(t,\boldsymbol{x}) + 2c(\boldsymbol{x})\delta c(\boldsymbol{x})\nabla u(t,\boldsymbol{x})\right]\mathrm{d}\boldsymbol{x}, \end{aligned} \tag{6}$$

which is equal to

$$\begin{aligned} &\int_S \mathrm{d}\boldsymbol{x}\int_0^T\left\{\frac{\partial}{\partial t}\left[q(t,\boldsymbol{x})\frac{\partial}{\partial t}\delta u(t,\boldsymbol{x}) - \delta u(t,\boldsymbol{x})\frac{\partial}{\partial t}q(t,\boldsymbol{x})\right]\right. \\ &\left.+\delta u(t,\boldsymbol{x})\frac{\partial^2}{\partial t^2}q(t,\boldsymbol{x})\right\}\mathrm{d}t \\ &= \int_0^T \mathrm{d}t\int_S \delta u(t,\boldsymbol{x})\nabla \\ &\cdot\left[c^2(\boldsymbol{x})\nabla q(t,\boldsymbol{x})\right]\mathrm{d}\boldsymbol{x} - \int_0^T \mathrm{d}t\int_S \nabla \\ &\cdot\left[\delta u(t,\boldsymbol{x})c^2(\boldsymbol{x})\nabla q(t,\boldsymbol{x})\right]\mathrm{d}\boldsymbol{x} \\ &+\int_0^T \mathrm{d}t\int_S \nabla\cdot\{q(t,\boldsymbol{x}) \\ &\left[c^2(\boldsymbol{x})\nabla\delta u(t,\boldsymbol{x}) + 2c(\boldsymbol{x})\delta c(\boldsymbol{x})\nabla u(t,\boldsymbol{x})\right]\}\mathrm{d}\boldsymbol{x} \\ &-\int_0^T \mathrm{d}t\int_S 2c(\boldsymbol{x})\delta c(\boldsymbol{x})\nabla q(t,\boldsymbol{x})\cdot\nabla u(t,\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \end{aligned} \tag{7}$$

As travel-time residual δt in Eq. (3) is measured at the receiver location $\boldsymbol{x}_\mathrm{r}$, Eq. (3) can be alternatively expressed as

$$\delta t = \frac{1}{N_\mathrm{r}}\int_0^T w(t)\int_S \frac{\partial u(t,\boldsymbol{x})}{\partial t}\delta u(t,\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{x}_\mathrm{r})\mathrm{d}\boldsymbol{x}\mathrm{d}t. \tag{8}$$

Summing up Eq. (7) and Eq. (8), using the second and third relationships in Eq. (5), and assuming that

$$\begin{cases} \frac{\partial^2}{\partial t^2}q(t,\boldsymbol{x}) - \nabla\cdot\left[c^2(\boldsymbol{x})\nabla q(t,\boldsymbol{x})\right] = \frac{1}{N_\mathrm{r}}w(t)\frac{\partial u(t,\boldsymbol{x})}{\partial t}\delta(\boldsymbol{x}-\boldsymbol{x}_\mathrm{r}), \quad \boldsymbol{x}\in S, \\ q(T,\boldsymbol{x}) = \frac{\partial q(T,\boldsymbol{x})}{\partial t} = 0, \quad \boldsymbol{x}\in S, \\ \hat{\boldsymbol{n}}\cdot c^2(\boldsymbol{x})\nabla q(t,\boldsymbol{x}) = 0, \quad \boldsymbol{x}\in\partial S, \end{cases} \tag{9}$$

we can get a relationship

$$\delta t = -\int_0^T \mathrm{d}t\int_S\left[2c^2(\boldsymbol{x})\nabla q(t,\boldsymbol{x})\cdot\nabla u(t,\boldsymbol{x})\right]\frac{\delta c(\boldsymbol{x})}{c(\boldsymbol{x})}\mathrm{d}\boldsymbol{x}. \tag{10}$$

Note that $q(t,\boldsymbol{x})$ is not longer an arbitrary function. Instead, it satisfies Eq. (9) and represents a wavefield generated reversely in time by backpropagating the windowed and normalized velocity signal recorded at the receiver for the velocity model $c(\boldsymbol{x})$, also known as the adjoint wavefield (e.g., Liu et al., 2004; Tromp et al., 2005; Fichtner et al., 2006).

By defining the travel-time sensitivity kernel

$$K(\boldsymbol{x};\boldsymbol{x}_\mathrm{r},\boldsymbol{x}_\mathrm{s}) = -\int_0^T\left[2c^2(\boldsymbol{x})\nabla q(t,\boldsymbol{x})\cdot\nabla u(t,\boldsymbol{x})\right]\mathrm{d}t, \tag{11}$$

Eq. (10) provides a concise mathematical expression of the relationship between travel-time residual δt and relative velocity perturbation $\delta c(\boldsymbol{x})/c(\boldsymbol{x})$

$$\delta t = \int_{\Omega} K(\boldsymbol{x}; \boldsymbol{x}_{\mathrm{r}}, \boldsymbol{x}_{\mathrm{s}}) \frac{\delta c(\boldsymbol{x})}{c(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}. \tag{12}$$

The travel-time kernel $K(\boldsymbol{x}; \boldsymbol{x}_{\mathrm{r}}, \boldsymbol{x}_{\mathrm{s}})$ is a weighted convolution of forward wavefield gradient $\nabla u(t, \boldsymbol{x})$ and the adjoint wavefield gradient $\nabla q(t, \boldsymbol{x})$, which can be obtained by solving Eqs. (4) and (9). Assuming small perturbations, we can set Δt in Eq. (1) equal to δt, and Eq. (12) becomes

$$T^{\mathrm{obs}} - T^{\mathrm{syn}} = \int_{\Omega} K(\boldsymbol{x}; \boldsymbol{x}_{\mathrm{r}}, \boldsymbol{x}_{\mathrm{s}}) \frac{\delta c(\boldsymbol{x})}{c(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}. \tag{13}$$

We call relation (13) the *tomographic equation* of wave-equation-based travel-time seismic tomography. Once the observed arrival time T^{obs} and synthetic arrival time T^{syn} are measured or calculated, tomographic Eq. (13) can be inverted to infer the relative velocity perturbation $\delta c(\boldsymbol{x})/c(\boldsymbol{x})$.

Before proceeding to the next section, we should keep in mind that $T^{\mathrm{obs}} - T^{\mathrm{syn}}$ in Eq. (13) is theoretically needed to be a finite-frequency travel-time residual. Due to the finite-frequency and dispersive effects, seismic waves of different frequencies may have different sensitivities to the subsurface structures and arrive at different times. To measure physically meaningful travel-time residual $T^{\mathrm{obs}} - T^{\mathrm{syn}}$, we should first ensure that the observed data $d(t)$ are filtered through a frequency band, which is generally selected to be consistent with the frequency spectrum of the synthetic seismogram $s(t)$. Hereinafter, $d(t)$ will represent the bandpass-filtered data.

3 Arrival time picking

We first discuss how to pick the arrival times of a particular seismic phase on observed and synthetic seismograms, i.e., T^{obs} and T^{syn} in Eq. (13). Since any errors in arrival times will distort the velocity anomalies, this step is crucial for travel-time seismic tomography. Although manual arrival-picking is time-consuming and labor intensive, it is still one of the most reliable and stable techniques to determine the arrival times of specific seismic phases on observed seismograms. For example, the first-arrivals picked by analysts of the combined seismic network in Japan (known as the JMA Unified Catalogue) have accuracies of about 0.1 s for P arrival and 0.1–0.2 s for S arrival (Tong et al., 2012). Since there is not yet an automatic, accurate, and robust arrival we prefer to use manually picked arrival times T^{obs} on observed seismograms for tomographic inversion purpose.

Regarding the arrival time T^{syn} of a particular phase on synthetic seismograms, we could also use manual picking. However, extra subjective errors will be introduced into the travel-time residual Δt and further affect final tomographic results. Since synthetic seismograms are generated by numerical methods, the errors come mainly from numerical dispersion and can be controlled (but cannot be avoided) by employing accurate forward solver or fine meshes in forward numerical modeling. For low-noise seismograms, automatic time-picking schemes such as the STA/LTA method have been proven to be accurate and efficient for detecting the arrivals of different seismic phases (e.g., Saari, 1991; Han et al., 2010). In this study, we present a new envelope energy ratio method to pick up the arrival times on synthetic seismograms, which has a better performance than the STA/LTA method. On the other hand, if the starting model $\boldsymbol{m}_0$ for a tomographic inversion has (or is near) a simple geometry where traveling paths can be easily and accurately determined, the combined ray and cross-correlation method developed later can be used to obtain arrival times of particular phases on synthetic seismograms.

3.1 Envelop energy ratio method

We give a brief introduction to the STA/LTA method and then discuss the envelop energy ratio (EER) method, which is an improved version of the STA/LTA algorithm. Let $u(t)$ represent a seismogram with a dominant period of T_0 in the time window $[0, T]$, then the average energies in the short- and long-term windows preceding the time t are defined as

$$\begin{aligned} S(t) &= \frac{1}{\alpha T_0} \int_{t-\alpha T_0}^{t} u^2(\tau) \mathrm{d}\tau, \\ L(t) &= \frac{1}{\beta T_0} \int_{t-\beta T_0}^{t} u^2(\tau) \mathrm{d}\tau, \end{aligned} \tag{14}$$

where $t \in [0, T]$ and $0 < \alpha < \beta$ are coefficients determining the lengths of the short and long-term time windows and should be determined by the user. Usually, α and β are chosen to be $2 \le \alpha \le 3$ and $5 \le \beta \le 10$, respectively (Earle and Shearer, 1994). $u(\tau)$ is assumed to be zero for $\tau < 0$. Let us define the ratio

$$R(t) = \frac{S(t)}{L(t)}, \tag{15}$$

and if the ratio $R(t)$ first exceeds a user-defined threshold at t_0, t_0 is considered to be the approximate onset time of the first arrival on the seismogram $u(t)$ (Munro, 2004). It is also claimed that the maximum value of the derivative $\mathrm{d}R(t)/\mathrm{d}t$ may be closer to the break time of the first arrival (Wong et al., 2009).

