Pure Appl. Geophys. 168 (2011), 1527–1552

DOI 10.1007/s00024-010-0240-8

Pure and Applied Geophysics

# Full-Wave Seismic Data Assimilation: Theoretical Background and Recent Advances

Po Chen[1]

*Abstract*—The seismological inverse problem has much in common with the data assimilation problem found in meteorology and oceanography. Using the data assimilation methodology, I will formulate the seismological inverse problem for estimating seismic source and Earth structure parameters in the form of weak-constraint generalized inverse, in which the seismic wave equation and the associated initial and boundary conditions are allowed to contain errors. The resulting Euler–Lagrange equations are closely related to the adjoint method and the scattering-integral method, which have been successfully applied in full-3D, full-wave seismic tomography and earthquake source parameter inversions. I will review some recent applications of the full-wave methodology in seismic tomography and seismic source parameter inversions and discuss some challenging issues related to the computational implementation and the effective exploitation of seismic waveform data.

## 1. Introduction

The study of the solid Earth system using seismological methods has much in common with the study of the Earth's atmosphere, ocean and other natural systems. The proper use of simulation-based predictions to make valid scientific inferences requires a continually iterated cycle of model formulation and verification, simulation-based predictions, validation against observations, and where the model is deficient, data assimilation to improve the model and reinitiate the cycle at a higher level. As we iterate the inference cycle, our model becomes more complex and the data encompassed by the model contain more information.

The melding of observations with the underlying dynamic principles to produce realistic estimations of the state variables and/or model parameters is known as data assimilation, which is providing rapid advances in important areas such as oceanography (e.g. Bennett 1992; Malanotte-Rizzoli 1996; Wunsch 1996) and meteorology (e.g. Bengtsson *et al*. 1981; Daley 1993; Kalnay 2003). Recent advances in parallel computing technology and numerical methods (Olsen 1994; Bao *et al*. 1998; Graves 1996; Akcelik *et al*. 2003; Olsen *et al*. 2003; Komatitsch *et al*. 2004; Olsen *et al*. 2006) have made large-scale, three-dimensional numerical simulations of seismic wavefields much more affordable, and they have opened up the possibility of full-3D, full-wave seismic data assimilation, in which we combine seismological observations of ground motions with the underlying dynamic principles described by the 3D seismic wave equation to produce realistic estimations of the properties of the seismic sources that generate seismic waves and the geological structures through which seismic waves propagate.

Schemes for solving data assimilation problems can be related either to estimation theory (e.g. Simon 2006), which can be formulated within a Bayesian statistical framework (Wikle and Berliner 2006), or to control theory that solves optimization problems constrained by partial differential equations (e.g. Schulz 2006). In this article, I will formulate the seismic data assimilation problem in the generalized inverse framework (Evensen 2009). The physics of the underlying dynamic principles is accounted for using the 3D seismic wave equation. The starting point of my derivation is the Bayes' theorem, which defines the posterior probability density functions of the poorly known seismic source and structure parameters conditioned on a set of seismological observations. The adjoint method, which was adopted to solve seismic imaging problems in Bamberger *et al*. (1977, 1982) and Tarantola (1984, 1988) and later extended to solve both seismic source and structural inverse problems in Tromp *et al*. (2005),

[1] Department of Geology and Geophysics, University of Wyoming, Laramie, USA. E-mail: pseudopochen@gmail.com

can be considered as a simplification of the generalized inverse, in which the dynamic model is assumed to have zero model uncertainties. A slight alteration in the derivation of the generalized inverse naturally leads to the scattering-integral method used for full-3D waveform tomography in CHEN *et al*. (2007a, b) and for earthquake moment tensor inversion in ZHAO *et al*. (2006). Through my derivation, it is possible to interpret different seismic data assimilation methods in a unified context and to understand the assumptions and approximations they rely on.

The formulation derived in this article rests upon some important assumptions. The stochastic noise processes used to account for model uncertainties in the wave equation and its initial and boundary conditions are treated as additive and the statistical properties of the model residuals and model parameters are assumed to be Gaussian. For non-Gaussian statistics, the formulation derived in this article still has applicability if a Gaussian distribution provides an adequate approximation. New techniques that can represent and propagate uncertainties with non-Gaussian statistics also exist. A promising new development is the generalized Polynomial Chaos (gPC) theory (WIENER 1938; XIU and KARNIADAKIS 2003; SANDU *et al*. 2007), in which any probability densities are represented using a complete, polynomial-based stochastic space. Formulations applicable to general probability densities will be presented in future publications.

In this article, I will also analyze the computational issues involved in full-wave seismic data assimilation in terms of compute cycles and data storage. Capabilities of the full-wave methodology will be demonstrated using examples and several challenging issues will be discussed in the end.

## 2. Generalized Inverse

We consider elastic waves propagating inside a 3D domain $V$ with boundary $\partial V$. Points inside $V$ or on $\partial V$ are denoted as $\mathbf{x}$ and time is denoted as $t$. The medium is described by the mass density $\rho(\mathbf{x})$ and a fourth-order rate-of-relaxation tensor $\boldsymbol{\psi}(\mathbf{x}, t)$ with components $\psi^{ijkl}(\mathbf{x}, t)$. The forcing of the system is described by a body-force density $\mathbf{f}_s(\mathbf{x}, t)$ with components $f_s^i(\mathbf{x}, t)$ and a second-order moment tensor density $\mathbf{M}_s(\mathbf{x}, t)$ with components $M_s^{ij}(\mathbf{x}, t)$. Here we have introduced an index $s$ to reference different seismic sources. The wave-field is described by the source-specific displacement field $\mathbf{u}_s(\mathbf{x}, t)$ with components $u_s^i(\mathbf{x}, t)$ and the second-order stress tensor field $\boldsymbol{\sigma}_s(\mathbf{x}, t)$ with components $\sigma_s^{ij}(\mathbf{x}, t)$. We assume that the dynamic model described by the seismic wave equation has an additive model residual,

$$\rho(\mathbf{x})\partial_t^2 u_s^i(\mathbf{x},t) - \sum_j \partial_j \left[ \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} \psi^{ijkl}(\mathbf{x}, t-\tau) \times \partial_l u_s^k(\mathbf{x},\tau) + M_s^{ij}(\mathbf{x},t) \right] - f_s^i(\mathbf{x},t) = q_s^i(\mathbf{x},t), \quad (1)$$

where we have assumed a constitutive relation of the form

$$\sigma_s^{ij}(\mathbf{x},t) = \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} \psi^{ijkl}(\mathbf{x},t-\tau)\partial_l u_s^k(\mathbf{x},\tau) + M_s^{ij}(\mathbf{x},t), \quad (2)$$

and $q_s^i(\mathbf{x}, t)$ is a component of an additive stochastic noise process $\mathbf{q}_s(\mathbf{x}, t)$ with Gaussian statistics and zero mean, $\partial_t$ represents the partial derivative with respect to time $t$, and $\partial_j$ represents the partial derivative with respect to the $j$-th component of $\mathbf{x}$, $\partial/\partial x^j$. In this article, we express summations over indices explicitly and do not use repeated indices to imply Einstein summation convention, unless explicitly stated otherwise. We consider the homogeneous initial condition and the free-surface boundary condition, also with additive residuals,

$$u_s^i(\mathbf{x},0) = a_s^i(\mathbf{x}), \partial_t u_s^i(\mathbf{x},0) = b_s^i(\mathbf{x}), \quad (3)$$

$$\sum_j \hat{n}^j(\mathbf{x}) \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} \psi^{ijkl}(\mathbf{x}, t-\tau)\partial_l u_s^k(\mathbf{x},\tau) = c_s^i(\mathbf{x},t), \quad \mathbf{x} \in \partial V, \quad (4)$$

where $a_s^i(\mathbf{x})$, $b_s^i(\mathbf{x})$ and $c_s^i(\mathbf{x}, t)$ are components of additive stochastic noise processes $\mathbf{a}_s(\mathbf{x})$, $\mathbf{b}_s(\mathbf{x}, t)$ and $\mathbf{c}_s(\mathbf{x}, t)$ with Gaussian statistics and zero mean, vector $\hat{\mathbf{n}}(\mathbf{x})$, with its components $\hat{n}^j(\mathbf{x})$, denotes the normal direction on $\partial V$. The extension to other types of initial and boundary conditions is straight forward.

Some common origins of model uncertainties, in addition to the uncertainties in the model parameters, include but are not limited to, the errors in the mathematical model, the numerical method used for solving the mathematical model, and errors in the initial and boundary conditions. Possible reasons that could cause deviations from the traction-free boundary condition might include lithosphere-atmosphere coupling, deviation from the continuum model for materials in the near-surface environment, numerical errors caused by, for instance, errors in the numerical representation of the actual topography, etc. The Earth is constantly in motion. The quiescent-past initial condition for one seismic event could be violated in practice if we consider motions caused by, for instance, other seismic events, the Earth's ambient noise field and the constant hum of the Earth caused by atmosphere–ocean-seafloor coupling (RHIE and ROMANOWICZ 2004).

For each seismogram, we consider a finite set of data functionals, indexed by $n$, that measure the misfit between the observed seismogram $\bar{u}_s^i(\mathbf{x}_r, t)$ and the model-predicted (i.e. synthetic) seismogram $u_s^i(\mathbf{x}_r, t)$,

$$d_{srn}^i = D_n\left[\bar{u}_s^i(\mathbf{x}_r, t), u_s^i(\mathbf{x}_r, t)\right], \tag{5}$$

where $r$ is an index to reference different receivers. We assume $D_n$ is constructed to satisfy

$$D_n\left[u_s^i(\mathbf{x}_r, t), u_s^i(\mathbf{x}_r, t)\right] = 0. \tag{6}$$

The measurement process generally involves nonlinear operations both on observed seismograms and on synthetic seismograms.

### 2.1. Prior Probability Densities

Assume that we have available a prior estimate for mass density ${}_0\rho(\mathbf{x})$. Furthermore, we assume that the poorly known mass density $\rho(\mathbf{x})$ has distributed errors with Gaussian statistics and an error covariance $C_{\rho\rho}(\mathbf{x}, \mathbf{x}')$. We define the inverse of the error covariance as $W_{\rho\rho}(\mathbf{x}, \mathbf{x}')$, which satisfies

$$\int_V \mathrm{d}V(\mathbf{x}'')C_{\rho\rho}(\mathbf{x}, \mathbf{x}'')W_{\rho\rho}(\mathbf{x}'', \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'), \tag{7}$$

where $\delta(\mathbf{x}-\mathbf{x}')$ is the delta function. The prior probability density function for mass density $\rho(\mathbf{x})$ can then be expressed as

$$p(\rho) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x}) \int_V \mathrm{d}V(\mathbf{x}') \right.$$
$$\left.[\rho(\mathbf{x}) - {}_0\rho(\mathbf{x})]W_{\rho\rho}(\mathbf{x}, \mathbf{x}')[\rho(\mathbf{x}') - {}_0\rho(\mathbf{x}')]\right\}. \tag{8}$$

We have not included a denominator that normalizes the right-hand-side, thereby using proportional to, $\propto$, rather than equal to, $=$ .

Assume that the component of the fourth-order rate-of-relaxation tensor, $\psi^{ijkl}(\mathbf{x},t)$, follows Gaussian statistics with a prior estimate ${}_0\psi^{ijkl}(\mathbf{x},t)$ and error covariance $C_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}, t; \mathbf{x}', t')$. We define the inverse of the error covariance as $W_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}, t, \mathbf{x}', t')$, which satisfies

$$\int_V \mathrm{d}V(\mathbf{x}'') \int_0^T \mathrm{d}t'' C_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}, t; \mathbf{x}'', t'')$$
$$\times W_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}'', t''; \mathbf{x}', t') = \delta(\mathbf{x} - \mathbf{x}')\delta(t - t'), \tag{9}$$

where [0, $T$] is the time interval of our analysis. The prior probability density for $\psi^{ijkl}(\mathbf{x}, t)$ can then be expressed as

$$p\left(\psi^{ijkl}\right) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x}) \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t \right.$$
$$\times \int_0^T \mathrm{d}t' \left[\psi^{ijkl}(\mathbf{x}, t) - {}_0\psi^{ijkl}(\mathbf{x}, t)\right] W_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}, t; \mathbf{x}', t')$$
$$\left. \times \left[\psi^{ijkl}(\mathbf{x}', t') - {}_0\psi^{ijkl}(\mathbf{x}', t')\right]\right\}. \tag{10}$$

For notational convenience in the following derivation, we express the prior probability density for $\boldsymbol{\psi}(\mathbf{x}, t)$ as a multiplication of the probability density for each component

$$p(\boldsymbol{\psi}) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x}) \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t \int_0^T \mathrm{d}t' \right.$$
$$\times \sum_{ijkl} \left( \left[\psi^{ijkl}(\mathbf{x}, t) - {}_0\psi^{ijkl}(\mathbf{x}, t)\right] \cdot \right.$$
$$\left.\left. W_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}, t; \mathbf{x}', t')\left[\psi^{ijkl}(\mathbf{x}', t') - {}_0\psi^{ijkl}(\mathbf{x}', t')\right]\right)\right\}. \tag{11}$$

Because of the symmetry of the rate-of-relaxation tensor, out of the 81 components, only 21 of them are independent. To incorporate equality constraints among elastic parameters, the delta distribution can be introduced to represent the corresponding conditional probabilities. The delta distribution can be considered as the Gaussian distribution with its variance approaching zero. We note that by modifying Eq. 11 and the structure of $W_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x},t,\mathbf{x}';t')$ to account for cross-dependences among components of $\boldsymbol{\psi}(\mathbf{x}, t)$, the derivations in the following can be extended to account for cases where the errors in the components of $\boldsymbol{\psi}(\mathbf{x}, t)$ are not independent. The elastic parameters also need to satisfy certain stability requirements (Backus 1962), which can result in positivity constraints. In this case, the Gaussian assumption is valid only locally when the reference elastic tensor satisfies the stability requirements and the variances are not too large.

We also define probability density functions for the body-force density $\mathbf{f}_s(\mathbf{x}, t)$ and the moment tensor density $\mathbf{M}_s(\mathbf{x}, t)$ using an assumption of Gaussian statistics and we obtain

$$p(\mathbf{f}_s) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x})\int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t\int_0^T \mathrm{d}t' \times \sum_i \left(\left[f_s^i(\mathbf{x},t) - {}_0f_s^i(\mathbf{x},t)\right] \times W_{f_s^i f_s^i}(\mathbf{x},t;\mathbf{x}',t')\left[f_s^i(\mathbf{x}',t') - {}_0f_s^i(\mathbf{x}',t')\right]\right)\right\}, \tag{12}$$

$$p(\mathbf{M}_s) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x})\int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t\int_0^T \mathrm{d}t' \times \sum_{ij} \left(\left[M_s^{ij}(\mathbf{x},t) - {}_0M_s^{ij}(\mathbf{x},t)\right] W_{M_s^{ij}M_s^{ij}}(\mathbf{x},t;\mathbf{x}',t')\left[M_s^{ij}(\mathbf{x}',t') - {}_0M_s^{ij}(\mathbf{x}',t')\right]\right)\right\}, \tag{13}$$

where ${}_0f_s^i(\mathbf{x}, t)$ and ${}_0M_s^{ij}(\mathbf{x}, t)$ are components of ${}_0\mathbf{f}_s(\mathbf{x}, t)$ and ${}_0\mathbf{M}_s(\mathbf{x},t)$, which are prior estimates of $\mathbf{f}_s(\mathbf{x}, t)$ and $\mathbf{M}_s(\mathbf{x}, t)$, $W_{f_s^i f_s^i}(\mathbf{x},t;\mathbf{x}',t')$ and $W_{M_s^{ij}M_s^{ij}}(\mathbf{x},t;\mathbf{x}',t')$ are inverses of the corresponding error covariance functions with definitions similar to Eq. 9.

For the model residual introduced in Eq. 1, as well as residuals in the initial and boundary conditions in Eqs. 3 and 4, we also define Gaussian probability density functions and the values of the prior estimates are 0,

$$p(\mathbf{q}_s) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x})\int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t\int_0^T \mathrm{d}t' \times \sum_i q_s^i(\mathbf{x},t)W_{q_s^i q_s^i}(\mathbf{x},t;\mathbf{x}',t')q_s^i(\mathbf{x}',t')\right\}, \tag{14}$$

$$p(\mathbf{a}_s) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x})\int_V \mathrm{d}V(\mathbf{x}') \times \sum_i a_s^i(\mathbf{x})W_{a_s^i a_s^i}(\mathbf{x},\mathbf{x}')a_s^i(\mathbf{x}')\right\}$$
$$p(\mathbf{b}_s) \propto \exp\left\{-\frac{1}{2}\int_V \mathrm{d}V(\mathbf{x})\int_V \mathrm{d}V(\mathbf{x}') \times \sum_i b_s^i(\mathbf{x})W_{b_s^i b_s^i}(\mathbf{x},\mathbf{x}')b_s^i(\mathbf{x}')\right\}, \tag{15}$$

$$p(\mathbf{c}_s) \propto \exp\left\{-\frac{1}{2}\int_{\partial V} \mathrm{d}S(\mathbf{x})\int_{\partial V} \mathrm{d}S(\mathbf{x}')\int_0^T \mathrm{d}t \times \int_0^T \mathrm{d}t' \sum_i c_s^i(\mathbf{x},t)W_{c_s^i c_s^i}(\mathbf{x},t;\mathbf{x}',t')c_s^i(\mathbf{x}',t')\right\}, \tag{16}$$

Here we have introduced $W_{q_s^i q_s^i}$, $W_{a_s^i a_s^i}$, $W_{b_s^i b_s^i}$ and $W_{c_s^i c_s^i}$, which are inverses of their corresponding covariance functions. The model residuals are assumed to have zero biases, but the derivation in the following can also be extended to model residuals with non-zero biases without too much difficulty.

### 2.2. Bayes' Theorem

We assume the misfit measurement $d_{srn}^i$ follows Gaussian statistics with an estimate given by $D_n[u_s^i(\mathbf{x}_r, t), u_s^i(\mathbf{x}_r, t)]$. Considering Eq. 6, we can express the conditional probability density function for $d_{srn}^i$ as

$$p\left(d_{srn}^i|\rho,\boldsymbol{\psi},\mathbf{f}_s,\mathbf{M}_s,\mathbf{q}_s,\mathbf{a}_s,\mathbf{b}_s,\mathbf{c}_s\right) \propto \exp\left\{-\frac{W_{d_{srn}^i d_{srn}^i}}{2}(d_{srn}^i)^2\right\}, \tag{17}$$

where $W_{d^i_{srm}d^i_{srm}}$ is the inverse of the variance of misfit measurement $d^i_{srm}$. Here we introduce a measurement vector $\mathbf{d}_s$ with its components given by all misfit measurements for source index $s$. We assume the measurements $d^i_{srm}$ are uncorrelated, the conditional probability for $\mathbf{d}_s$ can be expressed as

$$p(\mathbf{d}_s|\rho,\boldsymbol{\psi},\mathbf{f}_s,\mathbf{M}_s,\mathbf{q}_s,\mathbf{a}_s,\mathbf{b}_s,\mathbf{c}_s) \propto \exp\left\{-\frac{\mathbf{d}_s^{\mathrm{T}}\mathbf{W}_{\mathbf{d}_s\mathbf{d}_s}\mathbf{d}_s}{2}\right\}, \tag{18}$$

where $\mathbf{W}_{\mathbf{d}_s\mathbf{d}_s}$ is a diagonal matrix with its diagonal elements given by $W_{d^i_{sm}d^i_{sm}}$. We note that by modifying the form of $\mathbf{W}_{\mathbf{d}_s\mathbf{d}_s}$, Eq. 18 can be generalized to account for correlated $d^i_{srm}$ also. For seismic data assimilation problems involving multiple seismic sources, the conditional probability density can be expressed as

$$\begin{aligned} &p(\mathbf{d}_1,\cdots,\mathbf{d}_{N_s}|\rho,\boldsymbol{\psi},\mathbf{f}_1,\mathbf{M}_1,\mathbf{q}_1,\mathbf{a}_1,\mathbf{b}_1,\mathbf{c}_1,\cdots,\mathbf{f}_{N_s},\\ &\quad \mathbf{M}_{N_s},\mathbf{q}_{N_s},\mathbf{a}_{N_s},\mathbf{b}_{N_s},\mathbf{c}_{N_s})\\ &\propto \exp\left\{-\frac{1}{2}\sum_{s=1}^{N_s}\mathbf{d}_s^{\mathrm{T}}\mathbf{W}_{\mathbf{d}_s\mathbf{d}_s}\mathbf{d}_s\right\}, \end{aligned} \tag{19}$$

where $N_s$ is the total number of seismic sources used. The likelihood function can then be expressed as

$$\begin{aligned} &l(\rho,\boldsymbol{\psi},\mathbf{f}_1,\mathbf{M}_1,\mathbf{q}_1,\mathbf{a}_1,\mathbf{b}_1,\mathbf{c}_1,\cdots,\mathbf{f}_{N_s},\\ &\quad \mathbf{M}_{N_s},\mathbf{q}_{N_s},\mathbf{a}_{N_s},\mathbf{b}_{N_s},\mathbf{c}_{N_s}|\mathbf{d}_1,\cdots,\mathbf{d}_{N_s})\\ &\propto \exp\left\{-\frac{1}{2}\sum_{s=1}^{N_s}\mathbf{d}_s^{\mathrm{T}}\mathbf{W}_{\mathbf{d}_s\mathbf{d}_s}\mathbf{d}_s\right\}. \end{aligned} \tag{20}$$

Following Bayes' theorem, we obtain the posterior probability density

$$\begin{aligned} &p(\rho,\boldsymbol{\psi},\mathbf{f}_1,\mathbf{M}_1,\mathbf{q}_1,\mathbf{a}_1,\mathbf{b}_1,\mathbf{c}_1,\cdots,\mathbf{f}_{N_s},\\ &\quad \mathbf{M}_{N_s},\mathbf{q}_{N_s},\mathbf{a}_{N_s},\mathbf{b}_{N_s},\mathbf{c}_{N_s}|\mathbf{d}_1,\cdots,\mathbf{d}_{N_s})\\ &\propto l(\rho,\boldsymbol{\psi},\mathbf{f}_1,\mathbf{M}_1,\mathbf{q}_1,\mathbf{a}_1,\mathbf{b}_1,\mathbf{c}_1,\cdots,\mathbf{f}_{N_s},\\ &\quad \mathbf{M}_{N_s},\mathbf{q}_{N_s},\mathbf{a}_{N_s},\mathbf{b}_{N_s},\mathbf{c}_{N_s}|\mathbf{d}_1,\cdots,\mathbf{d}_{N_s})\\ &\cdot p(\rho)p(\boldsymbol{\psi})\prod_{s=1}^{N_s}p(\mathbf{f}_s)p(\mathbf{M}_s)p(\mathbf{q}_s)p(\mathbf{a}_s)p(\mathbf{b}_s)p(\mathbf{c}_s)\\ &\quad = \exp\left(-\frac{1}{2}\mathbf{Y}\right) \end{aligned} \tag{21}$$

where we have defined an objective function $\mathbf{Y}$, which has the expression

$$\begin{aligned} \mathbf{Y} = &\sum_{s=1}^{N_s}\mathbf{d}_s^{\mathrm{T}}\mathbf{W}_{\mathrm{d},\mathrm{d}_s}\mathbf{d}_s\\ &+\sum_{s=1}^{N_s}\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')\int_0^T\mathrm{d}t\int_0^T\mathrm{d}t'\\ &\times\sum_i q_s^i(\mathbf{x},t)W_{q_s^iq_s^i}(\mathbf{x},t;\mathbf{x}',t')q_s^i(\mathbf{x}',t')\\ &+\sum_{s=1}^{N_s}\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')\sum_i a_s^i(\mathbf{x})W_{a_s^ia_s^i}(\mathbf{x},\mathbf{x}')a_s^i(\mathbf{x}')\\ &+\sum_{s=1}^{N_s}\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')\sum_i b_s^i(\mathbf{x})W_{b_s^ib_s^i}(\mathbf{x},\mathbf{x}')b_s^i(\mathbf{x}')\\ &+\sum_{s=1}^{N_s}\int_{\partial V}\int_{\partial V} \mathrm{d}S(\mathbf{x})\mathrm{d}S(\mathbf{x}')\int_0^T\mathrm{d}t\int_0^T\mathrm{d}t'\\ &\times\sum_i c_s^i(\mathbf{x},t)W_{c_s^ic_s^i}(\mathbf{x},t;\mathbf{x}',t')c_s^i(\mathbf{x}',t')\\ &+\sum_{s=1}^{N_s}\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')\int_0^T\mathrm{d}t\int_0^T\mathrm{d}t'\\ &\times\sum_i\left(\left[f_s^i(\mathbf{x},t)-{}_0f_s^i(\mathbf{x},t)\right]\right.\\ &\left.W_{f_s^if_s^i}(\mathbf{x},t;\mathbf{x}',t')\left[f_s^i(\mathbf{x}',t')-{}_0f_s^i(\mathbf{x}',t')\right]\right)\\ &+\sum_{s=1}^{N_s}\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')\int_0^T\mathrm{d}t\int_0^T\mathrm{d}t'\\ &\times\sum_{ij}\left(\left[M_s^i(\mathbf{x},t)-{}_0M_s^i(\mathbf{x},t)\right]\right.\\ &\left.W_{M_s^{ij}M_s^{ij}}(\mathbf{x},t;\mathbf{x}',t')\left[M_s^{ij}(\mathbf{x}',t')-{}_0M_s^{ij}(\mathbf{x}',t')\right]\right)\\ &+\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')[\rho(\mathbf{x})-{}_0\rho(\mathbf{x})]W_{\rho\rho}(\mathbf{x},\mathbf{x}')[\rho(\mathbf{x}')-{}_0\rho(\mathbf{x}')]\\ &+\int_V\int_V \mathrm{d}V(\mathbf{x})\mathrm{d}V(\mathbf{x}')\int_0^T\mathrm{d}t\int_0^T\mathrm{d}t'\sum_{ijkl}\left(\left[\psi^{ijkl}(\mathbf{x},t)-{}_0\psi^{ijkl}(\mathbf{x},t)\right]\right.\\ &\left.W_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x},t;\mathbf{x}',t')\left[\psi^{ijkl}(\mathbf{x}',t')-{}_0\psi^{ijkl}(\mathbf{x}',t')\right]\right). \end{aligned} \tag{22}$$

Thus, for Gaussian priors and likelihood, maximization of the posterior probability density in Eq. 21 is equivalent to minimization of $\mathbf{Y}$ as defined in Eq. 22. The objective function $\mathbf{Y}$ will have a global minimum, but it may not be unique. It may also possess a number of local minima and there is a risk of converging into one of those local minima.

### 2.3. *Euler–Lagrange Equations*

To minimize the objective function $\mathbf{Y}$ in Eq. 22, we can invoke Hamilton's principle (i.e. principle of stationary action) by calculating the variational derivative of $\mathbf{Y}$ and requiring that it approaches zero while the arbitrary perturbations of the generalized coordinates go to zero. The resulting equations are called Euler–Lagrange equations.

We first define an "adjoint" vector field ${}^{\dagger}\mathbf{u}_s(\mathbf{x},t)$ with its components given by ${}^{\dagger}u_s^i(\mathbf{x},t)$ and satisfying

$$ {}^{\dagger}u_s^i(\mathbf{x},T-t)=\int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t' W_{q_s^i q_s^i}(\mathbf{x},t;\mathbf{x}',t')q_s^i(\mathbf{x}',t') \tag{23}$$

If we multiply $C_{q_s^i q_s^i}(\mathbf{x},t;\mathbf{x}'',t'')$ and carry out the spatial integration on $\mathbf{x}$ and the temporal integration on $t$ on both sides of Eq. 23, we obtain

$$\rho(\mathbf{x})\partial_t^2 u_s^i(\mathbf{x},t)-\sum_j \partial_j \times\left[\int_{-\infty}^{+\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l u_s^k(\mathbf{x},\tau)+M_s^{ij}(\mathbf{x},t)\right] -f_s^i(\mathbf{x},t)=\int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t' C_{q_s^i q_s^i}(\mathbf{x},t;\mathbf{x}',t')^{\dagger}u_s^i(\mathbf{x}',T-t'), \tag{24}$$

which is the original seismic wave equation with a representation of the model residual involving a product between the model residual covariance and the adjoint variable.

The variational derivative of $\mathbf{Y}$ with respect to the displacement field $\mathbf{u_s}(\mathbf{x},t)$ involves the first two terms on the right-hand-side of Eq. 22. We introduce the seismogram perturbation kernel, $J_{srn}^i(t)$, which is defined as the Fréchet kernel of the misfit measurement $d_{srn}^i$ with respect to the synthetic seismogram

$$\delta d_{srn}^i=\int_0^T \mathrm{d}t J_{srn}^i(t)\delta u_s^i(\mathbf{x}_r,t). \tag{25}$$

A perturbation of the first term on the right-hand-side of Eq. 22 can now be expressed as

$$\sum_{s=1}^{N_s}\delta\mathbf{d}_s^{\mathrm{T}}\mathbf{W}_{\mathbf{d}_s\mathbf{d}_s}\mathbf{d}_s=\sum_{s=1}^{N_s}\sum_i\int_V \mathrm{d}V(\mathbf{x})\int_0^T \mathrm{d}t \times\sum_m \delta(\mathbf{x}-\mathbf{x}_r)J_{srn}^i(t)W_{d_{srn}^i d_{srn}^i}d_{srn}^i\delta u_s^i(\mathbf{x},t). \tag{26}$$

Using the adjoint variable defined in Eq. 23 and considering the seismic wave equation as defined in Eq. 1, a perturbation of the second term on the right-hand-side of Eq. 22 can be expressed as

$$\sum_{s=1}^{N_s}\int_V \mathrm{d}V(\mathbf{x})\int_0^T \mathrm{d}t\sum_i \delta q_s^i(\mathbf{x},t)^{\dagger}u_s^i(\mathbf{x},T-t)=\sum_{s=1}^{N_s}\int_V \mathrm{d}V(\mathbf{x})\int_0^T \mathrm{d}t \sum_i\left(\left\{\rho(\mathbf{x})\partial_t^2\delta u_s^i(\mathbf{x},t)-\sum_j\partial_j\left[\int_{-\infty}^{+\infty}d\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l\delta u_s^k(\mathbf{x},\tau)\right]\right\}{}^{\dagger}u_s^i(\mathbf{x},T-t). -\sum_j\partial_j M_s^{ij}(\mathbf{x},t)^{\dagger}u_s^i(\mathbf{x},T-t)-f_s^i(\mathbf{x},t)^{\dagger}u_s^i(\mathbf{x},T-t)\right) \tag{27}$$

We can transfer the second-order temporal derivative on $\delta u_s^i(\mathbf{x},t)$ to the adjoint variable through integration-by-parts and we obtain

$$\sum_{s=1}^{N_s}\int_V \mathrm{d}V(\mathbf{x})\rho(\mathbf{x})\sum_i\int_0^T \mathrm{d}t\partial_t^2\delta u_s^i(\mathbf{x},t)^{\dagger}u_s^i(\mathbf{x},T-t) =\sum_{s=1}^{N_s}\sum_i\left\{\int_V \mathrm{d}V(\mathbf{x})\rho(\mathbf{x})\left[\delta\dot{u}_s^i(\mathbf{x},t)^{\dagger}u_s^i(\mathbf{x},T-t)|_0^T +\delta u_s^i(\mathbf{x},t)^{\dagger}(\dot{u})_s^i(\mathbf{x},T-t)|_0^T\right]. +\int_V \mathrm{d}V(\mathbf{x})\rho(\mathrm{x})\int_0^T \mathrm{d}t^{\dagger}\ddot{u}_s^i(\mathbf{x},T-t)\delta u_s^i(\mathbf{x},t)\right\} \tag{28}$$