Into the envelop function of seismogram $e(t) = |u(t) + iH[u(t)]|$, where $H[u(t)]$ denotes the Hilbert transform of $u(t)$, can be also used in seismic data analysis (e.g., Baer and Kradolfer, 1987; Maggi et al., 2009). Since the envelope function remains positive at zero crossings among different phase arrivals, average energy taken from an envelope

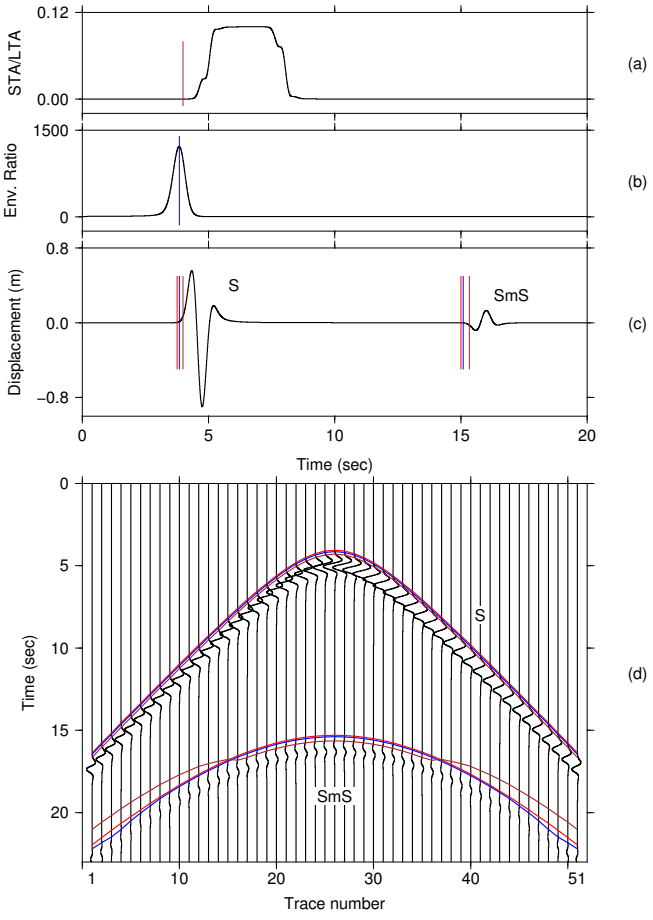

Figure 1. S arrival time-picking using **(a)** the STA/LTA method and **(b)** the envelop energy ratio (EER) method for the synthetic seismogram in **(c)**, which is the seismogram for trace number 26 in **(d)**. Panel **(d)** displays synthetic seismograms recorded by 51 stations with an equal spacing of 2 km at the surface, which are generated by an earthquake at the depth 12.0 km directly below the 26th station. The computational domain is a crust-over-mantle model. The crust has a thickness of 30.0 km and is homogeneous with the S wave velocity 3.2 km s^{-1} in the crust and 4.5 km s^{-1} in the mantle. In **(c)** and **(d)**, the arrival times of S and SmS phases determined based on the STA/LTA and EER methods are labeled with brown and blue lines, respectively. The theoretical arrivals are marked by red lines.

function may be a better measure of the signal strength (Baer and Kradolfer, 1987; Earle and Shearer, 1994). Meanwhile, Wong et al. (2009) proposed the modified energy method which has excellent performance in determining the break time of first arrival. By incorporating the envelope function and the modified energy method (Wong et al., 2009), we define the following envelop energy ratio function to determine the arrival time of the considered seismic phase filtered by the window function $w(t)$

$$r(t) = \frac{\int_{t-\beta T_0}^{t+\alpha T_0} w(\tau) e^2(\tau) \mathrm{d}\tau}{\int_{t-\gamma T_0}^{t} w(\tau) e^2(\tau) \mathrm{d}\tau}, \quad (16)$$

where $\alpha \geq 0$ and $\beta \geq \gamma \geq 1$. The peak of the ratio function $r(t)$ is very close to the onset time of the considered seismic phase.

To show the performance of the EER method, we apply it to shear-wave synthetic seismograms generated by an earthquake at 12.0 km depth in a homogeneous crust with a thickness of 30.0 km based on a high-order finite-difference method (in Appendix). Fifty-one surface stations with an equal spacing of 2.0 km are used to record seismograms. $\alpha = \beta = \gamma = 1.0$ are chosen in Eq. (16). Figure 1a–c shows the S wave arrival-time picking using the STA/LTA and EER methods on the seismogram for trace number 26 (Fig. 1d). We can see that S arrival time determined by the EER method is very close to the theoretical arrival time with an error smaller than 0.05 s (Fig. 1b and c). For the STA/LTA method, the threshold value is set to be 1.0×10^{-8}, and the obtained S arrival time is 0.24 s later than the theoretical arrival time (Fig. 1a and c). Note that an error of 0.24 s is unacceptable in travel-time inversion for local structures. We further show S and SmS arrival times on all 51 seismograms in Fig. 1d. For the direct S wave, results of both STA/LTA and EER methods are relatively close to the theoretical arrival times, with errors around 0.3 s and less than 0.1 s, respectively. However, for SmS phase, the STA/LTA algorithm is not able to give accurate estimates on the breaking times. In comparison, the EER method gives picked arrivals with accuracy similar to the direct S wave case, and 70 % of the errors are still less than 0.1 s. We have fixed all parameters for the STA/LTA and EER methods in picking the S and SmS arrival times. Actually, the accuracy of time picking on any single seismogram can be improved by slightly tuning some parameters, such as the threshold value for the STA/LTA method and the lengths of the time windows for both methods. We also find that the accuracy of the STA/LTA method is very sensitive to the threshold value and it is not an easy task to determine an appropriate threshold in practice. For the EER method, however, it is simple to locate the peak of the ratio function $r(t)$. This implies that the EER method could be a better choice for arrival-time picking on synthetic seismograms.

3.2 Combined ray and cross-correlation method

Because of the nonlinearity of seismic inverse problems, seismic tomography usually relies on an iterative method to find the optimal model. If the starting model $\boldsymbol{m}_0$ for travel-time seismic tomography is simple (e.g., a 1-D layered model) and traveling paths of particular phases can be easily traced, the arrival times T_0^{syn} of synthetics in $\boldsymbol{m}_0$ can be accurately determined based on ray theory. Meanwhile, we may expect that synthetic seismograms in the $(i+1)$th model $\boldsymbol{m}_{i+1}$ are reasonably similar to those in the ith model $\boldsymbol{m}_i$ $(i \geq 0)$, and the arrival-time shift $\delta t_{i+1,i}$ of a particular phase in models $\boldsymbol{m}_{i+1}$ and $\boldsymbol{m}_i$ can be calculated with high

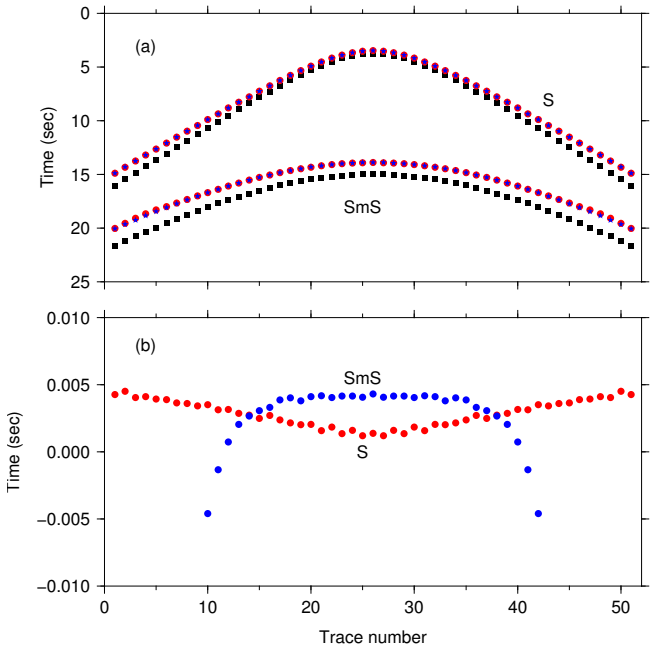

Figure 2. S and SmS arrival times on seismograms computed in two models, $\boldsymbol{m}_0$ and $\boldsymbol{m}_1$ **(a)**. Numerical computation in $\boldsymbol{m}_0$ is the same as the example shown in Fig. 1. S wave velocity in $\boldsymbol{m}_1$ has a perturbation of 8 % with respect to $\boldsymbol{m}_0$. Black squares, red circles, and blue stars correspond to theoretical arrival times in $\boldsymbol{m}_0$, theoretical arrival times in $\boldsymbol{m}_1$, and arrival times $\boldsymbol{m}_1$ computed by using the combined ray and cross-correlation method, respectively. Errors of S (red circles) and SmS (blue circles) arrival times determined by using the combined ray and cross-correlation method **(b)**.

accuracy by maximizing the cross-correlation equation,

$$\max_{\delta t_{i+1,i}} \frac{\int_0^T w(\tau)s(\tau;\boldsymbol{m}_{i+1})s(\tau-\delta t_{i+1,i};\boldsymbol{m}_i)\mathrm{d}\tau}{\left[\int_0^T w(\tau)s^2(\tau;\boldsymbol{m}_{i+1})\mathrm{d}\tau\int_0^T w(\tau)s^2(\tau-\delta t_{i+1,i};\boldsymbol{m}_i)\mathrm{d}\tau\right]^{1/2}}, \quad (17)$$

where $w(t)$ is the time window function used to isolate the considered phase (Liu et al., 2004). Consequently, the arrival time T_{i+1}^{syn} of the synthetic seismogram in model $\boldsymbol{m}_{i+1}$ satisfies the following relationship,

$$T_{i+1}^{\mathrm{syn}} = T_0^{\mathrm{syn}} + \sum_{j=0}^{i} \delta t_{j+1,j}. \quad (18)$$

Since T_0^{syn} and $\delta t_{j+1,j}$ are calculated with ray theory and cross-correlation method, respectively, Eq. (18) is called the combined ray and cross-correlation method.

Continuing the numerical example shown in the section of the EER method, we intend to verify the validity of the combined ray and cross-correlation method. Let $\boldsymbol{m}_0$ be the crust model with an S wave velocity of $3.2\,\mathrm{km\,s^{-1}}$ and a thickness of 30.0 km. S wave velocity in $\boldsymbol{m}_1$ is assumed to be $3.456\,\mathrm{km\,s^{-1}}$ which has a perturbation of 8.0 % with respect to $\boldsymbol{m}_0$. Synthetic seismograms generated by an earthquake at 12.0 km depth are calculated and recorded by 51 stations at the surface in both $\boldsymbol{m}_0$ and $\boldsymbol{m}_1$. For models $\boldsymbol{m}_0$ and $\boldsymbol{m}_1$, theoretical arrival times of S and SmS phases at each station can be calculated based on the ray theory (see solid red circles and black squares in Fig. 2a). Based on Eq. (17), we also measure arrival shifts of S and SmS in $\boldsymbol{m}_1$ from those in $\boldsymbol{m}_0$. Adding S and SmS arrival shifts to their corresponding arrival times for $\boldsymbol{m}_0$ (solid black squares in Fig. 2a), we get the approximated arrival times of S and SmS in model $\boldsymbol{m}_1$ (blue stars in Fig. 2a). Figure 2b shows the errors of the combined ray and cross-correlation method in determining the arrival times of S and SmS in model $\boldsymbol{m}_1$. It can be observed that the errors of direct S arrivals are less than 0.005 s, and 70 % errors of the SmS phases are smaller than 0.005 s with maximum error around about 0.175 s occurring at the 4th and 48th stations. Considering that the travel-time differences of the SmS phase in the two models are about 1.7 s at the two stations, these picking errors are relatively small. This numerical example suggests that the combined ray and cross-correlation method could serve as an efficient tool for high-accuracy arrival-time picking on synthetic seismograms in the iterative wave-equation-based tomographic inversions.