The spatial derivatives with respect to $\delta u_s^i(\mathbf{x},t)$ in Eq. 27 can also be transferred to the adjoint variable using Green's identity. After some algebraic manipulations, the second term on the right-hand-side of Eq. 27 can be expressed as

$$
\begin{aligned}
&\sum_{s=1}^{N_s}\sum_i\int_V \mathrm{d}V(\mathbf{x})\int_0^T \mathrm{d}t\sum_j \partial_j\left[\int_{-\infty}^{+\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l\delta u_s^k(\mathbf{x},\tau)\right]{}^{\dagger}u_s^i(\mathbf{x},T-t)\\
&=\sum_{s=1}^{N_s}\sum_i\int_{\partial V}\mathrm{d}S(\mathbf{x})\int_0^T \mathrm{d}t\left\{{}^{\dagger}u_s^i(\mathbf{x},T-t)\sum_j\hat{n}^j(\mathbf{x})\left[\int_{-\infty}^{+\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l\delta u_s^k(\mathbf{x},\tau)\right]\right.\\
&\qquad\left.-\sum_j\hat{n}^j(\mathbf{x})\left[\int_{\infty}^{+\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},T-t-\tau)\partial_l^{\dagger}u_s^k(\mathbf{x},\tau)\right]\delta u_s^i(\mathbf{x},t)\right\}\\
&\qquad+\int_V \mathrm{d}V(\mathbf{x})\int_0^T \mathrm{d}t\sum_j\partial_j\left[\int_{\infty}^{+\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},T-t-\tau)\partial_l^{\dagger}u_s^k(\mathbf{x},\tau)\right]\delta u_s^i(\mathbf{x},t)
\end{aligned}
\tag{29}
$$

Collecting terms containing $\delta u_s^i(\mathbf{x},t)$ in Eqs. 26–29, we obtain the variational derivative of $\mathbf{Y}$ with respect to $\mathbf{u}_s(\mathbf{x},t)$ as

$$
\begin{aligned}
\frac{\delta\mathbf{Y}}{2\delta u_s^i(\mathbf{x},t)}&=\rho(\mathbf{x}){}^{\dagger}\ddot{u}_s^i(\mathbf{x},T-t)\\
&-\sum_j\partial_j\left[\int_{-\infty}^{+\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},T-t-\tau)\partial_j^{\dagger}u_s^k(\mathbf{x},\tau)\right].\\
&+\sum_{rm}\delta(\mathbf{x}-\mathbf{x}_r)J_{srn}^i(t)W_{d_{srn}^i d_{srn}^i}d_{srn}^i
\end{aligned}
\tag{30}
$$

If we require $\delta\mathbf{Y}\to 0$ for arbitrary $\delta u_s^i(\mathbf{x},t)\to 0$, we must have

$$
\begin{aligned}
&\rho(\mathbf{x})\partial_t^{2\dagger}u_s^i(\mathbf{x},t)\\
&-\sum_j\partial_j\left[\int_{\infty}^{\infty}\mathrm{d}\tau\sum_{kl}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l^{\dagger}u_s^k(\mathbf{x},\tau)\right]\\
&=-\sum_m\delta(\mathbf{x}-\mathbf{x}_r)J_{sm}^i(T-t)W_{d_{sm}^i d_{sm}^i}d_{sm}^i.
\end{aligned}
\tag{31}
$$

Here we have reversed the sign of time by making the change $T-t\to t$. Equation 31 is the adjoint wave equation, which is forced by misfit measurements located at receiver locations and weighted by the corresponding time-reversed seismogram perturbation kernels and the inverse of the data covariance.

The variational derivatives of $\mathbf{Y}$ with respect to the initial conditions $\mathbf{u}_s(\mathbf{x},0)$ and $\dot{\mathbf{u}}_s(\mathbf{x},0)$ can be obtained by collecting terms containing $\delta u_s^i(\mathrm{x},0)$ and $\delta\dot{u}_s^i(\mathbf{x},0)$ from Eq. 28 and considering the third and fourth terms on the right-hand-side of Eq. 22. The variational derivatives with respect to initial conditions are

$$
\begin{aligned}
\frac{\delta\mathbf{Y}}{2\delta u_s^i(\mathbf{x},0)}&=\int_V \mathrm{d}V(\mathbf{x}')W_{a_s^i a_s^i}(\mathbf{x},\mathbf{x}')u_s^i(\mathbf{x}',0)\\
&\quad-{}^{\dagger}\dot{u}_s^j(\mathbf{x},T),
\end{aligned}
\tag{32}
$$

$$
\begin{aligned}
\frac{\delta\mathbf{Y}}{2\delta\dot{u}_s^i(\mathbf{x},0)}&=\int_V \mathrm{d}V(\mathbf{x}')W_{b_s^i b_s^i}(\mathbf{x},\mathbf{x}')\dot{u}_s^i(\mathbf{x}',0)\\
&\quad-{}^{\dagger}u_s^j(\mathbf{x},T).
\end{aligned}
\tag{33}
$$

Requiring $\delta\mathbf{Y}\to 0$ for arbitrary $\delta u_s^i(\mathbf{x},t)\to 0$ and $\delta\dot{u}_s^i(\mathbf{x},0)\to 0$, we obtain

$$
u_s^i(\mathbf{x},0)=\int_V \mathrm{d}V(\mathbf{x}')C_{a_s^i a_s^i}(\mathbf{x},\mathbf{x}'){}^{\dagger}\dot{u}_s^i(\mathbf{x}',T),
\tag{34}
$$

$$
\dot{u}_s^i(\mathbf{x},0)=\int_V \mathrm{d}V(\mathbf{x}')C_{b_s^i b_s^i}(\mathbf{x},\mathbf{x}'){}^{\dagger}u_s^i(\mathbf{x}',T).
\tag{35}
$$

Considering Eq. 28, the variational derivatives of $\mathbf{Y}$ with respect to $\mathbf{u}_s(\mathbf{x},T)$ and $\dot{\mathbf{u}}_s(\mathbf{x},T)$ provide us a set of initial conditions for the adjoint wave-field at time 0,

$${}^{\dagger}u_s^i(\mathbf{x},0)=0,\quad \partial_t^{\dagger}u_s^i(\mathbf{x},0)=0.
\tag{36}$$

The variational derivative of $\mathbf{Y}$ with respect to the boundary condition involves the fifth term in the

right-hand-side of Eq. 22 and the first surface integral on the right-hand-side of Eq. 29. Requiring this variational derivative to be zero, we obtain

$$\sum_j \hat{n}^j(\mathbf{x}) \left[ \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} \psi^{ijkl}(\mathbf{x}, t-\tau) \partial_l u_s^k(\mathbf{x}, \tau) \right] = \int_{\partial V} \mathrm{d}S(\mathbf{x}') \int_0^T \mathrm{d}t' C_{c_s^i c_s^i}(\mathbf{x}, t; \mathbf{x}', t')^\dagger u_s^i(\mathbf{x}', T-t'). \tag{37}$$

The variational derivative of $\mathbf{Y}$ with respect to $u_s^i(\mathbf{x}, t)$ for $\mathbf{x} \in \partial V$ is provided by the second surface integral on the right-hand-side of Eq. 29. Requiring this variational derivative to be zero, we obtain the boundary condition for the adjoint wave equation,

$$\sum_j \hat{n}^j(\mathbf{x}) \left[ \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} \psi^{ijkl}(\mathbf{x}, T-t-\tau) \partial_l^\dagger u_s^k(\mathbf{x}, \tau) \right] = 0. \tag{38}$$

The variational derivative of $\mathbf{Y}$ with respect to the body-force density $\mathbf{f}_s$ involves the second and the sixth terms on the right-hand-side of Eq. 22. Requiring this variational derivative to be zero, we obtain

$$f_s^i(\mathbf{x}, t) = {}_0 f_s^i(\mathbf{x}, t) + \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t' C_{f_s^i f_s^i}(\mathbf{x}, t; \mathbf{x}', t')^\dagger u_s^i(\mathbf{x}', T-t'). \tag{39}$$

The variational derivative of $\mathbf{Y}$ with respect to the moment tensor density $\mathbf{M}_s$ involves the second and the seventh terms on the right-hand-side of Eq. 22. Requiring this variational derivative to be zero and applying the divergence theorem, we obtain

$$M_s^{ij}(\mathbf{x}, t) = {}_0 M_s^{ij}(\mathbf{x}, t) - \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t' \times C_{M_s^{ij} M_s^{ij}}(\mathbf{x}, t; \mathbf{x}', t') \partial_j^\dagger u_s^i(\mathbf{x}', T-t'). \tag{40}$$

The variational derivative of $\mathbf{Y}$ with respect to the mass density $\rho(\mathbf{x})$ involves the second and the eighth terms on the right-hand-side of Eq. 22. Collecting terms containing $\delta\rho(\mathbf{x})$, we obtain

$$\rho(\mathbf{x}) = {}_0\rho(\mathbf{x}) - \int_V \mathrm{d}V(\mathbf{x}') C_{\rho\rho}(\mathbf{x}, \mathbf{x}') \sum_{s=1}^{N_s} \int_0^T \mathrm{d}t \times \sum_i {}^\dagger u_s^i(\mathbf{x}', T-t) \partial_t^2 u_s^i(\mathbf{x}', t) \tag{41}$$

The variational derivative of $\mathbf{Y}$ with respect to the rate-of-relaxation tensor $\boldsymbol{\psi}(\mathbf{x},t)$ involves the second and the last terms on the right-hand-side of Eq. 22. Collecting terms containing $\delta\psi^{ijkl}(\mathbf{x},t)$and applying the divergence theorem, we obtain

$$\psi^{ijkl}(\mathbf{x}, t) = {}_0\psi^{ijkl}(\mathbf{x}, t) - \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t' \left\{ C_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x}, t; \mathbf{x}', t') \sum_{s=1}^{N_s} \int_o^T \mathrm{d}\tau \left[ \partial_j {}^\dagger u_s^i(\mathbf{x}', T-t'-\tau) \partial_l u_s^k(\mathbf{x}', \tau) \right] \right\}. \tag{42}$$

By now, we have derived the Euler–Lagrange equations for the full-wave seismic data assimilation problem. The equations can be solved iteratively, i.e., the solution that maximizes the posterior probability density can be used as the prior estimate for the next iteration. We summarize the equations in the following and use index $\gamma$ to indicate iteration number. The Euler–Lagrange equations consist of the adjoint wave equation, together with the appropriate initial and boundary conditions,

$$\begin{aligned} &{}_\gamma\rho(\mathbf{x}) \partial_t^2 {}^\dagger u_s^i(\mathbf{x}, t) \\ &\quad - \sum_j \partial_j \left[ \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} {}_\gamma\psi^{ijkl}(\mathbf{x}, t-\tau) \partial_l {}^\dagger u_s^k(\mathbf{x}, \tau) \right], \\ &\quad = -\sum_{rn} \delta(\mathbf{x} - \mathbf{x}_r) J_{srn}^i(T-t) W_{d_{srn}^i d_{srn}^i} d_{srn}^i \end{aligned} \tag{43}$$

$$ {}^\dagger u_s^i(\mathbf{x}, 0) = 0, \quad \partial_t {}^\dagger u_s^i(\mathbf{x}, 0) = 0, \tag{44}$$

$$\sum_j \hat{n}^j(\mathbf{x}) \left[ \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} {}_\gamma\psi^{ijkl}(\mathbf{x}, t-\tau) \partial_l {}^\dagger u_s^k(\mathbf{x}, \tau) \right] = 0, \tag{45}$$

a set of equations for updating the seismic source and structure parameters,

$$ {}_{\gamma+1}f_s^i(\mathbf{x},t) = {}_{\gamma}f_s^i(\mathbf{x},t) + \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t' C_{f_s^i f_s^i}(\mathbf{x},t;\mathbf{x}',t')^{\dagger} u_s^i(\mathbf{x}',T-t'), \tag{46} $$

$$ {}_{\gamma+1}M_s^{ij}(\mathbf{x},t) = {}_{\gamma}M_s^{ij}(\mathbf{x},t) - \int_V \mathrm{d}V(\mathbf{x}') \times \int_0^T \mathrm{d}t' C_{M_s^{ij}M_s^{ij}}(\mathbf{x},t;\mathbf{x}',t')\partial_j^{\dagger} u_s^i(\mathbf{x}',T-t'), \tag{47} $$

$$ {}_{\gamma+1}\rho(\mathbf{x}) = {}_{\gamma}\rho(\mathbf{x}) - \int_V \mathrm{d}V(\mathbf{x}')C_{\rho\rho}(\mathbf{x},\mathbf{x}') \sum_{s=1}^{N_s} \int_0^T \mathrm{d}t \sum_i {}^{\dagger}u_s^i(\mathbf{x}',T-t)\partial_t^2 u_s^i(\mathbf{x}',t), \tag{48} $$

$$ {}_{\gamma+1}\psi^{ijkl}(\mathbf{x},t) = {}_{\gamma}\psi^{ijkl}(\mathbf{x},t) - \int_V \mathrm{d}V(\mathbf{x}') \int_0^T \mathrm{d}t' \times \left\{ C_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x},t;\mathbf{x}',t') \sum_{s=1}^{N} \int_0^T \mathrm{d}\tau \left[ \partial_j^{\dagger} u_s^i(\mathbf{x}',T-t'-\tau)\partial_l u_s^k(\mathbf{x}',\tau) \right] \right\}, \tag{49} $$

and the forward seismic wave equation, together with the initial and boundary conditions,

$$ {}_{\gamma+1}\rho(\mathbf{x})\partial_t^2 u_s^i(\mathbf{x},t) - \sum_j \partial_j \left[ \int_{-\infty}^{+\infty} \mathrm{d}\tau \sum_{kl} {}_{\gamma+1}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l u_s^k(\mathbf{x},\tau) + {}_{\gamma+1}M_s^{ij}(\mathbf{x},t) \right] - {}_{\gamma+1}f_s^i(\mathbf{x},t) = \int_V \mathrm{d}V(\mathbf{x}') \int_0^{t_1} \mathrm{d}t' C_{q_s^i q_s^i}(\mathbf{x},t;\mathbf{x}',t')^{\dagger} u_s^i(\mathbf{x}',T-t'), \tag{50} $$

$$ u_s^i(\mathbf{x},0) = \int_V \mathrm{d}V(\mathbf{x}')C_{a_s^i a_s^i}(\mathbf{x},\mathbf{x}')^{\dagger}\dot{u}_s^i(\mathbf{x}',T), \dot{u}_s^i(\mathbf{x},0) = \int_V \mathrm{d}V(\mathbf{x}')C_{b_s^i b_s^i}(\mathbf{x},\mathbf{x}')^{\dagger} u_s^i(\mathbf{x}',T) \tag{51} $$

$$ \sum_j \hat{n}^j(\mathbf{x}) \left[ \int_{-\infty}^{+\infty} d\tau \sum_{kl} {}_{\gamma+1}\psi^{ijkl}(\mathbf{x},t-\tau)\partial_l u_s^k(\mathbf{x},\tau) \right] = \int_{\partial V} \mathrm{d}S(\mathbf{x}') \int_0^T \mathrm{d}t' C_{c_s^i c_s^i}(\mathbf{x},t;\mathbf{x}',t')^{\dagger} u_s^i(\mathbf{x}',T-t'). \tag{52} $$

These equations define the Euler–Lagrange equations for the weak-constraint problem. The forward wave equation is forced by terms representing model uncertainties and the adjoint wave equation is forced by misfit measurements that quantify the discrepancy between the observations and their corresponding model predictions. Extension to model residuals with non-zero biases is straightforward. Nonlinearity of the problem can be accounted for by solving Eqs. 43–52 iteratively for a number of times, thereby providing an effective means to assimilate seismic observations into dynamic ground-motion models.

### 2.4. Adjoint Method

A majority of previous work on full-wave seismic inversions (e.g. Tarantola 1984, 1988; Pratt *et al.* 1998; Tromp *et al.* 2005) solves a simplified version of the variational problem defined by Eq. 22. A common simplification is to assume the seismic wave equation and its associated initial and boundary conditions have zero model residuals. In this case, $C_{q_s^i q_s^i}$, $C_{a_s^i a_s^i}$, $C_{b_s^j b_s^j}$, and $C_{c_s^i c_s^i}$ the covariance functions, are set to zero, which corresponds to an infinite weight on the dynamic ground-motion model, i.e., the seismic wave equation and the associated initial and boundary conditions must be satisfied exactly. In the Euler–Lagrange Eqs. 43–52, this assumption eliminates the direct coupling between the forward model, Eqs. 50–52, and the adjoint wave-field ${}^{\dagger}\mathbf{u}_s(\mathbf{x}, t)$, although the seismic source and structure parameters, $\mathbf{f}_s(\mathbf{x}, t)$, $\mathbf{M}_s(\mathbf{x}, t)$, $\rho(\mathbf{x})$ and $\boldsymbol{\psi}(\mathbf{x}, t)$, still depend on the adjoint wave-field through Eqs. 46–49. The resulting formulation is named the "adjoint method" for solving the strong-constraint problem.

The adjoint method (i.e. Eqs. 43–52, with the right-hand-sides of Eqs. 50–52 set to zero) can also be derived using the Lagrange multiplier method (Liu

2006), which is commonly employed in optimal control theory to solve optimization problems with strong constraints imposed by partial differential equations (e.g. Biegler *et al*. 2003). In the Lagrange multiplier method, an objective function defined in terms of a weighted summation of the misfit measurements (i.e. the first term on the right-hand-side of Eq. 22) is minimized and the seismic wave equation, as well as its initial and boundary conditions, is included into the objective function using Lagrange multipliers. Variations with respect to the Lagrange multipliers return the dynamic model (i.e. Eqs. 50–52 with the right-hand-sides set to zero). Variations with respect to the seismic source and structure parameters return a set of equations similar to Eqs. 46–49 with the covariance functions set to the delta function. More detailed derivation of the adjoint method based on the Lagrange multiplier method can be found in Liu (2006).

The effects of neglecting errors in the dynamic model and/or the associated initial and boundary conditions have not yet been thoroughly investigated for full-wave seismic data assimilation problems. It has been found in data assimilation problems in meteorology and oceanography that one can obtain nonphysical values for model parameters that compensate for errors neglected in dynamic models and/or their conditions (e.g. Zupanski and Zupanski 2006 and references therein). It is clear that one will not be able to find the correct solution unless the approximations employed in the data assimilation procedure are valid. Deficiencies in dynamic models are usually difficult to eliminate, but the weak-constraint variational approach formulated in Eqs. 43–52 provides a systematic means to accommodate model residuals in full-wave seismic data assimilation processes.

### 2.5. *Scattering-Integral Method*

The equations for updating the seismic structure models, Eqs. 48, 49, involve a summation over the source index $s$. Considering the computational costs for numerically solving the three-dimensional forward and adjoint seismic wave equations in realistic geological media for each seismic source, $s$, for seismic data assimilation problems that involve a large number of distinct seismic sources, the generalized inverse as given in Eqs. 43–52, as well as the adjoint method for the strong-constraint problem, could become inefficient. An alternative derivation can be obtained as follows.

We introduce the Green's tensor for the adjoint system described by Eqs. 43–45, $G^{ij}(\mathbf{x},\ t-t';\mathbf{x}')$, where, following the convention in Aki and Richards (2002), the Green's tensor relates a unit impulsive force in direction $j$ acting at location $\mathbf{x}'$ to the displacement response in direction $i$ at location $\mathbf{x}$. The adjoint wave-field can then be expressed as

$$^{\dagger}u_s^i(\mathbf{x},t)=\int_V \mathrm{d}V(\mathbf{x}')\int \mathrm{d}t'\sum_j G^{ij}(\mathbf{x},t-t';\mathbf{x}')\,{}^{\dagger}f_s^i(\mathbf{x}',t'), \tag{53}$$

where the adjoint source field $^{\dagger}f_s^i(\mathbf{x}',t')$ is defined as

$$^{\dagger}f_s^j(\mathbf{x}',t')=-\sum_{rn}\delta(\mathbf{x}'-\mathbf{x}_r)J^i_{srn}(T-t')W_{d^i_{srn}d^i_{srn}}d^i_{srn}. \tag{54}$$

Evaluating the spatial integral on $\mathbf{x}'$, the adjoint wave-field can be expressed as

$$^{\dagger}u_s^i(\mathbf{x},t)=\sum_{r=1}^{N_r}\int \mathrm{d}t'\sum_j G^{ij}(\mathbf{x},t-t';\mathbf{x}_r)\,{}^{\dagger}f^j_{sr}(t'), \tag{55}$$

where $N_r$ is the total number of seismic receivers and the adjoint source field $^{\dagger}f^j_{sr}(t')$is defined as

$$^{\dagger}f^j_{sr}(t')=-\sum_n J^j_{srn}(T-t')W_{d^i_{srn}d^i_{srn}}d^i_{srn}. \tag{56}$$

The receiver-side Green's tensor (RGT) $G^{ij}(\mathbf{x},\ t-t';\ \mathbf{x}_r)$ in Eq. 55 does not depend upon the seismic source. It can be constructed by calculating the space–time volumes of the wave-fields generated by three orthogonal unit impulsive point forces acting at the receiver location $\mathbf{x}_r$ (Zhao *et al*. 2005). We can now replace the $N_s$ adjoint wave-propagation simulations for solving Eqs. 43–45 for all seismic sources with $3N_r$ wave-propagation simulations for constructing the RGTs for all seismic receivers.

An important benefit of explicitly constructing and storing the RGTs is that they allow us to obtain an approximate Hessian for the penalty function, which can be used to speed up convergence when

solving Eqs. 43–52 iteratively. Bringing Eq. 55 into Eqs. 48, 49, we obtain

$$
\begin{aligned}
{}_{\gamma+1}\rho(\mathbf{x}) &= {}_{\gamma}\rho(\mathbf{x}) \\
&+ \int_V \mathrm{d}V(\mathbf{x}')C_{\rho\rho}(\mathbf{x},\mathbf{x}')\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\sum_{rm}\left[\int \mathrm{d}t J^m_{srn}(t)\right. \\
&\left.\int_0^T d\tau \sum_i G^{mi}(\mathbf{x}_r, t-\tau;\mathbf{x}')\partial_t^2 u^i_s(\mathbf{x}',\tau)\right] W_{d^m_{srn}d^m_{srn}} d^m_{srn},
\end{aligned}
\tag{57}
$$

$$
\begin{aligned}
{}_{\gamma+1}\psi^{ijkl}(\mathbf{x},t) &= {}_{\gamma}\psi^{ijkl}(\mathbf{x},t) + \int_V \mathrm{d}V(\mathbf{x}')\int_0^T \\
&\times \mathrm{d}t'\{C_{\psi^{ijkl}\psi^{ijkl}}(\mathbf{x},t;\mathbf{x}',t')\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\sum_{nm} \\
&\times\left[\int \mathrm{d}t'' J^m_{srn}(t'')\int_0^T \mathrm{d}\tau\partial_j G^{mi}(\mathbf{x}_r,t''-t'-\tau;\mathbf{x}')\partial_l u^k_s(\mathbf{x}',\tau)\right] \\
&W_{d^m_{srn}d^m_{srn}} d^m_{srn}\},
\end{aligned}
\tag{58}
$$

where we have applied the reciprocity principle (Aki and Richards 2002)

$$
G^{ij}(\mathbf{x}',t;\mathbf{x}_r) = G^{ji}(\mathbf{x}_r,t;\mathbf{x}'). \tag{59}
$$

Considering the Born approximation (Dahlen and Tromp 1998), the terms inside the square brackets in Eqs. 57–58 are the Fréchet kernels of individual misfit measurements $d^m_{srn}$ with respect to structural parameters $\rho(\mathbf{x})$ and $\boldsymbol{\psi}(\mathbf{x}, t)$, and the temporal convolution over $\tau$ between the RGT and the forward wave-field is called the "scattering-integral" (Chen *et al*. 2007a). If we express the variational derivative of the misfit measurements with respect to structural parameters as

$$
\delta\mathbf{d} = \mathbf{A}\delta\mathbf{m}, \tag{60}
$$

where $\mathbf{d}$ is a vector composed of all misfit measurements, $\mathbf{m}$ is a vector of discretized $\rho(\mathbf{x})$ and $\boldsymbol{\psi}(\mathbf{x}, t)$, and $\mathbf{A}$ is a matrix with each row given by the Fréchet kernel of a misfit measurement, then the summations over $s$, $r$, $n$ and $m$ on the right-hand-sides of Eqs. 57–58 can be expressed as $\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathbf{dd}}\mathbf{d}$, an approximate Hessian of the penalty function with respect to $\mathbf{m}$ is given by $\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathbf{dd}}\mathbf{A} + \mathbf{W}_{\mathbf{mm}}$ and the Gauss–Newton normal equation is

$$
(\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathbf{dd}}\mathbf{A} + \mathbf{W}_{\mathbf{mm}})\delta\mathbf{m} = \mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathbf{dd}}\mathbf{d}, \tag{61}
$$

where $\mathbf{W}_{\mathbf{dd}}$ and $\mathbf{W}_{\mathbf{mm}}$ are the inverses of the covariance functions for $\mathbf{d}$ and $\mathbf{m}$, respectively. In practice, this normal equation does not need to be formed, because its solution can be computed by solving the linear system:

$$
\begin{bmatrix} \mathbf{W}^{1/2}_{\mathbf{dd}}\mathbf{A} \\ \mathbf{W}^{1/2}_{\mathbf{dd}} \end{bmatrix}\delta\mathbf{m} = \begin{bmatrix} \mathbf{W}^{1/2}_{\mathbf{dd}}\mathbf{d} \\ 0 \end{bmatrix} \tag{62}
$$

via a relaxation method such as LSQR (Paige and Saunders 1982). By solving Eq. 62 for $\delta\mathbf{m}$ in each iteration, it is possible to achieve quadratic convergence rate when minimizing the penalty function.

Another benefit for explicitly computing and archiving the RGTs is that they provide the Fréchet kernels for individual misfit measurements with respect to seismic source parameters $\mathbf{f}_s$ and $\mathbf{M}_s$, which also allows us to construct an approximate Hessian of the penalty function with respect to source parameters (Zhao *et al*. 2006). Bringing Eq. 55 into Eqs. 46, 47, we obtain

$$
\begin{aligned}
{}_{\gamma+1}f^i_s(\mathbf{x},t) &= {}_{\gamma}f^i_s(\mathbf{x},t) - \int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t'\left\{C_{f^i_s f^i_s}(\mathbf{x},t;\mathbf{x}',t')\right. \\
&\left.\sum_{rn}\sum_j\left[\int \mathrm{d}t'' J^j_{srn}(t'')G^{ij}(\mathrm{x}',t''-t';\mathbf{x}_r)\right]W_{d^j_{srn}d^j_{srn}} d^j_{srn}\},
\end{aligned}
\tag{63}
$$

$$
\begin{aligned}
{}_{\gamma+1}M^{ij}_s(\mathbf{x},t) &= {}_{\gamma}M^{ij}_s(\mathbf{x},t) + \int_V \mathrm{d}V(\mathbf{x}')\int_0^T \mathrm{d}t'\left\{C_{M^{ij}_s M^{ij}_s}(\mathbf{x},t;\mathbf{x}',t')\right. \\
&\left.\sum_{rn}\sum_m\left[\int \mathrm{d}t'' J^m_{srn}(t'')\partial_j G^{im}(\mathbf{x}',t''-t';\mathbf{x}_r)\right]W_{d^m_{srn}d^m_{srn}} d^m_{srn}\right\}
\end{aligned}
\tag{64}
$$

The terms in the square brackets in Eqs. 63, 64 provide the Fréchet kernels of individual misfit measurements with respect to the source parameters. The RGTs do not depend upon seismic source parameters; therefore, the same set of RGTs can be used for different seismic sources and no additional wave-propagation simulations are needed.

## 3. Data Functionals

The generalized inverse formulated in the previous section provides us with a unified methodology to solve two joint seismological inverse problems:

from observations of ground motions infer the structure of (a) the source that generates seismic waves, and (b) the geological medium through which seismic waves propagate. Although these two problems are coupled, seismologists usually attempt to separate variables through a judicious choice of ground-motion observations and appropriate averaging rules. For example, in certain classical methods, the travel-times of wave groups are used to constrain focal positions and seismic velocities, while their amplitudes (and polarities) are used to constrain the focal mechanisms and seismic attenuation parameters. These procedures work well when applied to first-motion data, but they must be generalized to capture the rich information now available from the complex waveforms recorded by modern broadband instruments.

### 3.1. Direct Waveform Inversion

Perhaps the most obvious data functional for waveform inversion is the difference between the observed and the synthetic seismograms at a set of sampling time $t_n$ (Nolet 1987; Tarantola 1988; Akcelik *et al*. 2002)

$$d^i_{srn} = \bar{u}^i_s(\mathbf{x}_r, t_n) - u^i_s(\mathbf{x}_r, t_n). \tag{65}$$

In this case, the seismogram perturbation kernel is just the Dirac delta function, $J^i_{srn}(t) = -\delta(t–t_n)$, and the Fréchet derivative of the misfit measurement with respect to structural parameters $\rho(\mathbf{x})$ and $\boldsymbol{\psi}(\mathbf{x}, t)$ is given directly by the Born approximation. However, the validity of the Born approximation $\bar{u}^i_s(\mathbf{x}_r, t_n) \approx u^i_s(\mathbf{x}_r, t_n) + \delta u^i_s(\mathbf{x}_r, t_n)$ is limited by the weak scattering condition (Wu and Aki 1988; Wu 2003). Using the data functional in Eq. 65 effectively involves the linearization of the Fourier shift operator, $\exp(i\omega t)$. For an observed waveform with centroid frequency $\omega_0$, the Born approximation is valid only when $\omega_0 \Delta T^i_{srn} \ll 1$, where $\Delta T^i_{srn}$ is the travel-time shift relative to the synthetic waveform.

An alternative data functional for waveform inversion is the difference between the frequency-domain observed and synthetic seismograms at a set of sampling frequencies $\omega_n$ (Pratt 1990; Pratt *et al*. 1998),

$$d^i_{srn} = \int \mathrm{d}t \, \exp(-i\omega_n t)\left[\bar{u}^i_s(\mathbf{x}_r, t) - u^i_s(\mathbf{x}_r, t)\right]. \tag{66}$$

In this case, the seismogram perturbation kernel is given by the Fourier transform kernel $J^i_{srn}(t) = -\exp(-i\omega_n t)$. Both the forward and the adjoint systems can be solved in the frequency domain, which allows us to avoid a linearization of the Fourier shift operator $\exp(i\omega t)$ effectively. But a drawback of this approach is that by using the Fourier transform, we lose resolution in the time domain and misfit information from different types of arrivals is mixed together, which may introduce nonlinear effects caused by the interference among different wave groups.

### 3.2. Time- and Frequency-Dependent Phase and Amplitude Misfit Measurements

One major difficulty in effectively utilizing waveform data in seismic tomography is that, unlike the travel-time data, the waveform data are much more nonlinear with respect to the seismic velocity model. The Fréchet derivatives provided by the generalized inverse or its variations are exact, but the quality with which the misfit measurement $d^i_{srn}$ can be approximated by a linear perturbation $\delta d^i_{srn}$, as well as the linearity of the seismological inverse problem, depends upon the type of the data functional.