4 Model parameterization

Tomographic Eq. (13) needs invariably to be discretized for actual inversions (Nolet et al., 2005). This gives rise to model parameterization, which is an approximation to the true Earth structure. Model parameterization determines the accuracy of forward modeling and hence affects the final form of tomographic inversion results. Most commonly, functional approaches with a set of basis functions or an a prior functional form, such as cells and grid nodes, have been adopted to represent the Earth's structure (e.g., Dziewonski, 1984; Aki and Lee, 1976; Thurber, 1983). Each approach has its own advantages and drawbacks (e.g., Zhao, 2009; Rawlinson et al., 2010). To guarantee accurate computation of synthetic seismograms and travel-time kernels and to adapt to local variations in data coverage, we use two sets of grid nodes (i.e., forward modeling grid and inversion grid) to parameterize the Earth's structure for forward modeling and inversion algorithms in this study.

4.1 Forward modeling grid

As discussed in Sect. 2, we need to solve wave Eqs. (4) and (9) to obtain synthetic seismogram $u(t)$ and travel-time kernel $K(\boldsymbol{x})$. Many numerical methods such as the staggered-grid finite-difference (FD) method (e.g., Virieux, 1984; Graves, 1996) and spectral-element method (Komatitsch and Tromp, 1999) are well suited for this kind of forward modeling. In this study, we choose a FD scheme called high-order central difference method (see Appendix) to conduct forward modeling. The prominent feature of this high-order central difference method is that it simultaneously

computes the displacement $u(t,\boldsymbol{x})$ and the spatial gradient field $\nabla u(t,\boldsymbol{x})$, making the computation of the travel-time kernel $K(\boldsymbol{x})$ very straightforward. It is also easier to implement the high-order central difference method than the staggered-grid finite-difference (FD) method and spectral-element method. When sensitivity kernels are calculated by solving the full wave equation, there are spurious amplitudes in the immediate vicinity of the sources and receivers (Tape et al., 2007; Tong et al., 2014a). An efficient way of removing these spurious amplitudes is to smooth the travel-time kernel $K(\boldsymbol{x}) = K(x,z)$ with a 2-D Gaussian function

$$G(x,z) = \frac{4}{\pi\sigma_x\sigma_z} e^{-4\left(\frac{x^2}{\sigma_x^2}+\frac{z^2}{\sigma_z^2}\right)}, \tag{19}$$

where σ_x and σ_z are the averaging scale lengths along x and z directions, respectively. Usually, σ_x and σ_z are chosen to be less than the main wavelength of the seismic waves (Tape et al., 2007). The smoothed travel-time kernel $\widetilde{K}(x,z)$ is given by

$$\widetilde{K}(x,z) = \iint\limits_S K(x-x',z-z')G(x',z')\mathrm{d}x'\mathrm{d}z', \tag{20}$$

i.e., the smoothed kernel value at a given point is obtained by averaging the unsmoothed kernel values at its neighboring points.

For the 2-D FD numerical simulation, the continuous area S is sampled by a set of n discrete nodes $\boldsymbol{x}_i$ $(i = 1, 2, \cdots, n)$. By choosing a corresponding set of n basis functions $L_i(\boldsymbol{x})$ $(i = 1, 2, \cdots, n)$, the smoothed travel-time kernel $\widetilde{K}(\boldsymbol{x})$ and the relative velocity perturbation $\delta c(\boldsymbol{x})/c(\boldsymbol{x})$ can be expanded into linear combinations of the basis functions as

$$\widetilde{K}(\boldsymbol{x};\boldsymbol{x}_\mathrm{r},\boldsymbol{x}_\mathrm{s}) = \sum_{i=1}^{n} K_i L_i(\boldsymbol{x}), \tag{21}$$

where K_i and C_i are the corresponding coefficients related to the basis function $L_i(\boldsymbol{x})$. Substituting Eq. (22) into Eq. (13) results in the discrete form of the tomographic equation

$$\delta c(\boldsymbol{x})/c(\boldsymbol{x}) = \sum_{i=1}^{n} C_i L_i(\boldsymbol{x}), \tag{22}$$

A general way to define a basis function $L_i(\boldsymbol{x})$ is to construct a local interpolation function on knot node $\boldsymbol{x}_i$ and its neighbors. The possibility of different choices for the basis functions $L_i(\boldsymbol{x})$ $(i = 1, \cdots, n)$ has led to various inversion algorithms (Nolet et al., 2005). As the high-order central difference method discussed in this study simulates seismic wave propagation on a 2-D regular mesh, we assume the spatial increments along x and z directions are Δx and Δz, respectively. Let the knot node $\boldsymbol{x}_i$ with a global index i be the grid node (x_m, z_l) on the 2-D mesh. In this scenario, the simplest basis function may be the piecewise constant function

$$L_i(\boldsymbol{x}) = L_i(x,z) = \begin{cases} 1, & \text{if } (x,z) \in \left[x_{m-1/2}, x_{m+1/2}\right] \\ & \times \left[z_{l-1/2}, z_{l+1/2}\right]; \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

And the coefficient of the unknown C_i in Eq. () is

$$\sum_{j=1}^{n} K_j \int\limits_\Omega L_j(\boldsymbol{x}) L_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \Delta x \Delta z K_i. \tag{24}$$

However, the interpolation function with the basis functions (Eq. 23) is not even continuous. To make the interpolation function continuous, we can use bilinear interpolation to fit the perturbation field $\delta c(\boldsymbol{x})/c(\boldsymbol{x})$ and the travel-time kernel $K(\boldsymbol{x})$. Bilinear interpolation performs linear interpolation first in one direction and then in the other direction. The basis function $L_i(\boldsymbol{x})$ for bilinear interpolation takes the following form

$$L_i(\boldsymbol{x}) = L_i(x,z) = \begin{cases} \frac{x-x_{m-1}}{x_m-x_{m-1}} \frac{z-z_{l-1}}{z_l-z_{l-1}}, & \text{if } (x,z) \in \left[x_{m-1}, x_m\right] \times \left[z_{l-1}, z_l\right]; \\ \frac{x-x_{m-1}}{x_m-x_{m-1}} \frac{z_{l+1}-z}{z_{l+1}-z_l}, & \text{if } (x,z) \in \left[x_{m-1}, x_m\right] \times \left[z_l, z_{l+1}\right]; \\ \frac{x_{m+1}-x}{x_{m+1}-x_m} \frac{z-z_{l-1}}{z_l-z_{l-1}}, & \text{if } (x,z) \in \left[x_m, x_{m+1}\right] \times \left[z_{l-1}, z_l\right]; \\ \frac{x_{m+1}-x}{x_{m+1}-x_m} \frac{z_{l+1}-z}{z_{l+1}-z_l}, & \text{if } (x,z) \in \left[x_m, x_{m+1}\right] \times \left[z_l, z_{l+1}\right]; \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

Correspondingly, the coefficient for the unknown C_i in Eq. (22) becomes

$$\sum_{j=1}^{n} K_j \int\limits_\Omega L_j(\boldsymbol{x}) L_i(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = \Delta x \Delta z \begin{pmatrix} \frac{1}{36} & \frac{4}{36} & \frac{1}{36} \\ \frac{4}{36} & \frac{16}{36} & \frac{4}{36} \\ \frac{1}{36} & \frac{4}{36} & \frac{1}{36} \end{pmatrix} \tag{26}$$

$$\circ \begin{pmatrix} K_{m-1,l+1} & K_{m,l+1} & K_{m+1,l+1} \\ K_{m-1,l} & K_{m,l} & K_{m+1,l} \\ K_{m-1,l-1} & K_{m,l-1} & K_{m+1,l-1} \end{pmatrix},$$

where "$\circ$" denotes a two-step operation which first gets the entry-wise product of two matrices and then sums up all the entries of the produced matrix. Kernel values for global and local grids are linked by $K_{i+pM+q} = K_{m+p,l+q}$ (M is the number of grid nodes along x direction, and $p, q = -1, 0, 1$). To have a smoother fitting function, we could further use bicubic interpolation, which is an extension of cubit interpolation on 2-D regular mesh. Actually, in the framework of piecewise constant interpolation (Eq. 23), both bilinear interpolation and bicubic interpolation can be achieved by replacing C_i's coefficient $\Delta x \Delta z K_i$ in Eq. (24) with a weighted average value $\Delta x \Delta z \widetilde{K}_i$ around the knot node $\boldsymbol{x}_i$ and its neighbors such as shown in Eq. (26). Since we have previously smoothed the kernel by convolving it with a Gaussian function, using piecewise constant interpolation or bilinear interpolation to construct tomographic Eq. (22) is accurate enough for practical applications.

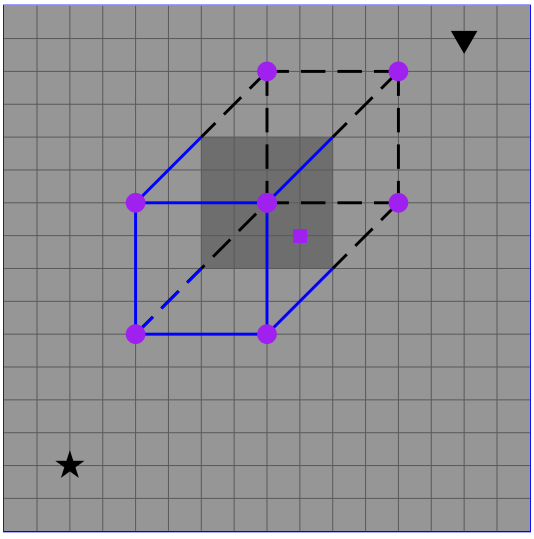

Figure 3. Linear interpolation of the material properties on one forward modeling grid node (purple square) with material properties on its eight surrounding inversion grid nodes (purple circles). Forward modeling grid is a regular 2-D mesh with fixed grid intervals (formed by grey lines), and inversion grid is a 3-D regular mesh with variable grid intervals. Black star and black inverse triangle denote the locations of the earthquake and seismic station in the 2-D vertical plane, respectively.

4.2 Inversion grid

For the high-order central difference scheme, we assume that seismic waves propagate in 2-D vertical planes, and hence sensitivity kernels are restricted to the same 2-D planes. For a single pair of source $\boldsymbol{x}_\mathrm{s}$ and receiver $\boldsymbol{x}_\mathrm{r}$, the forward grid nodes and equally the velocity model parameters C_i are distributed on a 2-D regular mesh in Eq. (22). An additional set of grid nodes needs to be introduced to characterize the actual 3-D tomographic region. For simplicity, we use a regular grid with variable grid intervals to represent the final tomographic results, which has the advantage of allowing a fine grid for a target volume with dense data coverage (mostly depending on spatial distribution of source and receivers) to be embedded in coarse grid nodes.

To be consistent with the realistic application in the second paper, we directly set up the inversion grid in a geographical coordinate system (d, ϕ, λ), where d, ϕ, and λ are depth, latitude, and longitude, respectively. If the Cartesian coordinate system is adopted for the inversion grid, the following derivation procedure is almost the same. In a 3-D regular inversion grid, each forward modeling grid node $\boldsymbol{x}_i$ $(i = 1, 2, \cdots, n)$ is located within a cube formed by eight inversion grid nodes (Fig. 3). It is natural and straightforward to use trilinear interpolation between the eight grid nodes (Zhao et al., 1992). Note that the Cartesian coordinate $\boldsymbol{x}_i$ should be transformed into geographical coordinate $\widetilde{\boldsymbol{x}}_i$ prior to locating it in a cube. If assuming that $\widetilde{\boldsymbol{x}}_i$ is located within the cube formed by $(d_{r+j_1}, \phi_{p+j_2}, \lambda_{q+j_3})$ $(j_1, j_2, j_3 = 0, 1;\ 1 \le r + j_1 \le R;\ 1 \le p + j_2 \le P;\ 1 \le q + j_3 \le Q;\ R, P, Q$ are the numbers of inversion grid nodes along depth, latitude, and longitude, respectively), the unknown velocity model parameter C_i corresponding to $\widetilde{\boldsymbol{x}}_i$ can be expressed as a linear combination of the parameters $X_{r+j_1,p+j_2,q+j_3}$ $(j_1, j_2, j_3 = 0, 1)$ at the eight inversion grid nodes:

$$C_i = \sum_{j_1,j_2,j_3=0}^{1} \left(1 - \left|\frac{d - d_{r+j_1}}{d_{r+1} - d_\mathrm{r}}\right|\right) \left(1 - \left|\frac{\phi - \phi_{p+j_2}}{\phi_{p+1} - \phi_p}\right|\right) \left(1 - \left|\frac{\psi - \psi_{q+j_3}}{\psi_{q+1} - \psi_q}\right|\right) X_{r+j_1,p+j_2,q+j_3}, \tag{27}$$

and it defines a continuously varying velocity perturbation field $\delta c(\boldsymbol{x})/c(\boldsymbol{x})$. Note that the velocity field $c(\boldsymbol{x})$ itself can be discontinuous. Substituting Eq. (27) into Eq. (22) gives the tomographic equation on the inversion grid

$$T^\mathrm{obs} - T^\mathrm{syn} = \sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{q=1}^{Q} a_{r,p,q} X_{r,p,q}, \tag{28}$$

where $a_{r,p,q}$ is the coefficient for the unknown $X_{r,p,q}$ and pre-determined, the accuracy of which relies on not only the accurate calculation of the travel-time kernel $K(\boldsymbol{x})$ but also on the choice of the inversion grid. For the convenience of this discussion, we convert the 3-D array index (r, p, q) of the inversion grid to 1-D index $n = (r-1)PQ + (q-1)P + q$ $(1 \le n \le N = RPQ)$. Tomographic Eq. (28) can be rewritten as

$$T^\mathrm{obs} - T^\mathrm{syn} = \sum_{n=1}^{N} a_n X_n \tag{29}$$

for a single pair of source $\boldsymbol{x}_\mathrm{s}$ and receiver $\boldsymbol{x}_\mathrm{r}$, which relates the travel-time residual $T^\mathrm{obs} - T^\mathrm{syn}$ linearly to the unknown relative velocity perturbation X_n $(1 \le n \le N)$ on the inversion grid.

5 Regularization and inversion method

With a significant increase in both quantity and quality of seismic data from the proliferation of dense seismic arrays, an increasing number of seismic data will be involved in seismic tomography, which may result in higher-resolution tomographic models. Certainly, more data will increase the complexity of the seismic inverse problem.

When M seismic measurements are used to explore the subsurface structure, M tomographic equations take the form of Eq. (29) and form a linear system $\boldsymbol{b} = \mathbf{A}\boldsymbol{X}$ at each iteration, where $\boldsymbol{b} = [b_m]_{M\times 1}$ and $b_m = T_m^\mathrm{obs} - T_m^\mathrm{syn}$ is the iterative travel-time residual vector, $\mathbf{A} = [a_{m,n}]_{M\times N}$ is the Fréchet or Jacobin matrix calculated in the current iterative model and $\boldsymbol{X} = [X_n]_{N\times 1}$ is the unknown model vector. Since

the problem $\boldsymbol{b} = \mathbf{A}\boldsymbol{X}$ is always ill-posed (either because of non-uniqueness or non-existence of $\boldsymbol{X}$), the general way to solve it is to seek a solution that minimizes the following regularized objective function

$$\chi(\boldsymbol{X}) = \frac{1}{2}(\mathbf{A}\boldsymbol{X} - \boldsymbol{b})^T \mathbf{C}_\mathrm{d}^{-1}(\mathbf{A}\boldsymbol{X} - \boldsymbol{b}) + \frac{\epsilon^2}{2}\boldsymbol{X}^T \mathbf{C}_\mathrm{m}^{-1}\boldsymbol{X} + \frac{\eta^2}{2}\boldsymbol{X}^T \mathbf{D}^T \mathbf{D}\boldsymbol{X}, \tag{30}$$

where $\mathbf{C}_\mathrm{d}$ and $\mathbf{C}_\mathrm{m}$ are the a prior data and model covariance matrix which reflect the uncertainties in the data and the initial model (Rawlinson et al., 2010), $\mathbf{D}$ is a derivative smoothing operator for model vector $\boldsymbol{X}$, and ϵ and η are the damping parameter and smoothing parameter, respectively (e.g., Tarantola, 2005; Li et al., 2008; Rawlinson et al., 2010). The last two terms on the right-hand side of Eq. (30) are regularization terms, which are included to improve the conditioning of the inverse problem $\boldsymbol{b} = \mathbf{A}\boldsymbol{X}$ and are designed to give preference to solutions with desirable properties (Aster et al., 2012): damping favors a result that is close to the reference model, while smoothing reduces the differences between adjacent nodes and thus produces smooth model variations (Li et al., 2006). Generally speaking, the objective function (Eq. 30) tries to strike a balance between how well the solution satisfies the data, the variations of the solution from the reference model, and the smoothness of the solution model.

Calculating the gradient (Fréchet derivative) of the objective function $\chi(\boldsymbol{X})$ is often a key step in finding an optimal solution to the minimization problem (Eq. 30) (Rawlinson et al., 2010). Here, the Fréchet derivative of the objective function $\chi(\boldsymbol{X})$ can be expressed as

$$\frac{\partial \chi(\boldsymbol{X})}{\partial \boldsymbol{X}} = \left(\mathbf{A}^T \mathbf{C}_\mathrm{d}^{-1}\mathbf{A} + \epsilon^2 \mathbf{C}_\mathrm{m}^{-1} + \eta^2 \mathbf{D}^T\mathbf{D}\right)\boldsymbol{X} - \mathbf{A}^T \mathbf{C}_\mathrm{d}^{-1}\boldsymbol{b}. \tag{31}$$

Based on the Fréchet derivative $\partial\chi(\boldsymbol{X})/\partial\boldsymbol{X}$, we describe two different approaches to solve the optimization (minimization) problem (Eq. 30).

5.1 LSQR solver

The minimizer $\widetilde{\boldsymbol{X}}$ of Eq. (30) satisfies $\partial\chi(\widetilde{\boldsymbol{X}})/\partial\boldsymbol{X} = 0$ and formally can be expressed as

$$\widetilde{\boldsymbol{X}} = \left(\mathbf{A}^T \mathbf{C}_d^{-1}\mathbf{A} + \epsilon^2 \mathbf{C}_m^{-1} + \eta^2 \mathbf{D}^T\mathbf{D}\right)^{-1} \mathbf{A}^T \mathbf{C}_d^{-1}\boldsymbol{b}. \tag{32}$$

Clearly, to explicitly obtain $\widetilde{\boldsymbol{X}}$ we need to invert an $N \times N$ matrix. There are various methods available to fulfil this goal, such as LU decomposition, singular value decomposition (SVD), and conjugate-gradient types of methods such as LSQR algorithm. Among these methods, LSQR algorithm may be one of the most efficient and widely used methods to solve a linear system, especially when N is very large (Paige and Saunders, 1982). Additionally, the minimization problem (Eq. 30) is equivalent to solving the following linear system in a least square sense

$$\begin{pmatrix} \mathbf{C}_d^{-1/2}\mathbf{A} \\ \epsilon \mathbf{C}_m^{-1/2} \\ \eta \mathbf{D} \end{pmatrix} \boldsymbol{X} = \begin{pmatrix} \mathbf{C}_d^{-1/2}\boldsymbol{b} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \tag{33}$$

and the application of LSQR or SVD to Eq. (33) will give the same solution as that of Eq. (32) (Rawlinson et al., 2010). Once we obtain a perturbation velocity field $\widetilde{\boldsymbol{X}}$, the velocity model can be updated from the current velocity model on inversion grid, C, to $C + \widetilde{\boldsymbol{X}}$. Because of the nonlinearity of the inverse problem, further iteration may be needed to update the velocity model until the objective function $\chi(\boldsymbol{X})$ drops below a tolerance level.

5.2 Nonlinear conjugate gradient method

Once we have the Fréchet derivative of the objective function computed in Eq. (31), instead of inverting the matrix in Eq. (32), we can alternatively use a nonlinear conjugate-gradient method to iteratively improve the model (e.g., Fletcher and Reeves, 1964; Tromp et al., 2005). Previous studies have shown the feasibility and efficiency of this nonlinear conjugate-gradient method in recovering seismic properties of the Earth's interior (e.g., Tape et al., 2007, 2009; Zhu et al., 2012). Here, we summarize the step-by-step process of this nonlinear conjugate-gradient method, which starts from $k = 0$ (Tape et al., 2007; Kim et al., 2011):

1. Calculate the objective function $\chi(\boldsymbol{X}^k)$, compute the gradient $\boldsymbol{g}^k = \partial\chi/\partial\boldsymbol{X}^k$.
2. Compute the model update direction $\boldsymbol{p}^k = -\boldsymbol{g}^k + \beta_k \boldsymbol{p}^{k-1}$. For the first iteration $k = 0$, set $\beta_0 = 0$ and $\boldsymbol{p}^0 = -\boldsymbol{g}^0$; otherwise calculate β_k based on the equation
$$\beta_k = \max\left(0, \frac{\boldsymbol{g}^k \cdot (\boldsymbol{g}^k - \boldsymbol{g}^{k-1})}{\boldsymbol{g}^{k-1} \cdot \boldsymbol{g}^{k-1}}\right). \tag{34}$$
3. Determine the step length λ_k in the model update direction:
   - Let $f_1 = \chi(\boldsymbol{X}^k)$, $g_1 = \boldsymbol{g}^k \cdot \boldsymbol{p}^k$ and compute a test step length $\lambda_t = -2f_1/g_1$.
   - Calculate the test perturbation model $\boldsymbol{X}_t^k = \boldsymbol{X}^k + \lambda_t \boldsymbol{p}^k$.
   - Compute the objective function $\chi(\boldsymbol{X}_t^k)$ and let $f_2 = \chi(\boldsymbol{X}_t^k)$. Note that we generally have $f_1 > f_2 > 0$.
   - Compute
   $$\gamma = [(f_2 - f_1) - g_1\lambda_t]/\lambda_t^2, \quad \xi = g_1 \tag{35}$$
   and then λ_k is given by
   $$\lambda_k = \begin{cases} -\xi/(2\gamma), & \gamma \neq 0; \\ \text{error otherwise.}, \end{cases} \tag{36}$$

4. Update the perturbation model $\boldsymbol{X}^{k+1} = \boldsymbol{X}^k + \lambda_k \boldsymbol{p}^k$.

5. If $||\boldsymbol{g}^k||_{L_2} = (\boldsymbol{g}^k \cdot \boldsymbol{g}^k)^{1/2} \leq \epsilon$, the tolerance level, then $\boldsymbol{X}^{k+1}$ is the optimal perturbation model; otherwise reiterate from the first step (1) with $k+1$.