In designing data functionals, we would like our measurement operator $D_n$ to have two properties. First, the measurement operator should allow us to selectively fit certain portions of the observed seismograms in time domain or frequency domain or time–frequency domain. This localization capability is important because it reduces possible nonlinear effects due to the interference among different wave groups and allows us to make incremental changes to our Earth model. We can start from fitting portions of the observed seismograms that are not too different from our synthetic seismograms and gradually improve our Earth model and try to fit more observed waveforms through iterations. Second, for those selected portions of the observed waveforms, the measurement operator should allow us to transform the raw waveform information, which is usually

recorded in the form of displacement (or its temporal derivatives) at a set of time samples, into other forms that are more linear with respect to the Earth structure model without loss of information or introducing bias.

The generalized seismological data functionals (GSDF) of Gee and Jordan (1992) quantify waveform differences using frequency-dependent phase-delay times and amplitude-reduction times measured on time-localized arrivals and have been successfully applied to tomographic inversions at different geographic scales (e.g. Gaherty *et al*. 1996; Katzman *et al*. 1998; Chen *et al*. 2007b) as well as to inversions for earthquake source parameters (e.g. McGuire *et al*. 2001; Chen *et al*. 2010a). In particular, the linearization of GSDF measurements depends on the Rytov approximation, which is valid for large phase-shifts as long as the phase perturbation per wavelength is small (Chernov 1960; Snieder and Lomax 1996). This is far less restrictive than the Born approximation, which requires small phase-shifts.

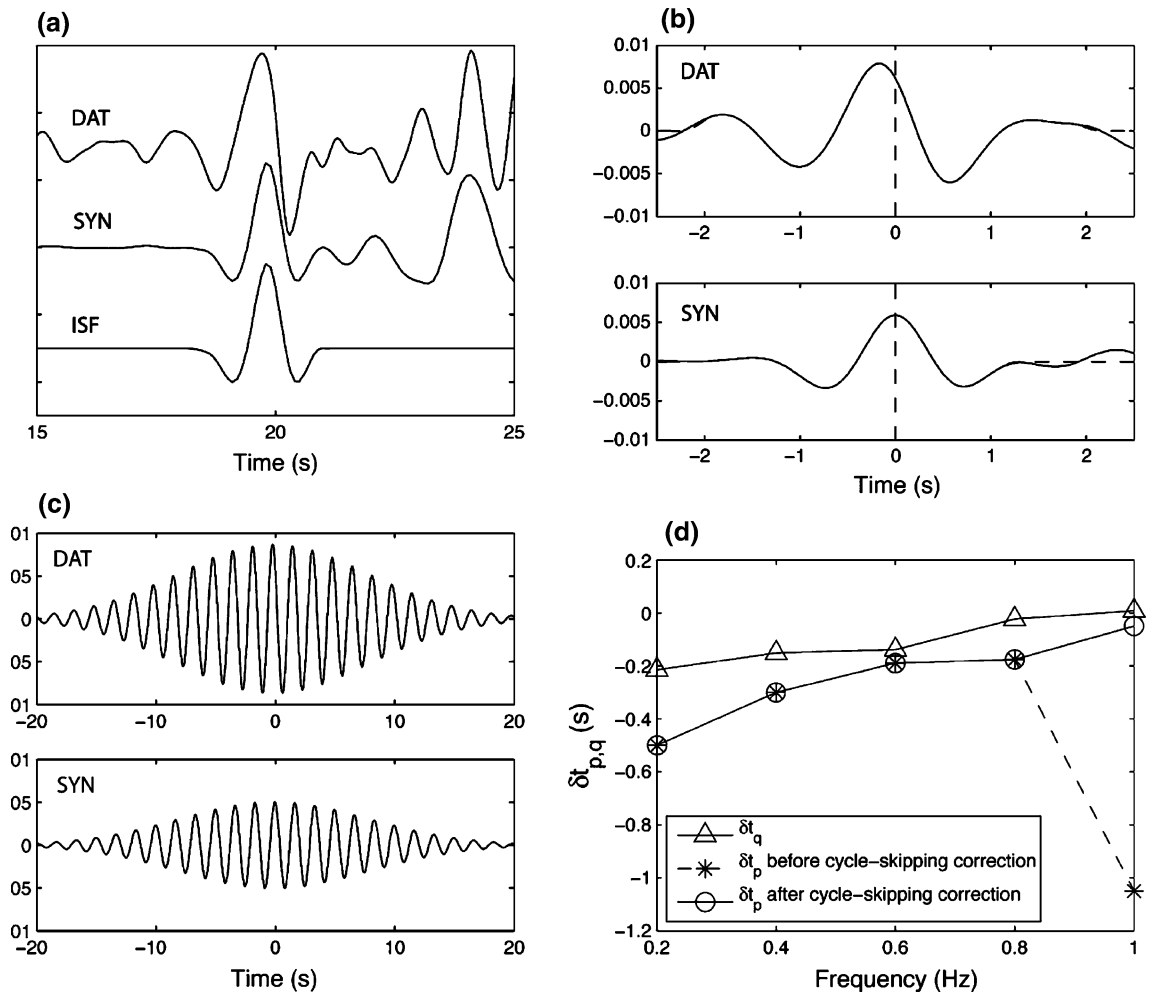


Figure 1

An example of GSDF processing. **a** The observed seismogram (DAT), synthetic seismogram (SYN) and the isolation filter (ISF) for the *S*-wave waveform we are analyzing. **b** Cross-correlogram between the isolation filter and the observed seismogram (*upper panel*) and cross-correlogram between the isolation filter and the complete synthetic seismogram (*lower panel*). Windowed cross-correlograms are shown as dashed lines. **c** Examples of narrowband-filtered windowed cross-correlograms for the data (*upper panel*) and the synthetic (*lower panel*). Centre frequency of the narrowband filter is 0.6 Hz; half-width is 0.1 Hz. **d** GSDF measurements made at five sampling frequencies. *Triangles*, amplitude-reduction times; *stars*, phase-delay times before correcting for cycle-skipping errors; *circles*, phase-delay times after correcting for cycle-skipping errors

The GSDF data processing consists of several steps (Fig. 1). We isolate the target wave group using an isolation filter $\tilde{u}_s^i(\mathbf{x}_r, t)$, which is obtained by windowing the complete synthetic seismogram,

$$\tilde{u}_s^i(\mathbf{x}_r, t) = W(t) u_s^i(\mathbf{x}_r, t). \tag{67}$$

In practice, we often apply a Tukey (i.e. cosine-tapered) window, which has a flat part in the middle of the window. The effect of applying such a time window is that the target waveform is distorted as little as possible by the windowing operation, while extraneous phases are excluded from the time window to reduce interference. We then cross-correlate the isolation filter with the complete synthetic seismogram and the observed seismogram and we window the resulting synthetic and data cross-correlograms around the zero-lag. The windowed correlograms are then narrowband-filtered at a set of frequencies $\omega_i$. When certain conditions about windowing and narrowband-filtering are enforced (Gee and Jordan 1992; Chen *et al*. 2010b), the resulting narrowband-filtered windowed correlograms can always be well matched by five-parameter Gaussian wavelets, which are cosine functions with frequencies at around $\omega_i$ and modulated by Gaussian envelopes. The differences in the phase and the amplitude between the synthetic and data Gaussian wavelets give us the phase-delay time $\delta t_{\mathrm{p}}$ and amplitude-reduction time $\delta t_{\mathrm{q}}$ at each narrowband-filtering frequency $\omega_i$. A practical issue of the GSDF analysis is that the phase-delay measurements need to be corrected for possible cycle-skipping errors before they can be used in inversions. These cycle-skipping errors can usually be corrected by bootstrapping the phase from low frequencies to high frequencies (Ekström *et al*. 1997).

In traditional broadband cross-correlation analysis, the travel-time shift $\Delta T$ of an isolated waveform is estimated using the location of the cross-correlogram peak (Luo and Schuster 1991; Woodward and Masters 1991; Dahlen *et al*. 2000; Zhao *et al*. 2000) and the amplitude anomaly can be determined from the maximum amplitudes of the cross-correlograms (Dahlen and Baig 2002; Ritsema *et al*. 2002). The seismogram perturbation kernels for broadband cross-correlation travel-time shift and amplitude anomaly are presented in Chen *et al*. (2007a). For band-limited signals, these measurements provide good estimates around the dominant frequency but they do not characterize the differences in the shape of the waveforms. In the GSDF analysis, we can account for differences in waveform shapes by making measurements at several frequencies across the frequency bandwidth. In Fig. (2), we illustrate this point using an example. By correcting the phase and amplitude of the synthetic waveform using the GSDF measurements made at five frequencies evenly distributed over the frequency band, we were able to recover the observed waveform almost perfectly.

Seismogram perturbation kernels $J^i_{srn}(t)$ for GSDF measurements were derived in Chen *et al*. (2010b). In this article I give a much-simplified derivation by assuming zero bandwidth for the narrowband-filtering operation in the GSDF analysis, which essentially results in a windowed Fourier analysis. The Fourier transform of the isolation filter can be expressed as

$$\hat{\tilde{u}}_s(\mathbf{x}_r, \omega) = \int \mathrm{d}t \, \exp(-i\omega t) W(t) u_s^i(\mathbf{x}_r, t). \tag{68}$$

We can also express the left-hand-side of Eq. 68 using two frequency-dependent, time-like quantities, $\tau_{\mathrm{p}}(\mathbf{x}_r, \omega)$ and $\tau_{\mathrm{q}}(\mathbf{x}_r, \omega)$,

$$\begin{aligned} &\exp\left[-i\omega\tau_{\mathrm{p}}(\mathbf{x}_r, \omega) - \omega\tau_{\mathrm{p}}(\mathbf{x}_r, \omega)\right] \\ &\quad = \int \mathrm{d}t \, \exp(-i\omega t) W(t) u_s^i(\mathbf{x}_r, t). \end{aligned} \tag{69}$$

By linearizing the left-hand-side of Eq. 69, we obtain the variational derivatives of $\tau_{\mathrm{p,q}}(\mathbf{x}_r, \omega)$ with respect to the synthetic seismogram $u_s^i(\mathbf{x}_r, t)$ as

$$\delta\tau_{\mathrm{p}}(\mathbf{x}_r, \omega) = \int_{-\infty}^{+\infty} \mathrm{d}t \, \mathrm{Im}\left[\frac{W(t)\mathrm{e}^{-\mathrm{i}\omega t}}{-\hat{\tilde{u}}_s^i(\mathbf{x}_r, \omega)\omega}\right] \delta u_s^i(\mathbf{x}_r, t), \tag{70}$$

$$\delta\tau_{\mathrm{q}}(\mathbf{x}_r, \omega) = \int_{-\infty}^{+\infty} \mathrm{d}t \, \mathrm{Re}\left[\frac{W(t)\mathrm{e}^{-\mathrm{i}\omega t}}{-\hat{\tilde{u}}_s^i(\mathbf{x}_r, \omega)\omega}\right] \delta u_s^i(\mathbf{x}_r, t). \tag{71}$$

The seismogram perturbation kernels for $\tau_{\mathrm{p,q}}(\mathbf{x}_r, \omega)$ can then be expressed as

$$J^i_{\mathrm{p}srn}(t) = \mathrm{Im}\left[\frac{W(t)\mathrm{e}^{-\mathrm{i}\omega t}}{-\hat{\tilde{u}}_s^i(\mathbf{x}_r, \omega)\omega}\right], \tag{72}$$

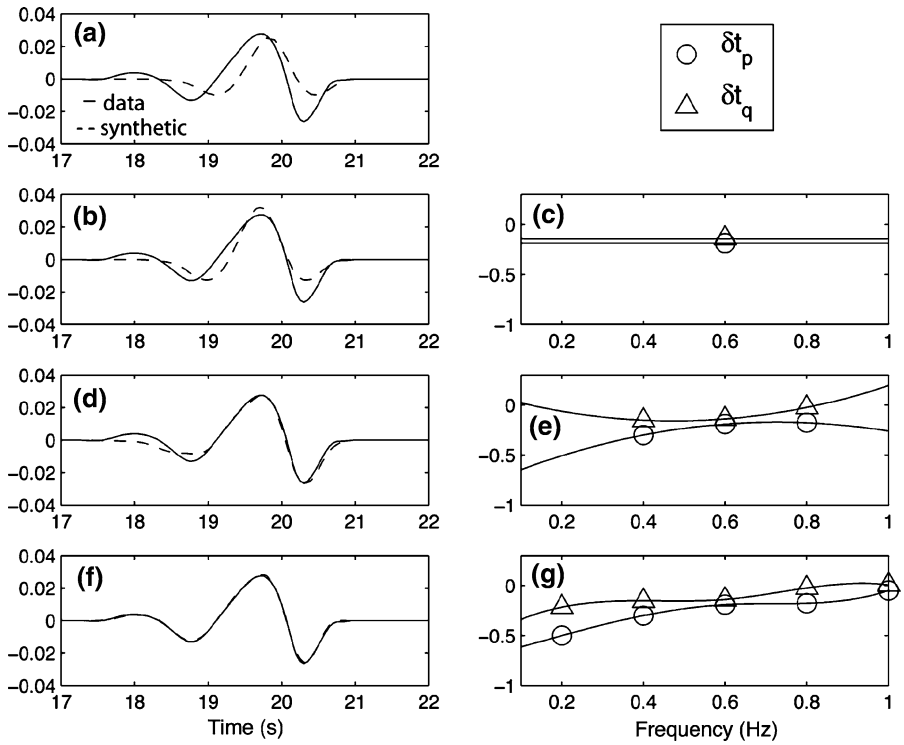


Figure 2

An example of fitting the synthetic waveform (*dash lines*) to the observed waveform (*solid lines*) by correcting the phase and amplitude of the synthetic waveform using GSDF measurements of phase-delay (*circles*) and amplitude-reduction times (*triangles*). The observed and synthetic waveforms are the same as in Fig. 1. **a** The original waveforms without perturbation. **b** and **c** correcting the synthetic waveform using GSDF measurements made at one sampling frequency 0.6 Hz. **d** and **e** correcting synthetic waveform using GSDF measurements made at three sampling frequencies, 0.4, 0.6, and 0.8 Hz. **f** and **g** correcting the synthetic waveform using GSDF measurements made at five sampling frequencies 0.2, 0.4, 0.6, 0.8, and 1.0 Hz. Cubic splines are used to interpolate and extrapolate phase and amplitude perturbations to all other frequencies. This example demonstrates that minimizing the frequency-dependent GSDF measurements is equivalent to fitting the waveforms

$$J^i_{\mathrm{q}srm}(t) = \mathrm{Re}\left[\frac{W(t)\mathrm{e}^{-\mathrm{i}\omega t}}{-\hat{\tilde{u}}^i_s(\mathbf{x}_r,\omega)\omega}\right]. \qquad (73)$$

The Fréchet kernels for $\tau_{\mathrm{p,q}}(\mathbf{x}_r, \omega)$ with respect to seismic velocity can be constructed using either the adjoint method or the scattering-integral method as described previously. Examples of the frequency-dependent Fréchet kernels for a direct-arriving *P*, the free surface reflected *pP* and *pS* waves for a source-receiver pair buried in the half-space model are shown in Fig. 3. The sensitivity for the *P*-wave is not concentrated on the ray path but extends to as far as 8 km away from the ray path. For the kernels of the frequency-dependent phase-delay times, we also see the "banana-doughnut" phenomena of vanishing sensitivity on the ray path (Marquering *et al*. 1999; Dahlen *et al*. 2000; Hung *et al*., 2000; Zhao *et al*., 2000; Zhao *et al*. 2005). The width of the first Fresnel zone and the spatial oscillation of the sensitivity depend on the frequency $\omega$. The sensitivity kernels for the surface-reflected phases *pP* and *pS* have similar characteristics to the direct-arriving *P*-wave. However, the width of the first Fresnel zone for these surface-reflected phases is larger than the direct-arriving *P*-wave due to longer propagation distances than the *P*-wave.

In addition to GSDF analysis, there are a number of different techniques for extracting time- and frequency-dependent phase and amplitude misfit measurements. Holschneider *et al*. (2005) developed a technique for extracting frequency-dependent phase-delay, group-delay and amplitude anomalies by modeling the deformation of the wavelet transform of the analyzed seismogram. The continuous

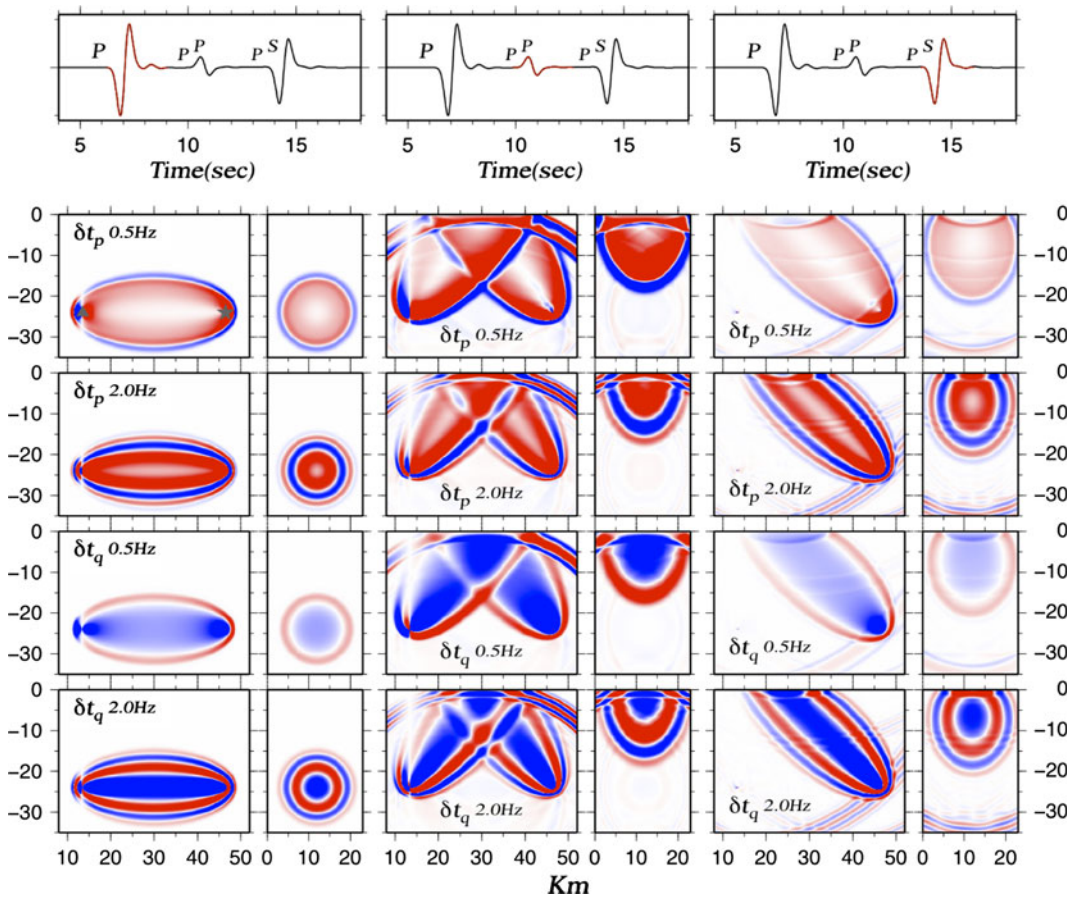


Figure 3

Fréchet kernels for the direct-arriving *P*-wave (*left column*), surface-reflected *pP*-wave (*middle column*) and *pS*-wave (*right column*) from an explosive source in a uniform half-space model. Plotted are the radial component of the synthetic seismogram, the isolation filters are highlighted in *red*, and the Fréchet kernels of GSDF $\delta t_{p,q}$ measurements at 0.5 and 2.0 Hz with respect to *P*-wave speed $\alpha$ in the source-receiver *vertical plane* and in the transverse plane midway (*dash line*) between the source and the receiver. The source (*grey star*) and the receiver (*grey triangle*) are buried at 24-km depth and have a distance of 32.2 km. In all the plots for the kernels, the color schemes are such that *white* represents zero, warm colors (*yellow* to *orange* to *red*) represent negative amplitudes indicating that a velocity increase leads to an advance in arrival time and an increase in amplitude and cool colors (*light* to *dark blue*) represent positive amplitudes indicating a velocity increase leads to a delay in arrival time and a reduction in amplitude

wavelet transform allows them to localize in time–frequency domain and selectively fit portions of the seismogram. Using the Gabor transform, Fichtner *et al*. (2008) extended GSDF analysis to continuous time–frequency and quantified waveform differences between observed and synthetic seismograms using an envelope misfit and a phase misfit that are continuous functions of time and frequency. The multi-taper technique (e.g. Thomason 1982) uses orthogonal windows to provide statistically independent estimates of spectral phases and amplitudes of a signal. This approach was later extended by Lilly and Park (1995) to provide a formal uncertainty estimates for the extracted spectral parameters.

In general, it has been recognized both in theoretical studies and in practical applications that the appropriate choice of data functionals plays an essential role in overcoming the nonlinearity of waveform inversions. In particular, we need to separate phase information, which is quasi-linearly related to structural parameters, from the amplitude information, which can be highly nonlinear with respect to seismic velocity. Separation of different wave groups on the seismogram through localization

in time domain and/or frequency domain allows us to reduce nonlinearity caused by interference and provides a pathway to bootstrap our Earth model through iteratively conditioning the data used in the inversion.

## 4. Applications

The advantage of the full-wave seismic data assimilation methodology over classical methods used in solving seismological inverse problems is two-folded. First, the data functionals defined in Eq. 5 and explained in more detail in previous sections allows us to treat different types of misfit between observation and model-prediction (e.g. travel-time, amplitude, waveform) in a unified and self-consistent framework based on the wave equation. Misfit measures commonly used in classical methods, such as travel-times of body waves, dispersion curves of surface waves, and frequencies of the Earth's free oscillation, usually have quite different back-propagation mechanisms based on approximations appropriate for different types of phases on the seismogram. In the full-wave methodology, they can all be back-propagated to provide modifications to the seismic source and structure models through the use of the wave equation and the seismogram perturbation kernel. Explicit expressions of the seismogram perturbation kernels for several classical types of misfit measures were presented Chen *et al*. (2007a). Moreover, this general wave-equation-based framework opens up the possibility of using the complete seismic waveform data, which contain much more information about the propagation medium and the excitation source than classical data functionals such as travel-times or amplitudes of certain waveforms on the seismogram. Such full waveform information, which can be decoded using the full-wave methodology, might provide us with a seismic source and structure models with much higher resolution than classical methods.

Second, in the full-wave methodology, either the adjoint method or the scattering-integral implementation, the mechanism for back-propagating misfit information to modifications of source and structure models is exact, therefore eliminating possible errors caused by approximate back-propagation mechanisms commonly found in classical methods and improving the accuracy of the obtained source and structure models. A widely used classical back-propagation method is based on the ray theory for the wave equation (e.g. Beylkin 1985). In such ray-theoretic framework, the earthquake wave-fields and the adjoint wave-fields in the adjoint method or the Green tensors in the scattering-integral implementation are usually approximated using solutions of the eikonal equation, which provides an approximation to the wave equation and is valid at high frequencies (i.e. the Fresnel zone size is much smaller than the scale of the heterogeneities). In the full-wave methodology, the Fréchet kernels of the objective function and the Fréchet kernels of the data functionals are all computed using purely numerical solutions to the seismic wave equation, thereby eliminating possible errors caused by kernel inaccuracy.

### 4.1. Enhancing the Resolution of Earth Structure Models

The capability of the full-wave methodology in enhancing the resolution and accuracy of the obtained Earth structure models was demonstrated quite convincingly in a "blind test" for the two-dimensional (2D) acoustic case in Brenders and Pratt (2006). In the "blind test", pseudo-data, which were synthetic seismograms computed by solving the acoustic wave equation in a complex true structure model of a 250 km-long, 40 km-deep, 2D section of the lithosphere, were provided to the authors, but the true model itself was not revealed until after the presentation of the full-wave inversion result. Compared with the result obtained using the classical ray-theoretic method, the full-wave model provided a much better correspondence with the true model, from large- to intermediate-wavelength scales. The source-receiver coverage was exactly the same for both the full-wave and the ray-theoretic inversions. In the full-wave inversion, frequency-domain waveform misfits similar to Eq. 66 were back-propagated using exact kernels computed from the acoustic wave equation, while in the ray-theoretic inversion, travel-times of the first-arriving *P*-waves were back-propagated using ray-theoretic, approximate kernels. The lenticular low-velocity zone near the surface, which is almost invisible in the ray-theoretic result, was correctly

recovered in the full-wave model. The full-wave inversion result provided a high-resolution image of the complex structure of the entire crust as well as the well-defined crust-mantle transition zone. Some inaccuracies were observed near the edges of the full-wave model due to the limited source-receiver coverage inherent to the geometry of the seismic survey. Within the central portion of the model, the full-wave method has a resolution of approximately 1 km, which was not attainable by the ray-theoretical method. The resolution of ray-theoretic inversions using travel-time data is generally limited to the size of the first Fresnel zone (Williamson 1991; Williamson and Worthington 1993; Schuster 1996; Dahlen 2004), while full-wave inversions can potentially provide sub-wavelength resolution with good source-receiver coverage (e.g. Pratt and Shipp 1999).

The first successful full-3D, full-wave seismic tomography using real waveform data from regional earthquakes was presented in Chen *et al*. (2007b). In that study, waveform misfits quantified using the frequency-dependent GSDF measurements were inverted to improve the 3D elastic starting model, the Southern California Earthquake Center (SCEC) Community Velocity Model version 3.0 (CVM3.0) (Fig. 4), in a 142 km-long, 84 km-wide and 26 km-deep volume around the Los Angeles Basin region using the scattering-integral implementation. The revised 3D model, LAF3D (Fig. 4), provides substantially better fit to the observed waveform data than the

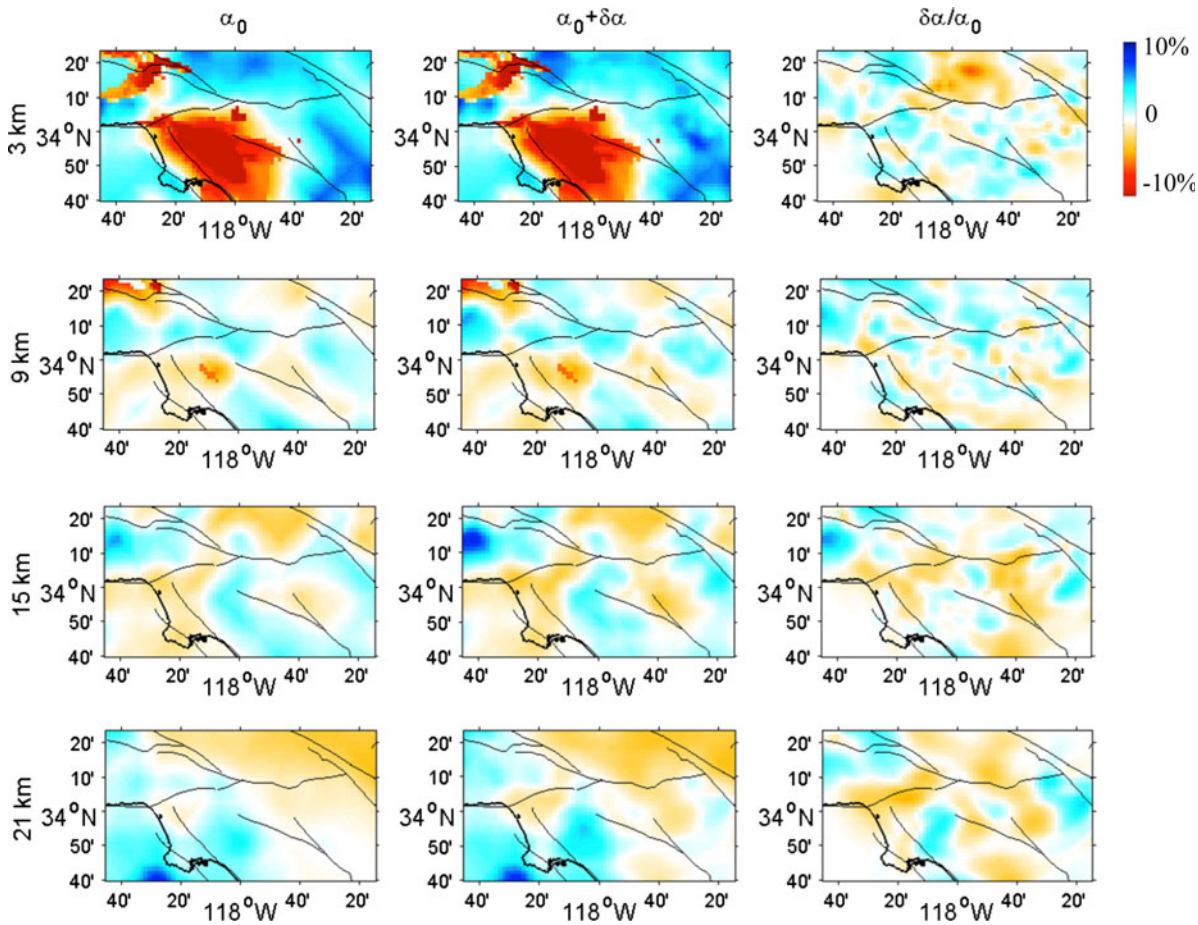


Figure 4
Map views of the *P*-velocity model for the starting model CVM3.0 (*left column*), the improved model LAF3D (*middle column*), and the relative perturbation of LAF3D relative to CVM3.0 (*right column*). For the first and second columns, color indicates the variation with respect to the average velocity at each depth; for the third column, color indicates relative perturbation between LAF3D and CVM3.0

3D starting model for frequencies up to 1.2 Hz. The perturbations in *P*- and *S*-wave speeds with respect to the starting model are up to 10%. In the interior of the Los Angeles Basin, wave speeds are slightly higher in LAF3D than in CVM3.0, while at the edge of the basin wave speeds are lower in LAF3D than in CVM3.0. This finding is consistent with the Harvard model (Süss and Shaw 2003), another 3D seismic velocity model in this area, whose *P*-velocity is slightly faster in the center of the basin than CVM3.0 and slightly slower than CVM3.0 on the border of the basin. The inverted model perturbation seems to enhance the velocity gradient that already exists in the starting model. In particular, the overall north–south gradient (from low to high) in both *P*- and *S*-velocities in the middle to lower crust is stronger in LAF3D than in CVM3.0. And the inverted *P*-velocity perturbation seems to enhance the west–east gradient in the lower crust. The inverted model perturbation has good spatial correlation with the distribution of major faults in the area. For example there is a negative perturbation of about 1–2% in *P* velocity from about 3 km to about 9 km depth along the San Andreas fault (Fig. 4, upper-right corner), suggesting that the *P*-velocity in CVM3.0 is too fast in this segment of the San Andreas fault. Another example is that the Sierra Madre fault zone and the Santa Monica fault seem to coincide with the locations of some negative perturbations in both *P*- and *S*-velocities.