For the current model $\boldsymbol{m}_k$ which has a perturbation $\boldsymbol{X}^k$ from the starting model $\boldsymbol{m}_0$, we can rewrite the gradient of the objective function as

$$\frac{\partial \chi(\boldsymbol{X}^k)}{\partial \boldsymbol{X}} = -(\mathbf{A}^k)^T \mathbf{C}_d^{-1} \boldsymbol{b}^k + \left(\epsilon^2 \mathbf{C}_m^{-1} + \eta^2 \mathbf{D}^T \mathbf{D}\right) \boldsymbol{X}^k, \quad (37)$$

where $\mathbf{A}^k$ and $\boldsymbol{b}^k$ are respectively the Fréchet matrix and travel-time residuals in the kth model. The first term on the right-hand side of Eq. (37) is actually the sum of all travel-time kernels (negatively) weighted by their corresponding travel-time residuals. That is to say, if no damping and smoothing operations are applied, the gradient (Eq. 37) is simply the sum of all weighted individual travel-time kernels. Since operators $\mathbf{C}_d^{-1}$, $\mathbf{C}_m^{-1}$, and $\mathbf{D}$ remain constant throughout the whole process, to update the model from $\boldsymbol{m}_k$ to $\boldsymbol{m}_{k+1}$ we only need to compute the Fréchet matrix and travel-time residuals in model $\boldsymbol{m}_k$. This is different from the approach using the LSQR algorithm as a linear system is solved at each iteration. Generally speaking, the model update with the LSQR algorithm may be larger than the nonlinear conjugate-gradient method and the LSQR approach probably requires fewer iterations.

Besides the LSQR solver and the nonlinear conjugate-gradient method, the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm (Nocedal , 1980) is another good candidate for solving the optimization problem of Eq. 30 (Luo, 2012). The L-BFGS algorithm is a quasi-Newton method that only involves vector operations and storage and is therefore well suited for optimization problems with a large number of variables. More details on L-BFGS can be found in Nocedal (1980) and Luo (2012).

6 Numerical examples

As discussed in Sect. 5, computing travel-time sensitivity kernel or the Fréchet derivative of the objective function is one of the key components of wave-equation-based travel-time seismic tomography. In this section, we show examples of Fréchet kernel for one earthquake. These examples provide insights into sensitivities of various seismic phases and the future applications of wave-equation-based travel-time seismic tomography involving tens of thousands of seismic record.

A two-layer S wave velocity model with the Moho discontinuity at a depth of 30.0 km is used as a reference model. The size of the model is 100 km × 50 km. S wave velocities in the crust and the mantle are 3.2 and 4.5 km s^{-1}, respectively. The "true" model is the same two-layer S wave velocity model but with a −5.0 % low-velocity anomaly (red box in Fig. 5) and a +5.0 % high-velocity anomaly (blue box in Fig. 5) included in the mid-crust. An earthquake is placed at the horizontal distance $x = 50.0$ km and the depth of 12.0 km with the dominant frequency of the Gaussian source time function at 1.0 Hz. There are 51 stations equally spaced on the surface with an interval of 2.0 km. The high-order central difference method is used as the forward solver. Seismograms recorded at $x = 14.0$ km and $x = 86.0$ km on the surface are shown in Fig. 4a and 4b, respectively. Three main phases can be observed in these seismograms, including the direct S wave, the Moho reflected phase SmS and the surface reflected wave $sSmS$, which provide complementary information on the crustal structures. For example, D. Zhao et al. (2005) used S, SmS, and $sSmS$ arrivals to conduct crustal tomography in the 1992 Landers earthquake area with a ray-based tomographic method. Here, we compute Fréchet kernels for the three seismic phases. Because only sensitivity kernels are computed and no inversion is conducted, the two regularization terms at the right-hand side of Eq. (37) are not taken into account in this section.

For seismograms recorded at $x = 14.0$ km (Fig. 4a) the direct S wave and the Moho reflected SmS phase for the true model arrive closely following the corresponding phases in the reference model. As shown in Fig. 5a and b, the geometrical ray paths of both phases are partially within the low-velocity zone, and therefore it is reasonable to have delayed S and SmS arrivals in the true model. For the $sSmS$ phase, its geometrical ray path does not pass through the low-velocity zone but its first Fresnel zone partially coincides with the low-velocity anomaly (Fig. 4c). Due to the influence of the low-velocity zone, the arrival time of $sSmS$ is delayed by 0.0025 s obtained through the cross-correlation calculation. The Fréchet kernels for S, SmS, and $sSmS$ are shown in Fig. 5a–c, which closely follows their corresponding geometry ray paths (indicated by dashed lines). The positive Fréchet kernel values in the first Fresnel zones indicate that a reduction of velocity within these regions will result in the reduction of objective function χ. Figures 4b and 5d–f are for the case when seismic waves travel through a high-velocity region in the true model and seismograms are recorded at the station $x = 86.0$ km. Negative Fréchet kernel values in the first Fresnel zones suggest that an increase of velocity in this region of the reference model can reduce the objective function χ.

The Fréchet kernels displayed in Fig. 5a–f are associated with a particular seismic phase at one seismic station, i.e., the individual kernels. Of course one seismic record does not well constrain the subsurface heterogeneous structure. With the 51 stations on the surface, we can compute the Fréchet kernel for one seismic phase defined at all seismic stations, shown in Fig. 5g–i. These kernels are actually the sum of individual S, SmS, and $sSmS$ kernels computed at each station. Due to the increased data coverage and the constructive effect, both the low and high-velocity areas are sampled by

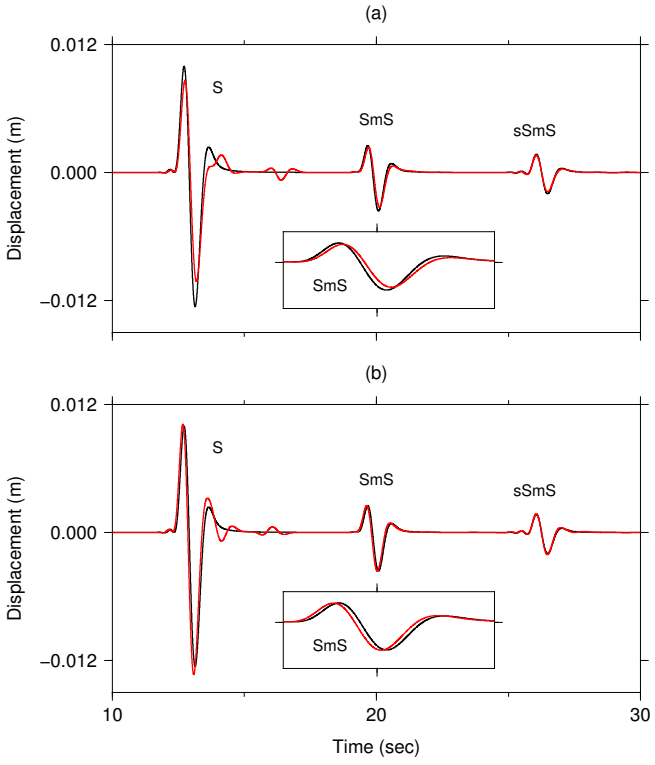

Figure 4. Seismograms recorded by the stations located at **(a)** $x = 14$ km and **(b)** $x = 86$ km on the surface. Seismograms computed in the reference model are shown as black curves, and those computed in the true model are illustrated by red curves. The computational domain is a crust-over-mantle model with a size of 100 km × 50 km. The crust has a thickness of 30 km containing one low and one high-velocity zone in the true model respect to the reference model (Fig. 5). The earthquake is located at $x = 50$ km at the depth of 12 km.

the bulk part of the kernels. The values of these three kernels are positive within the low-velocity zone and negative within the high-velocity area, which indicates that updating the velocity model in the opposite direction $-\partial\chi(\boldsymbol{X})/\partial\boldsymbol{X}$ would reduce the objective function χ. We can further define the objective function χ as the sum of S, SmS, and $sSmS$ phases at all seismic stations. The corresponding Fréchet kernel is shown in Fig. 6, which is the sum of the kernels in Fig. 5g–i. It can be observed that kernel values at the anomalous regions are not prominent in Fig. 5g–i, but are dominant in Fig. 6. This suggests that we may simultaneously use different seismic phase data to highlight anomalous structures in future studies. For demonstration purpose, we only worked with one event in this part. To increase the illumination, more seismic events should be included. Once the Fréchet kernels for all events and phases are computed, the LSQR solver or the nonlinear conjugate-gradient method can be used to iteratively improve the velocity model.

7 Discussion and conclusions

Wave-equation-based travel-time seismic tomography (WETST) involves 2-D forward modeling and 3-D tomographic inversion. Considering adjoint tomography based on 3-D spectral-element method or other 3-D forward modeling techniques as an approach for "3-D–3-D" seismic tomography (e.g., Tromp et al., 2005; Tape et al., 2009; Zhu et al., 2012), WETST can be viewed as a "2-D–3-D" adjoint tomography method. From the computation point of view, 2-D forward modeling with a high-order central difference scheme is computationally efficient and can be conducted on most single PCs. This makes it possible to handle large seismic data sets with WETST. Actually, increasing data amount and data coverage is the best way to improve the resolution of tomographic results, and sometimes may compensate for the approximations in the tomography technique itself. For example, it is well known that one main drawback of ray theory is that it does not consider the influence of off-ray structures (Dahlen et al., 2000); however, a good data set with a dense and even distribution of ray paths can greatly improve the resolution of ray tomography (Tong et al., 2011). A similar problem for the 2-D approximation in WETST is its ignorance of the off-plane influence on seismic arrivals. To what extent this approximation is valid and how it affects the final inversion results should be further investigated. However, by taking advantage of the computational efficiency of 2-D forward modeling, we may be able to reduce the effect of the 2-D approximation by increased data coverage in real applications.