Recently, Tape *et al*. (2009, 2010) have adapted the adjoint method to image the crustal structure in southern California using waveform data from local earthquakes. Frequency-dependent phase and amplitude misfit measurements extracted using the multitaper technique were used to quantify waveform misfits. After 16 iterations of the adjoint method, strong heterogeneities with local perturbations up to 30% with respect to the 3D starting model were recovered. High-resolution features such as seismic velocity contrasts across faults due to compositional changes were recovered with high fidelity. The adjoint method has also been adapted for continental-scale waveform tomography in Fichtner *et al*. (2009). The continuous time–frequency phase and envelope misfits (Fichtner *et al*. 2008) were used to quantify differences between observed and synthetic waveforms. The resulting seismic velocity models reveal structural features of Australasian upper-mantle with great detail. The same full-wave technique has been extended to infer radially anisotropic structure under Australasian in Fichtner *et al*. (2010). Detailed structural features with locally 2° lateral resolution were revealed. Strong radial anisotropy showing a clear ocean-continent dichotomy was recovered for depths above 150 km. For depths below 250 km, significant anisotropy with short wavelength (<500 km) was detected, indicating small-scale sublithospheric convection and/or a change in the dominant glide system of olivine.

### 4.2. *Improving Earthquake Source Parameter Estimation*

Accurate and rapid estimation of earthquake source parameters is important both for seismic hazard analysis and for realistic interpretation of geological and tectonic structures. The scattering-integral method provides a very efficient means for assimilating seismic waveform observations into better estimations of earthquake source parameters using realistic 3D Earth structure models. Once the RGTs for all the receivers are computed and stored on disk, seismic source parameters can be estimated very rapidly by retrieving a small volume around the reference source location from the RGT database and no additional wave-propagation simulations are needed. Moreover, the capability to use realistic 3D structure models in computing the RGTs not only allows us to make more accurate estimates of classical source parameters (Zhao *et al*. 2006) but also opens up the possibility of estimating new types of source parameters that could lead to better understanding of regional tectonics and earthquake source physics and provide new constraints for seismic hazard analysis.

The lowest-order representation of a finite rupture is the finite moment tensor (FMT) (Chen *et al*. 2005), which contains the second-order polynomial moments of $\mathbf{M}(\mathbf{x}, t)$ in addition to the zeroth- and first-order polynomial moments included in the centroid moment tensor (CMT) representation. The FMT resolves fault-plane ambiguity of the CMT and provides several additional parameters of seismological interest, including the characteristic length, width, and

duration of the faulting, as well as the directivity vector of the fault slip. FMT parameters have been successfully recovered for global large earthquakes ($M_w > 6$) using Green tensors computed in a 1D Earth structure model (McGuire *et al*. 2001). For regional small to medium-sized earthquakes ($2.5 < M_L < 5$), Green tensors computed in a 1D structure model are usually not accurate enough to warrant full recovery of FMT parameters due to stronger small-scale 3D heterogeneities in the crust. Some recent experiments in southern California have shown that by adopting more realistic 3D structure models, it is feasible to recover FMT parameters for small- to medium-sized earthquakes (Chen *et al*. 2005; Chen *et al*. 2010a). Figure 5 shows a 3D rendering of the source mechanisms for some small earthquakes around the Los Angeles Basin area. The two nodal planes of the CMT solutions are now represented as two intersecting disks, whose sizes are proportional to the likelihood of being the actual fault plane as determined from the FMT parameters (Chen *et al*. 2010a). The procedure does not rely on detecting directivity effects; therefore, it is applicable to any type of earthquake. For some earthquakes, the fault planes selected by this automated FMT inversion procedure were confirmed by the distributions of relocated aftershock hypocenters (Chen *et al*. 2010a). In regions where there are no precisely relocated aftershocks or for earthquakes with few aftershocks, this method provides the most convenient means for resolving fault plane ambiguity.

The success of those experiments suggest that improved versions of more accurate 3D structure models will eventually allow the automated recovery of FMTs for most small- to medium-sized earthquakes in well-instrumented regions. By recovering FMT parameters for regional small- to medium-sized earthquakes in a routine manner, we can provide clearer pictures on regional tectonics, reduce the uncertainties in inversions for crustal stress fields based on earthquake source mechanism data (Gephart 1985; Michael 1987; Gephart 1990; Yin 1996) and calculate Coulomb stress changes and provide a more robust description of probabilistic seismic hazard based on the stress transfer model (Harris 1998; McCloskey *et al*. 2003; Steacy *et al*. 2005; Parsons 2005).

The importance of using realistic 3D structure models in finite-source inversions is also highlighted in a recent study of the 1996 Bárdarbunga, Iceland, earthquake, which was caused by volcanic caldera subsidence (Fichtner and Tkalcic 2010). A 3D structure model was constructed by interpolating vertical profiles from receiver function studies and surface wave dispersion data and the spectral-element method was used to solve the elastic wave equation. The accuracy of their 3D structure model and the numerical method for wave propagation allows the authors to model observed waveforms for frequencies up to 0.2 Hz and obtain robust estimates of the caldera geometry and the kinematics of rupture initiation and propagation.

## 5. Discussion

The implementation of full-wave seismic data assimilation methodology is far from trivial. The computational efficiency of different types of implementations depends upon the overall problem geometry, particularly on the ratio of sources to receivers, as well as trade-offs in computational resources, such as the relative costs of compute cycles to data storage. The waveform data can potentially provide much more detailed information about the seismic source and the Earth structure, but to effectively utilize waveform information we need to overcome the nonlinearity associated with waveform inversions.

### 5.1. Computational Challenges

The computational challenges of full-wave seismic data assimilation largely lie in (1) the solution of the 3D seismic wave equations (forward or adjoint) for computing the synthetic seismograms and the Fréchet kernels and (2) the optimization algorithms used for finding the optimal structure and source models that minimize the objective function. Both the scattering-integral implementation and the adjoint method require a large number of simulations for problems of practical interest. To make full-wave seismic data assimilation feasible on a practical time scale, the efficiency of the underlying seismic wave equation solver becomes paramount, which poses significant computational challenges in terms of both

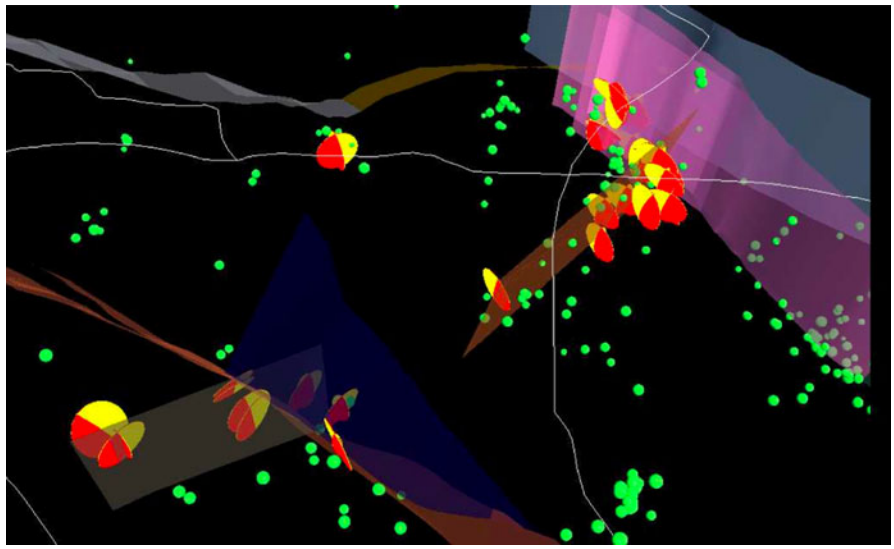

Figure 5
Projection of a 3D rendering of the focal mechanisms for earthquakes in the Yorba Linda cluster and the Fontana trend. Focal mechanisms are represented as two intersecting disks, whose sizes are proportional to the likelihood of being the actual fault plane as determined in this study. Extensive quadrants are painted in *yellow* and compressive quadrants are painted in *red*. Major faults in this area are represented as colored transparent surfaces. The surfaces denoted as 'Fontana Seismicity Best fit' (FSB) and 'Yorba Linda Extruded' (YLE) are linear least-square fit to the hypocenters of the earthquakes in the Fontana trend and the Yorba Linda cluster, respectively. Locations and orientations of the faults as well as the FSB and YLE surfaces are obtained from the SCEC Community Fault Model (CFM-A). Background seismicity is shown as green spheres. Major highways in this area are plotted as white solid lines and the interstate freeways I-10 and I-15 are labelled in *red*. The 3D rendering was done using the Java3D-based interactive 4D visualization software SCEC-VDO (http://scecdata.usc.edu/scecinterns/index.php?title=UseIT:_SCEC-VDO_User%27s_Page) developed by the SCEC Undergraduate Studies in Earthquake Information Technology (UseIT) interns under the supervision of Sue Perry (perry@usc.edu)

the design of the numerical algorithms and their implementations on homogeneous/heterogeneous multi-core hardware.

In the adjoint method, both the forward wave-field $\mathbf{u}_s$ and the adjoint wave-field $^{\dagger}\mathbf{u}_s$ are source-specific, therefore $2N_s$ simulations ($N_s$ forward and $N_s$ adjoint simulations) are needed for one calculation of the gradients of the objective function. The time integration in Eqs. 49, 50 is independent of the spatial location $\mathbf{x}$, and the total number of operations is proportional to $N_V N_s$, where $N_V$ is the total number of spatial grid points. If the conjugate-gradient algorithm (Press *et al*. 1992) is adopted for minimizing the objective function, each iteration requires one evaluation of the gradient to determine the conjugate-direction in the model space, followed by a line search to determine the optimal step length to take in the conjugate direction. There are a variety of algorithms for line search (Box and Swann 1969) and its implementation certainly affects the number of simulations needed. A cubic interpolation scheme for the line search (Tromp *et al*. 2005) requires an additional $N_s$ forward and $N_s$ adjoint simulations and the total number of time-integration operations is proportional to $2N_V N_s$. During the time-integration operations, we need simultaneous access to both the forward and the adjoint wave-fields and one of these two wave-fields needs to be time-reversed. If we choose to store the forward wave-field or the adjoint wave-field on disk and read it back during the time integration, the total number of simulations is $4N_s$ and the disk storage requirement is the space–time volume of one wave-field, which is proportional to $N_V N_T$, where $N_T$ is the total number of time steps for the simulation. To avoid storing the complete space–time wave-fields and the associated I/O cost, for non-dissipative wave equations, another approach is to reconstruct the time-reversed forward wave-field or adjoint wave-field on the fly. In this case, we need to store only the last time-frame of the forward or the adjoint wave-field as a function of space, which has a volume proportional to $N_V$, and additional $2N_s$ simulations are needed to

recover the time-reversed forward or adjoint wave-field by solving the forward or adjoint wave-equation backwards in time, taking the final state of the wave-field as its initial condition (GAUTHIER *et al*. 1986). The total number of simulations is then $6N_s$.

In the scattering-integral implementation, the calculation of the Fréchet kernels for individual data functionals requires simultaneous access to both the forward wave-field from the source and the Green tensor from the receiver. We can compute the RGTs once for each receiver, store it on disk (at cost $\sim N_V N_T$), and use it for every source in the dataset, which can be efficient when the number of sources is larger than the number of receivers. A complete calculation of the data-specific kernels requires $3N_r$ simulations (considering three orthogonal unit forces at each receiver) and one forward simulation for each source—a total $3N_r + N_s$, plus storage cost proportional to $3N_rN_VN_T$. The number of time integration operation is determined by the convolution between the forward wave-field from the source and the RGT and is proportional to $2N_t$, where $N_t$ is the number of time steps within the support of the seismogram perturbation kernel, the number of spatial grid points $N_V$ and the total number of seismograms $N_u$ used in the inversion. For three-component receivers, $N_u \leq 3N_rN_s$. Owing to noise and instrumental reasons for eliminating seismograms, $N_u$ is usually much less than this bound. In the scattering-integral method, the quadratic objective function is minimized using the Gauss–Newton algorithm, which requires additional computing time (though not too much compared with the wave-propagation simulations) to solve the linear system in Eq. 62.

In recent years, seismologists have developed parallel computer codes using a variety of numerical algorithms to solve the seismic wave equation. At present, most of the parallel earthquake simulation codes cannot exploit multiple cores beyond running additional MPI processes per node. A general impression is that once the message-passing library is made "SMP-aware" (i.e. message-passing between processors on the same node can bypass the network, therefore reducing network performance bottlenecks), the single-level, flat message-passing-only programming model might continue to be the best overall approach to a scalable design. However, despite the success of the single-level, message-passing-only approach, there are growing evidences that future earthquake simulation software will have to address increasing node-level complexities by introducing another level of parallelism. Message-passing might continue to be the dominant inter-node programming model in future high-performance scientific applications, but it might not be sufficient for driving all the cores on a single node for much longer, at least at the single-level, flat message-passing model. Seismologists and computer scientists are now exploring new node-level programming models that can effectively combine with the inter-node message-passing model and incorporating this new programming model into next-generation seismic wave equation solvers.

### 5.2. *Nonlinearity*

The nonlinearity of the structural inverse problem, which includes multiple scattering (higher-order terms in the Born series), is accounted for through iteration, in which the reference structure model is updated at the beginning of each iteration and the Green's functions (or the adjoint wave-fields in the adjoint method) are re-computed using the updated structure model. This approach of accounting for nonlinearity is different from inverse scattering based on reconstructing higher-order terms in the Born or Rytov series, in which the higher-order terms are computed using the same Green's function as the first-order term. The higher-order terms in the Born/Rytov series might be difficult to compute except for simple structure models for which analytical solutions of the Green's functions are available.

The convergence of the optimization algorithm depends upon the shape (number of local minima) of the objective function used to quantify misfit between observation and model prediction (TARANTOLA 2005). Objective functions defined in terms of the energy of differential waveforms (TARANTOLA 1984, 1988) are dominated by local minima because the waveform itself is a very nonlinear function of the velocity model. In this case, conjugate-gradient or other types of gradient- and/or Hessian-based algorithms will have difficulty converging unless the starting model is already very close to the global minimum. The gradients and/or Hessian, even though they are

computed exactly, are not very useful if the starting model is far from global minimum.

There are a number of different strategies for dealing with this problem in the full-wave domain, for instance, using a wave-equation travel-time tomography (Luo and Schuster 1991) to align the phases before doing a waveform inversion. One strategy that has been developed substantially is based on a scale-decomposition approach or "frequency bootstrapping" (i.e. starting from low-frequency waveform data to improve large-scale structure and move to higher frequencies while the model improves, please see Bunks *et al*. 1995; Akcelik *et al*. 2002; Sirgue and Pratt 2004), which can allow a descent algorithm to stay in the neighborhood of the global minimum or the "basin of attraction". This type of approach may fail in practice if the waveform data do not have sufficient signal-to-noise ratio at low frequencies that would allow us to bootstrap from a starting model far away from the global minimum. In this case, we have a scale gap in our approach.

Another strategy is to re-parameterize waveform information into another form that is more linear with respect to structure model. Experiences from decades of travel-time tomography show that objective functions defined in terms of differential travel-times are much smoother than those defined in terms of differential waveforms; therefore, more suitable for gradient-descent type optimization algorithms. To better capture waveform information, Gee and Jordan (1992), Holschneider *et al*. (2005), among others, proposed to use frequency-dependent phase-delay and amplitude anomalies (GSDF) to re-parameterize waveform misfit. The exact Fréchet derivatives of these misfit measurements (gradient and/or Hessian of the objective function) can be computed using the "seismogram perturbation kernel", which is the exact Fréchet derivative of the misfit measurement with respect to the waveform, and the Born approximation, which provides the Fréchet derivative of the waveform with respect to structural parameters, by applying the chain rule (Chen *et al*. 2007a).

In Chen *et al*. (2007a), we pointed out the connection between the exact Fréchet derivatives of these frequency-dependent phase and amplitude misfits (GSDF measurements) and the Rytov approximation, which can handle large accumulative phase-shifts much better than the Born approximation. Unlike the Born approximation, the Rytov approximation is not limited to the "small phase-shift" constraint (weak scattering); instead, it is limited by the "small-angle" (forward-scattering) constraint. The Rytov approximation has been widely used in long-distance propagations with primarily forward-scattering or small-angle scattering such as line-of-sight optical or radio wave propagation (Chernov 1960; Tatarskii 1971; Ishimaru 1978) and diffraction tomography (Devaney 1984; Wu and Toksöz 1987).