WETST only uses travel-time information for two main reasons. First, travel time is quasi-linear with respect to variations in the velocity structures, which greatly assists the convergence of gradient-based inversion methods as presented in Sect. 5. Second, compared with fitting waveforms, it is much easier to only predict the arrival times of particular phases on synthetic seismograms computed through 2-D forward modeling. The envelop energy method or the combined ray and cross-correlation method presented in this study can be easily implemented to pick the arrival times on synthetic seismograms. Additionally, the finite-frequency travel-time residuals $T^{\mathrm{obs}} - T^{\mathrm{syn}}$ in previous finite-frequency tomography studies (e.g., Hung et al., 2001, 2004, 2011) were determined with the cross-correlation travel-time measurement (Dahlen et al., 2000). However, in this study, the residuals are obtained after manually picking the onset times T^{obs} on observed seismograms (bandpass filtered) and calculating T^{syn} on synthetic seismograms with the energy envelop method or the combined ray and cross-correlation method. Considering that the derivations of the travel-time sensitivity kernel in Eq. (11) and previous finite-frequency sensitivity kernels (e.g., Dahlen et al., 2000; Zhao et al., 2000; Tromp et al., 2005; Fichtner et al., 2006) are mainly based on the Born approximation, which requires that the reference velocity model $c(\mathbf{x})$ for synthetic seismogram $s(t)$ is very

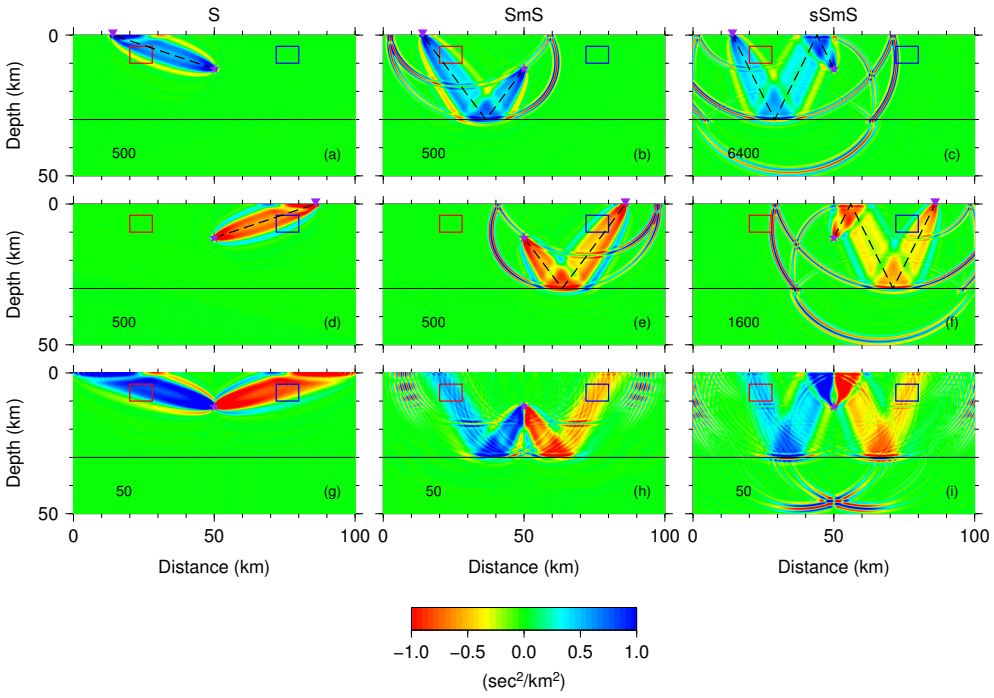

Figure 5. Fréchet kernels corresponding to direct S **(a, d, g)**, SmS **(b, e, h)**, and $sSmS$ **(c, f, i)** phases. Panels **(a)–(c)** are computed only using seismograms recorded at the station $x = 14$ km. These seismograms are influenced by the low-velocity zone in the red box in the true model. Panels **(d)–(f)** are only related to seismograms recorded at the station $x = 86$ km, which are influenced by the high-velocity zone in the blue box in the true model. Panels **(g)–(i)** are computed for all 51 stations on the surface. To use a uniform color bar indicated at the bottom for all panels, each kernel is amplified by multiplying the number at the bottom left of the corresponding panel.

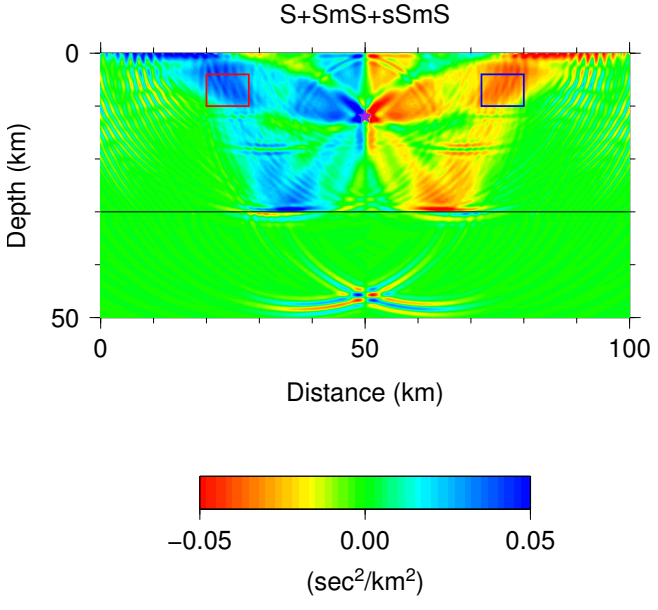

Figure 6. Summation of the Fréchet kernels corresponding to the kernels for S, SmS, and $sSmS$ phases in Fig. 5g–i.

close to the real model $c(\mathbf{x}) + \delta c(\mathbf{x})$ for real data $d(t)$, i.e., $|\delta c(\mathbf{x})| \ll c(\mathbf{x})$. Since $|\delta c(\mathbf{x})| \ll c(\mathbf{x})$, it is straightforward to get that $|s(t) - d(t)| \ll |s(t)|$ if both $s(t)$ and $d(t)$ satisfy the same wave equation. Providing that $|s(t) - d(t)| \ll |s(t)|$, we can say that the data and synthetic broad pulses $d(t)$ and $s(t)$ have very similar waveforms, then the time differences of onset, peak, and end times of the two pulses should be almost the same. The time difference of onsets on synthetic and data pulses should be identical or very close to the one calculated with the cross-correlation travel-time measurement, since cross-correlation travel-time difference is also a kind of travel-time difference. But in this study, we take a detour to get the onset time of synthetic seismogram $s(t)$. It is computationally expensive to simulate the propagation of seismic waves in 3-D models. To reduce the computation cost, we alternatively use 2-D simulation to get synthetic seismograms $u(t)$, which may only match the 3-D synthetic seismogram $s(t)$ against the onset time. Since we only need to know the exact onset time of the synthetic pulse, this 2-D approximation is accurate enough. Therefore, the approach of calculating the travel-time residuals $T^{\mathrm{obs}} - T^{\mathrm{syn}}$ proposed in this study can provide high-accuracy data for the wave-equation-based travel-time seismic tomography study.

If 3-D finite-frequency effects need to be taken into account and full waveform fitting is required, we suggest the use of 3-D–3-D tomographic techniques such as adjoint tomography based on the spectral-element method (Tromp et al., 2005; Fichtner et al., 2006). In this case, WETST may

be used to construct the starting models for 3-D–3-D seismic tomography. The hybrid approach could help reduce the total computational costs and speed up the convergence rate of the inverse algorithm as a "closer" initial model is used. How much efficiency can be gained depends on the problem itself, and a comparison of computation time costs between the 2-D–3-D and 3-D–3-D methods is presented in the companion paper (Tong et al., 2014b). Considering that ray-based seismic tomography methods are still the most prevalent tomographic methods and WETST has the advantage of more accurately computed sensitivity kernels, WETST may be a potentially useful compromise for 3-D tomographic inversions before the wider application of 3-D–3-D seismic tomography in the near future.

Forward modeling in WETST discussed in this paper is based on solving a 2-D acoustic wave equation in the Cartesian coordinates. If the source and the receiver are far apart and the curvature of the Earth cannot be neglected, the acoustic wave equation in Cartesian coordinates needs to be transformed into geographical coordinates, which may be necessary for the use of teleseismic data. Currently, WETST cannot use converted seismic phases such as P–S or simultaneously determine the P wave and S wave velocity structures in tomographic inversions. But these two goals can be achieved by replacing the 2-D acoustic wave equation with the 2-D elastic wave equation. Additionally, a regular grid with variable grid intervals is suggested to represent the final tomographic results in this paper. To automatically adapt the inversion grid to the data distribution, adaptive mesh using Delaunay triangles and Voronoi polyhedra can be alternatively adopted (e.g., Sambridge and Rawlinson, 2005; Zhang and Thurber, 2005; Rawlinson et al., 2010). Source inversion and discontinuity (such as the depth of Moho) determination may also be considered in the future (e.g., Liu and Tromp, 2008; Tong et al., 2014a).

In addition, WETST can include not only direct first arrivals (P wave and S wave) but also later reflected (e.g., PmP, SmS, $pPmP$, $sSmS$) and refracted (Pn, Sn) phases as the ray-based tomographic methods do (e.g., D. Zhao et al., 1992, 2005; Xia et al., 2007). Different seismic phases have different traveling paths and are influenced by structural anomalies differently. The combined use of various seismic phases can increase the illumination of the subsurface structures (Figs. 5 and 6). Given that WETST conducts forward modeling in 2-D vertical planes with an efficient high-order central difference scheme, it is possible to include a large set of seismic data in tomographic inversion. Two different inversion algorithms, LSQR solver and the nonlinear conjugate-gradient method, can be used to find the optimal tomographic results efficiently. Since individual kernels are computed, it is also straightforward and efficient to examine resolution in both data and model spaces (Luo, 2012). In a companion paper, we will use WETST to explore the heterogeneous structures beneath the 1992 Landers earthquake (M_w 7.3) area.

Appendix A: High-order central difference method

Yang et al. (2012) developed a finite-difference scheme, nearly analytic central difference (NACD) method to solve the 2-D acoustic wave equation. The NACD method has fourth-order accuracies in both space and time, and it uses only three grid nodes in each spatial direction. This method shows good performance in suppressing numerical dispersions. The essence of the NACD method is to use displacement and its spatial gradient to approximate second-and-higher order spatial derivatives of the displacement. To achieve this goal, the displacement gradient field is obtained by numerically solving some derived acoustic wave equations (Yang et al., 2012). For simplicity, we use a simplified version of the NACD method to simulate 2-D acoustic wave propagation. In this approach, the value of the spatial gradient along one axis at a particular node is interpolated by the displacement values at its neighboring grid nodes. We call the resultant numerical scheme the high-order central difference method. The detailed schemes of the high-order central difference method are summarized as follows:

$$\begin{aligned}
&\frac{u_{i,j}^{n+1}-2u_{i,j}^{n}+u_{i,j}^{n-1}}{\Delta t^2}-c_{i,j}^2 \qquad \text{(A1)}\\
&\left(\frac{u_{i+1,j}^{n}-2u_{i,j}^{n}+u_{i-1,j}^{n}}{\Delta x^2}+\frac{u_{i,j+1}^{n}-2u_{i,j}^{n}+u_{i,j-1}^{n}}{\Delta z^2}\right)\\
&+\left(\frac{c^2\Delta x^2}{12}-\frac{c^4\Delta t^2}{12}\right)\frac{\partial^4 u}{\partial x^4}\Big|_{i,j}^{n}\\
&+\left(\frac{c^2\Delta z^2}{12}-\frac{c^4\Delta t^2}{12}\right)\frac{\partial^4 u}{\partial z^4}\Big|_{i,j}^{n}-\frac{c^4\Delta t^2}{6}\frac{\partial^4 u}{\partial x^2\partial z^2}\Big|_{i,j}^{n}\\
&=\frac{\partial^2 u}{\partial t^2}\Big|_{i,j}^{n}-c^2\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial z^2}\right)\Big|_{i,j}^{n}+O(\Delta t^4+\Delta x^4+\Delta z^4),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4 u}{\partial x^4}\Big|_{i,j}^{n} &= \frac{1}{\Delta x^4}\left(u_{i+1,j}^{n}-2u_{i,j}^{n}+u_{i-1,j}^{n}\right) \qquad \text{(A2)}\\
&+\frac{6}{\Delta x^3}\left(\frac{\partial u}{\partial x}\Big|_{i+1,j}^{n}-\frac{\partial u}{\partial x}\Big|_{i-1,j}^{n}\right)+O(\Delta x^2),\\
\frac{\partial^4 u}{\partial z^4}\Big|_{i,j}^{n} &= \frac{1}{\Delta z^4}\left(u_{i,j+1}^{n}-2u_{i,j}^{n}+u_{i,j-1}^{n}\right) \qquad \text{(A3)}\\
&+\frac{6}{\Delta z^3}\left(\frac{\partial u}{\partial z}\Big|_{i,j+1}^{n}-\frac{\partial u}{\partial z}\Big|_{i,j-1}^{n}\right)+O(\Delta z^2),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^4 u}{\partial x^2\partial z^2}\Big|_{i,j}^{n} &= \frac{1}{\Delta x^2\Delta z^2} \qquad \text{(A4)}\\
&\Big[2\left(u_{i+1,j}^{n}+u_{i-1,j}^{n}+u_{i,j+1}^{n}+u_{i,j-1}^{n}-2u_{i,j}^{n}\right)\\
&-u_{i+1,j+1}^{n}\\
&-u_{i-1,j-1}^{n}-u_{i+1,j-1}^{n}-u_{i-1,j+1}^{n}\Big]\\
&+\frac{1}{2\Delta x\Delta z^2}\Big(\frac{\partial u}{\partial x}\Big|_{i+1,j+1}^{n}-\frac{\partial u}{\partial x}\Big|_{i-1,j-1}^{n}\\
&+\frac{\partial u}{\partial x}\Big|_{i+1,j-1}^{n}-\frac{\partial u}{\partial x}\Big|_{i-1,j+1}^{n}-2\frac{\partial u}{\partial x}\Big|_{i+1,j}^{n}\\
&+2\frac{\partial u}{\partial x}\Big|_{i-1,j}^{n}\Big)\\
&+\frac{1}{2\Delta x^2\Delta z}\Big(\frac{\partial u}{\partial z}\Big|_{i+1,j+1}^{n}\\
&-\frac{\partial u}{\partial z}\Big|_{i-1,j-1}^{n}+\frac{\partial u}{\partial z}\Big|_{i-1,j+1}^{n}-\frac{\partial u}{\partial z}\Big|_{i+1,j-1}^{n}\\
&-2\frac{\partial u}{\partial z}\Big|_{i,j+1}^{n}+2\frac{\partial u}{\partial z}\Big|_{i,j-1}^{n}\Big)+O(\Delta x^2+\Delta z^2),
\end{aligned}$$

$$\begin{aligned}
\frac{\partial u}{\partial x}\Big|_{i,j}^{n} &= \frac{1}{12\Delta x}\left(u_{i-2,j}^{n}-8u_{i-1,j}^{n}+8u_{i+1,j}^{n}-u_{i+2,j}^{n}\right) \qquad \text{(A5)}\\
&+O(\Delta x^4),\\
\frac{\partial u}{\partial z}\Big|_{i,j}^{n} &= \frac{1}{12\Delta z}\left(u_{i,j-2}^{n}-8u_{i,j-1}^{n}+8u_{i,j+1}^{n}-u_{i,j+2}^{n}\right) \qquad \text{(A6)}\\
&+O(\Delta z^4).
\end{aligned}$$

The high-order central difference method also has fourth-order temporal accuracy and fourth-order spatial accuracy. Additionally, the perfectly matched layer boundary condition is used to absorb the outgoing waves (Komatitsch and Tromp, 2003). To implement this numerical method, the gradients $\partial u/\partial x$ and $\partial u/\partial z$ should be explicitly computed based on Eqs. (A5) and (A6). Since the gradients of the displacement are computed in forward modeling, the computation of the travel-time sensitivity kernel (Eq. 11) becomes very straightforward, which shows that the high-order central difference method can be naturally adapted for kernel computations.

Acknowledgements. This work was supported by the National Natural Science Foundation of China (grant no. 41230210), Japan Society for the Promotion of Science (Kiban-S 11050123), and the Discovery Grants of the Natural Sciences and Engineering Research Council of Canada (NSERC, nos. 487237 and 490919). X. Yang was partially supported by the Regents Junior Faculty Fellowship of University of California, Santa Barbara. All figures are made with the Generic Mapping Tool (GMT) (Wessel and Smith, 1991). We thank K. Liu, Y. Luo, and another anonymous reviewer for providing constructive comments and suggestions that greatly improved the manuscript.

Edited by: K. Liu

References

Aki, K. and Lee, W.: Determination of the three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes 1. A homogeneous intial model, J. Geophys. Res., 81, 4381–4399, 1976.

Aki, K. and Richards, P. G.: Quantitative Seismology: Theory and Methods, 2nd edn., University Science Books, 2002.

Akram, J.: Automatic P-wave arrival time picking method for seismic and micro-seismic data, CSPG CSEG CWLS Convention, 2011.

Aster, R. C., Borchers, B., and Thurber, C. H.: Parameter Estimation and Inverse Problems, 2nd edn., Academic Press, 2012.

Baer, M. and Kradolfer, U.: An automatic phase picker for local and teleseismic events, B. Seismol. Soc. Am., 77, 1437–1445, 1987.

Chen, P., Jordan, T. H., and Zhao, L.: Full 3-D waveform tomography: a comparison between the scattering-integral and adjoint-wavefield methods, Geophys. J. Int., 170, 175–181, 2007a.

Chen, P., Zhao, L., and Jordan, T. H.: Full 3-D tomography for crustal structure of the Los Angeles Region, B. Seismol. Soc. Am., 97, 1094–1120, 2007b.

Coppens, F.: First arrival picking on common-offset trace collections for automatic estimation of static corrections, Geophys. Prospect., 33, 1212–1231, 1985.

Dahlen, F., Nolet, G., and Hung, S.: Fréchet kernels for finite-frequency traveltimes – I. Theory, Geophys. J. Int., 141, 157–174, 2000.

Dahlen, F. A. and Nolet, G.: Comment on "On sensitivity kernels for 'wave-equation' transmission tomography" by de Hoop and van der Hilst, Geophys. J. Int., 163, 949–951, 2005.

de Hoop, M. V. and van der Hilst, R. D.: On sensitivity kernels for "wave-equation" transmission tomography, Geophys. J. Int., 160, 621–633, 2005a.

de Hoop, M. V. and van der Hilst, R. D.: Reply to comment by F. A. Dahlen and G. Nolet on "On sensitivity kernels for 'wave-equation' transmission tomography", Geophys. J. Int., 163, 952–955, 2005b.

Dziewonski, A.: Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, J. Geophys. Res., 89, 5929–5952, 1984.

Dziewonski, A. M., Hager, B. H., and O'Connell, R. J.: Large-scale heterogeneities in the lower mantle, J. Geophys. Res., 82, 239–255, 1977.

Earle, P. S. and Shearer, P. M.: Characterization of global seismograms using an automatic-picking algorithms, B. Seismol. Soc. Am., 84, 366–376, 1994.

Fichtner, A. and Trampert, J.: Resolution analysis in full waveform inversion, Geophys. J. Int., 187, 1604–1624, 2011.

Fichtner, A., Bunge, H. P., and Igel, H.: The adjoint method in seismology I. Theory, Phys. Earth Planet. In., 157, 86–104, 2006.

Fichtner, A., Igel, H., Bunge, H.-P., and Kennett, B. L. N.: Simulation and inversion of seismic wave propagation on continental scales based on a spectral-element method, J. Numer. Anal. Indust. Appl. Math., 4, 11–22, 2009.

Fletcher, R. and Reeves, C.: Function minimization by conjugate gradients, Comput. J., 7, 149–154, 1964.

Gautier, S., Nolet, G., and Virieux, J.: Finite-frequency tomography in a crustal environment: application to the western part of the Gulf of Corinth, Geophys. Prospect., 56, 493–503, 2008.

Graves, R. W.: Simulating seismic wave propagation in 3-D elastic media using staggered-grid finite differences, B. Seismol. Soc. Am., 86, 1091–1106, 1996.

Han, L., Wong, J., and Bancroft, J.: Time picking on noisy microseismograms, GeoCanada, 2010.

Hung, S. H., Dahlen, F. A., and Nolet, G.: Wavefront healing: a banana–doughnut perspective, Geophys. J. Int., 146, 289–312, 2001.

Hung, S. H, Shen, Y., and Chiao, L.: Imaging seismic velocity structure beneath the Iceland hot spot: a finite frequency approach, J. Geophys. Res., 109, B08305, doi:10.1029/2003JB002889, 2004.

Hung, S. H., Chen, W. P., and Chiao, L. Y.: A data-adaptive, multiscale approach of finite-frequency, traveltime tomography with special reference to P and S wave data from central Tibet, J. Geophys. Res., 116, B06307, doi:10.1029/2010JB008190, 2011.

Kennett, B. L. N. and Engdah, E. R.: Traveltimes for global earthquake location and phase identification, Geophys. J. Int., 105, 429–465, 1991.

Kim, Y., Liu, Q., and Tromp, J.: Adjoint centroid-moment tensor inversions, Geophys. J. Int., 186, 264–278, 2011.

Komatitsch, D. and Tromp, J.: Introduction to the spectral element method for three-dimensional seismic wave propagation, Geophys. J. Int., 139, 806–822, 1999.

Komatitsch, D. and Tromp, J.: Spectral-element simulations of global seismic wave propagation – I. Validation, Geophys. J. Int., 149, 390–412, 2002a.

Komatitsch, D. and Tromp, J.: Spectral-element simulations of global seismic wave propagation – II. 3-D models, oceans, rotation, and self-gravitation, Geophys. J. Int., 150, 303–318, 2002b.

Komatitsch, D. and Tromp, J.: A perfectly matched layer absorbing boundary condition for the second-order seismic wave equation, Geophys. J. Int., 154, 146–153, 2003.

Komatitsch, D., Liu, Q., Tromp, J., Suss, M. P., Stidham, C., and Shaw, J. H.: Simulations of ground motion in the Los Angeles Basin based upon the spectral-element method, B. Seismol. Soc. Am., 94, 187–206, 2004.

Komatitsch, D., Erlebacher, G., Göddeke, D., and Michéa, D.: High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster, J. Comput. Phys., 229, 7692–7714, 2010.

Lee, H. Y., Koo, J. M., Min, D. J., Kwon, B. D., and Yoo, H. S.: Frequency-domain elastic full waveform inversion for VTI media, Geophys. J. Int., 183, 884–904, 2010.

Li, C., van der Hilst, R. D., and Toksoz, N. M.: Constraining spatial variations in P-wave velocity in the upper mantle beneath Southeast Asia, Phys. Earth Planet. In., 154, 180–195, 2006.

Li, C., van der Hilst, R. D., Engdahl, E. R., and Burdick, S.: A new global model for P wave speed variations in Earth's mantle, Geochem. Geophys. Geosyst., 9, Q05018, doi:10.1029/2007GC001806, 2008.

Liu, Q. and Gu, Y. J.: Seismic imaging: from classical to adjoint tomography, Tectonophysics, 566/567, 31–66, 2012.

Liu, Q. and Tromp, J.: Finite-frequency kernels based on adjoint methods, B. Seismol. Soc. Am., 96, 2283–2297, 2006.

Liu, Q. and Tromp, J.: Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, Geophys. J. Int., 174, 265–286, 2008.

Liu, Q., Polet, J., Komatitsch, D., and Tromp, J.: Spectral-element moment tensor inversions for earthquakes in southern California, B. Seismol. Soc. Am., 94, 1748–1761, 2004.

Luo, Y. and Schuster, G.: Wave equation inversion of skeletonized geophysical data, Geophys. J. Int., 105, 289–294, 1991.

Luo, Y.: Seismic imaging and inversion based on spectral-element and adjoint methods, PhD Thesis, Princeton University, 2012.

Maggi, A., Tape, C. H., Chen, M., Chao, D., and Tromp, J.: An automated time-window selection algorithm for seismic tomography, Geophys. J. Int., 178, 257–281, 2009.

Marquering, H., Dahlen, F., and Nolet, G.: Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, Geophys. J. Int., 137, 805–815, 1999.

Michéa, D. and Komatitsch, D.: Accelerating a three-dimensional finite-difference wave propagation code using GPU graphics cards, Geophys. J. Int., 182, 389–402, 2010.

Montelli, R., Nolet, G., Dahlen, A., Masters, G., Robert Engdahl, E., and Hung, S.: Finite-frequency tomography reveals a variety of plumes in the mantle, Science, 303, 338–343, 2004.

Munro, K. A.: Automatic event detection and picking P-wave arrivals, CREWES Research Report, 18, 12.1–12.10, 2004.

Nocedal, J.: Updating quasi-Newton matrices with limited storage, Math. Comp., 35, 773–782, 1980.

Nolet, G., Dahlen, F. A., and Montelli, R.: Traveltimes and amplitudes of seismic waves: a re-assessment, in: Seismic Earth: Array Analysis of Broadband Seismograms, edited by: Levander, A. and Nolet, G., 37–47, AGU, 2005.

Operto, S., Virieux, J., Dessa, J. X., and Pascal, G.: Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: application to the eastern Nankai trough, J. Geophys. Res., 111, B09306, doi:10.1029/2005JB003835, 2006.

Operto, S., Virieux, J., Amestoy, P., L'Excellent, J. Y., Giraud, L., and Ben Hadj Al, H.: 3-D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: a feasibility study, Geophysics, 72, SM195–SM211, doi:10.1190/1.2759835 2007.

Paige, C. and Saunders, M.: LSQR: An algorithm for sparse linear-equations and sparse least-squares, Trans. Maths Software, 8, 43–71, 1982.

Pratt, R. G. and Shipp, R. M.: Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data, Geophysics, 64, 902–914, 1999.

Rawlinson, N., Pozgay, S., and Fishwick, S.: Seismic tomography: a window into deep earth, Phys. Earth Planet. In., 178, 101–135, 2010.

Rickers, F., Fichtner, A., and Trampert, J.: The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: evidence from full-waveform inversion, Earth Planet. Sc. Lett., 367, 39–51, 2013.

Romanowicz, B.: Seismic tomography of the Earth's mantle, Annu. Rev. Earth Pl. Sci., 19, 77–99, 1991.

Saari, J.: Automated phase picker and source location algorithm for local distances using a single three-component seismic station, Tectonophysics, 189, 307–315, 1991.

Sambridge, M. and Rawlinson, N.: Seismic tomography with irregular meshes, in: Seismic earth: array analysis of broadband seismograms, edited by: Lavender, A. and Nolet, G., Geophysical Monograph, 49–65, AGU, Washington, DC, 2005.

Tape, C., Liu, Q., and Tromp, J.: Finite-frequency tomography using adjoint methods-methodology and examples using membrane surface waves, Geophys. J. Int., 168, 1105–1129, 2007.

Tape, C., Liu, Q., Maggi, A., and Tromp, J.: Adjoint tomography of the southern California crust, Science, 325, 988–992, 2009.

Tape, C., Liu, Q., Maggi, A., and Tromp, J.: Seismic tomography of the southern California crust based on spectral-element and adjoint methods, Geophys. J. Int., 180, 433–462, 2010.

Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation, Geophysics, 49, 1259–1266, 1984.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, 1st edn., Society for Industrial and Applied Mathematics, Philadelphia, 2005.

Thurber, C. H.: Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, J. Geophys. Res., 88, 8226–8236, 1983.

Tian, Y., Montelli, R., Nolet, G., and Dahlen, F.: Computing traveltime and amplitude sensitivity kernels in finite-frequency tomography, J. Comput. Phys., 226, 2271–2288, 2007.

To, A. and Romanowicz, B.: Finite frequency effects on global S diffracted traveltimes, Geophys. J. Int., 179, 1645–1657, 2009.

Tong, P., Zhao, D., and Yang, D.: Tomography of the 1995 Kobe earthquake area: comparison of finite-frequency and ray approaches, Geophys. J. Int., 187, 278–302, 2011.

Tong, P., Zhao, D., and Yang, D.: Tomography of the 2011 Iwaki earthquake (M 7.0) and Fukushima nuclear power plant area, Solid Earth, 3, 43–51, doi:10.5194/se-3-43-2012, 2012.

Tong, P., Chen, C.-W., Komatitsch, D., Basini, P., and Liu, Q.: High-resolution seismic array imaging based on an SEM-FK hybrid method, Geophys. J. Int., 197, 369–395, 2014a.

Tong, P., Zhao, D., Yang, D., Yang, X., Chen, J., and Liu, Q.: Wave-equation based traveltime seismic tomography – Part 2: Application to the 1992 Landers earthquake (M_W 7.3) area, Solid Earth, 5, 1–20, doi:10.5194/se-5-1-2014, 2014b.

Tong, P., Komatitsch, D., Tseng, T.-L., Hung, S.-H., Chen, C.-W., Basini, P., and Liu, Q.: A 3-D spectral-element and frequency-wavenumber (SEM-FK) hybrid method for high-resolution seismic array imaging, Geophys. Res. Lett., 41, 7025–7034, 2014c.

Tromp, J., Tape, C., and Liu, Q.: Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, Geophys. J. Int., 160, 195–216, 2005.

Virieux, J.: SH-wave propagation in heterogeneous media: velocity-stress finite difference method, Geophysics, 49, 1933–1942, 1984.

Virieux, J. and Operto, S.: An overview of full-waveform inversion in exploration geophysics, Geophysics, 74, WCC1–WCC26, 2009.

Wessel, P. and Smith, W. H. F.: Free software helps map and display data, EOS T. Am. Geophys. Un., 72, 441–448, 1991.

Wong, J., Han, L., Bancroft, J., and Stewart, R.: Automatic time-picking of first arrivals on noisy microseismic data, CSEG, 2009.

Xia, S., Zhao, D., Qiu, X., Nakajima, J., Matsuzawa, T., and Hasegawa, A.: Mapping the crustal structure under active volcanoes in central Tohoku, Japan using P and PmP data, Geophys. Res. Lett., 34, L10309, doi:10.1029/2007GL030026, 2007.

Yang, D., Tong, P., and Deng, X.: A central difference method with low numerical dispersion for solving the scalar wave equation, Geophys. Prospect., 60, 885–905, 2012.

Zhang, H. and Thurber, C.: Adaptive-mesh seismic tomography based on tetrahedral and Voronoi diagrams: application to Parkfield, California, J. Geophys. Res., 110, B04303, doi:10.1029/2004JB003186, 2005.

Zhang, H., Thurber, C., and Rowe, C.: Automatic P Wave arrival detection and picking with multiscale wavelet analysis for single-component recordings, B. Seismol. Soc. Am., 93, 1904–1912, 2003.

Zhao, D.: Multiscale seismic tomography and mantle dynamics, Gondwana Res., 15, 297–323, 2009.

Zhao, D.: Tomography and dynamics of Western-Pacific subduction zones, Monogr. Environ. Earth Planets, 1, 1–70, 2012.

Zhao, D., Hasegawa, A., and Horiuchi, S.: Tomographic imaging of P and S wave velocity structure beneath northeastern Japan, J. Geophys. Res., 97, 19909–19928, 1992.

Zhao, D., Todo, S., and Lei, J.: Local earthquake reflection tomography of the Landers aftershock area, Earth. Planet. Sc. Lett., 235, 623–631, 2005.

Zhao, D., Yanada, T., Hasegawa, A., Umino, N., and Wei, W.: Imaging the subducting slabs and mantle upwelling under the Japan Islands, Geophys. J. Int., 190, 816–828, 2012.

Zhao, L. and Jordan, T. H.: Structural sensitivities of finite-frequency seismic waves: a full-wave approach, Geophys. J. Int., 165, 981–990, 2006.

Zhao, L., Jordan, T. H., and Chapman, C. H.: Three-dimensional Fréchet differential kernels for seismic delay times, Geophys. J. Int., 141, 558–576, 2000.

Zhao, L., Jordan, T. H., Olsen, K. B., and Chen, P.: Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, B. Seismol. Soc. Am., 95, 2066–2080, 2005.

Zhu, H., Bozdag, E., Peter, D., and Tromp, J.: Structure of the European upper mantle revealed by adjoint tomography, Nat. Geosci., 5, 493–498, 2012.