In Chen *et al*. (2007b) we applied our full-3D, full-wave methodology to the Los Angeles Basin region. The misfit measure we used was frequency-dependent phase-delay on direct-arriving *P*- and *S*-waves (forward-scattered or transmitted waves). In one Gauss–Newton iteration, we were able to correct more than 80% of the large phase-shifts, which can be larger than one-quarter of a cycle, between observed and synthetic waveforms (Fig. 18, 19 in Chen *et al*. 2007b). This suggests that the objective function defined in terms of the frequency-dependent phase-delay measurements for forward-scattered waves is smoother (has a larger basin of attraction) with respect to seismic velocities than an objective function defined in terms of differential waveforms.

In Chen *et al*. (2007b), we obtained nearly 10% perturbation to the reference model in one Gauss–Newton iteration. In Tape *et al*. (2009, 2010), the authors obtained nearly 30% perturbation after 16 conjugate-gradient iterations. Their penalty function is also defined in terms of frequency-dependent phase-delay measurements. These large perturbations obtained in these full-3D, full-wave applications suggest that if the objective function is relatively smooth (have a large basin of attraction), descent-based optimization algorithms may not necessarily require the reference model to be very close to the global minimum.

## 6. Summary

In this article, I have derived the generalized inverse formulation for the full-wave seismic data assimilation problem to iteratively improve the

estimates of both seismic source parameters and the Earth structure models. The starting point was the Bayes' theorem where the observations, the dynamical model and the poorly known seismic source and structure parameters were introduced with an assumption of Gaussian priors. This led to an objective function in quadratic form. By applying the Hamilton's principle, we derived the Euler–Lagrange equations, which allowed us to obtain a solution using gradient-descent algorithms. The adjoint method and the scattering-integral method, which have been successfully applied in full-3D, full-wave seismic tomography and source parameter inversions, can be considered as different implementations of the same physical principles based on the generalized inverse. The design of appropriate data functionals is an important issue. The frequency-dependent phase and amplitude misfit measurements made on localized waveforms can not only quantify waveform misfits but also reduce the nonlinearity associated with the seismic inverse problem. Some recent applications of full-wave seismic tomography at different geographic scales and earthquake source parameter inversions were reviewed and some challenging issues related to the computational implementations and the nonlinearity of waveform inversions were discussed.

## References


Akcelik, V., Biros, G. and Ghattas, O., 2002. *Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation*. Proceedings of the 2002 ACM/IEEE conference on Supercomputing.

Akcelik, V., Bielak, J., Biros, G., Epanomeritakis, I., Fernandez, A., Ghattas, O., Kim, E. J., Lopez, J., O'Hallaron, D., Tu, T., Urbanic, J., 2003. *High-resolution forward and inverse earthquake modeling on terascale computers*. Proceedings of the 2003 ACM/IEEE conference on Supercomputing.

Aki, K. and Richards, P. G., 2002. Quantitative seismology. *University Science Books*, Sausalito, California, USA, 700 pp.

Backus, G. E., 1962. *Long-wave elastic anisotropy produced by horizontal layering*. Journal of Geophysical Research, *67*(11), 4427–4440.

Bamberger, A., Chavent, G., Lailly, P., 1977. *Une application de la théorie du contrôle à un problème inverse sismique*. Annales Geophysicae, *33*, 183–200.

Bamberger, A., Chavent, G., Hemons, Ch., Lailly, P., 1982. *Inversion of normal incidence seismograms*. Geophysics, *47*(5), 757–770.

Bao, H., Bielak, J., Ghattas, O., Kallivokas, L., O'Hallaron, D. R., Shewchuk, J. R. and Xu, J., 1998. *Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers*. Computer Methods in Applied Mechanics and Engineering, *152*, 85–102.

Bennett. A.F., 1992. *Inverse Methods in Physical Oceanography*. Cambridge Monographs on Mechanics and Applied Mathematics, *Cambridge University Press*, New York, USA, 368 pp.

Bengtsson, L., Ghil, M. and Källén, E., 1981. *Dynamic meteorology: data assimilation methods. Springer*, New York, USA, 330 pp.

Beylkin, G., 1985. *Imaging of discontinuities in the inversion scattering problem by inversion of a causal generalized Radon transform*. Journal of Mathematical Physics, *26*(1), 99–108.

Biegler, L. T., Ghattas, O., Heinkenschloss, M. and Van BloemenWaanders, B., 2003. *Large-scale PDE-constrained optimization*. *Springer*, New York, USA, 349 pp.

Box, M. J., Davies D. and Swann, W. H., 1969. Non-Linear Optimization Techniques, *Oliver & Boyd*, Edinburgh, UK, 60 pp.

Brenders, A., and Pratt, R. G., 2006. *Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model*. Geophysical Journal International, *168*, 133–151.

Bunks, C., Saleck, F. M., Zaleski, S. and Chavent, G., 1995. *Multiscale seismic waveform inversion*. Geophysics, Vol. 60, No. 5, p. 1457–1473.

Chen, P., Jordan, T. H., and Zhao, L., 2005. *Finite-moment tensor of the 3 September 2002 Yorba Linda earthquake*. Bulletin of the Seismological Society of America, *95* (3), 1170–1180.

Chen, P., Jordan, T. H., and Zhao, L., 2007a. *Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods*. Geophysical Journal International, *170*(1), 175–181.

Chen, P., Zhao, L., and Jordan, T. H., 2007b. *Full 3D tomography for the crustal structure of the Los Angeles region*. Bulletin of the Seismological Society of America, *97* (4), 1094–1120.

Chen, P., Jordan, T. H. and Zhao, L, 2010a. *Resolving fault plane ambiguity for small earthquakes*, Geophysical Journal International, *181*, 493–501, doi:10.1111/j.1365-246X.2010.04515.x

Chen, P., Jordan, T. H. and Lee, E., 2010b. *Perturbation kernels for Generalized Seismological Data Functionals (GSDF)*. Geophysical Journal International, *183*, 869–883, doi:10.1111/j.1365-246X.2010.04758.x

Chernov, L.A., 1960. *Wave Propagation in a Random Medium*. *McGrawHill*, New York, USA, 168 pp.

Dahlen, F. A., 2004. *Resolution limit of traveltime tomography*. Geophysical Journal International. *157*, 315–331.

Dahlen, F. A. and Tromp, J., 1998. *Theoretical global seismology*. *Princeton University Press*, Princeton, New Jersey. 1025 pp.

Dahlen, F. A., Hung, S., and Nolet, G. 2000. *Fréchet kernels for finite-frequency travel times-.Theory.*. Geophysical Journal International, *141*, 157–174.

Dahlen, F. A. and Baig, A. M., 2002. *Fréchet kernels for body wave amplitudes*. Geophysical Journal International, *150*, 440–466.

Daley, R., 1993. Atmospheric data analysis. *Cambridge University Press*, New York, USA, 472 pp.

Devaney, A. J., 1984. *Geophysical Diffraction Tomography*, IEEE Transactions on Geosciences and Remote Sensing, Vol. GE-22, No. 1, 3–13.

Ekström, G., Tromp, J. and Larson, E.W.F., 1997. *Measurements and models of global surface wave propagation*, Journal of Geophysical Research, *102*, 8137–8157.

Evensen, G., 2009. *Data assimilation: the ensemble Kalman filter* (2nd ed.). *Springer*, New York, USA, 307 pp.

Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P., 2008. *Theoretical background for continental- and global-scale full*

*waveform inversion in the time-frequency domain*. Geophysical Journal International, *175*, 665–685.

Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P., 2009. *Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods*, Geophysical Journal International, *179*, 1703–1725.

Fichtner, A., Kennett, B. L. N., Igel, H., Bunge, H.-P., 2010. *Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle*. Earth and Planetary Science Letters, *290*, 270–280.

Fichtner, A. and Tkalčić, H., 2010. *Insights into the kinematics of a volcanic caldera drop: Probabilistic finite-source inversion of the 1996 Bárdarbunga, Iceland, earthquake*. Earth and Planetary Science Letters, *297*, 607–615.

Gaherty, J., Jordan, T. H., and Gee, L., 1996. *Seismic structure of the upper mantle in a central Pacific corridor*. Journal of Geophysical Research, *101*, 22291–22309.

Gauthier, O., Virieux, J. and Tarantola, A., 1986. *Two-dimensional nonlinear inversion of seismic waveforms: numerical results*. Geophysics, *51*(7), 1387–1403.

Gee, L., and Jordan, T. H., 1992. *Generalized seismological data functionals*. Geophysical Journal International, *111*, 363–390.

Gephart, J.W., 1985. *Principal stress directions and the ambiguity in fault plane identification from focal mechanism*, Bulletin of the Seismological Society of America, *75*, 621–625.

Gephart, J.W., 1990. *Stress and the direction of slip on fault planes*. Tectonics, *9*, 845–858.

Graves, R., 1996. *Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences*. Bulletin of the Seismological Society of America, *86*, 1091–1106.

Harris, R.A., 1998. *Introduction to special session: stress triggers, stress shadows, and implications for seismic hazard*. Journal of Geophysical Research, *103*, 24347–24358.

Heidbreder, G. R. 1967. *Multiple scattering and the method of Rytov*. Journal of the Optical Society of America. Vol *57*, Number 12, 1477–1479.

Holschneider, M., Diallo, M. S., Kulesh, M., Ohrnberger, M., Lück, E. and Scherbaum, F., 2005. *Characterization of dispersive surface waves using continuous wavelet transform*. Geophysical Journal International, *163*, 463–478, doi: 10.1111/j.1365-246X.2005.02787.x

Hung, S.-H., Dahlen, F. A., and Nolet, G., 2000. *Fréchet kernels for finite-frequency travel times, part II: examples*. Geophysical Journal International, *141*, 175–203.

Ishimaru, A., 1978. *Wave Propagation and Scattering in Random Media*, vol. II *Academic Press*, New York, USA, 600 pp.

Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*, *Cambridge University Press*, New York, USA, 364 pp

Katzman, R., Zhao, L., and Jordan, T. H., 1998. *High-resolution, two-dimensional vertical tomography of the central Pacific mantle using ScS*. Journal of Geophysical Research, *103*(B8), 17933–17971.

Komatitsch, D., and Tromp, J., 2002. *Spectral-element simulations of global seismic wave propagation-I. Validation*. Geophysical Journal International, *149*, 390–412.

Komatitsch, D., Liu, Q., Tromp, J., Süss, P., Stidham, C. and Shaw, J. H., 2004. *Simulations of ground motion in the Los Angeles basin based upon the spectral-element method*, Bulletin of the Seismological Society of America, *94*, 187–206.

Laske, G. and Masters, G., 1996. *Constraints on global phase velocity maps from long-period polarization data*. Journal of Geophysical Research, *101*, 16059–16075.

Lilly, J. M. and Park, J., 1995, *Multiwavelet spectral and polarization analysis of seismic records*. Geophysical Journal International, *122*, 1001–1021.

Liu, Q., 2006. *Spectral-element simulations of 3-D seismic wave propagation and applications to source and structural inversions. Dissertation (Ph.D.)*, California Institute of Technology. http://resolver.caltech.edu/CaltechETD:etd-05262006-172954

Liu, Q., & Tromp, J., 2006. *Finite-frequency kernels based upon adjoint methods*. Bulletin of the Seismological Society of America, *96*, 2383–2397.

Luo, Y. and Schuster, G. T., 1991. *Wave equation travel-time inversion*. Geophysics, *56*, 645–653.

Malanotte-Rizzoli, P. (ed.), 1996. *Modern Approaches to Data Assimilation in Ocean Modeling*, Elsevier Oceanography Series, *Elsevier Science B.V.*, Amsterdam, 455 pp.

Marquering, H., Dahlen, F. A. and Nolet, G., 1999. *Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox*. Geophysical Journal International, *137*, 805–815.

McCloskey, J., Nalbant, S.S., Steacy, S., Nostro, C., Scotti, O. andBaumont, D., 2003. *Structural constraints on the spatial distribution of aftershocks*, Geophysical Research Letters, *30*(12), 1610, doi:10.1029/2003GL017225.

McGuire, J. J., Zhao, L. and Jordan, T. H., 2001. *Teleseismic inversion for the second-degree moments of earthquakes*. Geophysical Journal International, *145*, 661–678.

Michael, A.J., 1987. *Use of focal mechanisms to determine stress: a control study*. Journal of Geophysical Research, *92*, 357–368.

Nolet, G., 1987. *Seismic wave propagation and seismic tomography*, in Seismic Tomography: Theory and Practice, pp. 1–24, ed. Nolet, G., Reidel, Dordrecht, *Chapman & Hall*, London, UK, 842 pp.

Olsen, K. B., 1994, *Simulation of 3-D elastic wave propagation in the Salt Lake Basin, PhD Dissertation*, University of Utah, Salt Lake City, USA.

Olsen, K. B., Day, S.M. and Bradley, C.R., 2003. *Estimation of Q for long-period (>2 s) waves in the Los Angeles Basin*. Bulletin of the Seismological Society of America, *93*, 627–638.

Olsen, K. B., Day, S. M., Minster, J. B., Cui, Y., Chourasia, A., Faerman, M., Moore, R., Maechling, P. J. and Jordan,T. H., 2006. *Strong shaking in Los Angeles expected from southern San Andreas earthquake*. Geophysical Research Letters, *33*, L07305, doi:10.1029/2005GL025472

Paige, C., and Saunders, M., 1982. *Algorithm 583 LSQR: Sparse Linear Equations and Least Squares Problems*. ACM Transactions on Mathematical Software.

Parsons, T., 2005. *Significance of stress transfer in time-dependent earthquake probability calculations*. Journal of Geophysical Research, *110*, B05S02, doi:10.1029/2004JB003190.

Pratt, R. G., 1990, *Frequency-domain elastic wave modeling by finite differences: A tool for crosshole seismic imaging (short note)*. Geophysics, *55*, 626–632.

Pratt, R. G., Sin, C. and Hicks, G. J., 1998. *Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion*. Geophysical Journal International, *133*, 341–362.

Pratt, R. G., and Shipp, R., 1999. *Seismic waveform inversion in the frequency domain Part 2: Fault delineation in sediments using cross-hole data*, Geophysics vol. *64*, no. 3 pp. 902–914.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., 1992. *Numerical Recipes in C, the Art of Scientific Computing*, 2nd edition, *Cambridge University Press*, New York, USA, 994 pp

Rhie, J. and Romanowicz, B., 2004. *Excitation of Earth's continuous free oscillations by atmosphere-ocean-seafloor coupling*. Nature, Vol. *431*, pp. 552–556.

Ritsema, J., Rivera, L., Komatitsch, D., Tromp, J. and Van Heijst, H. J., 2002. *The effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios*. Geophysical Research Letters, *29*, L07305, doi:10.1029/2005GL025472.

Sandu, C., Blanchard, E. D. and Sandhu, A., 2007. *A polynomial chaos based Bayesian approach for estimating uncertain parameters of mechanical systems Part 1: Theoretical approach*. In *Computer Science Technical Report TR-07-39*.

Schulz, M., 2006. *Control theory in physics and other fields of science: concepts, tools and applications. Springer*, New York, USA, 300 pp.

Schuster, G., 1996. *Resolution limits for cross well migration and travel time tomography*, Geophysical Journal International, *127*(2), 427–440.

Sirgue, L. and Pratt, R. G., 2004. *Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies*. Geophysics, Vol. *69*, No. 1, p. 231–248.

Simon, D., 2006. *Optimal State Estimation: Kalman, H., and Nonlinear Approaches*. *Wiley-Interscience*, New Jersey, USA, 556 pp.

Snieder, R., and Lomax, A., 1996. *Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes*. Geophysical Journal International, *125* (3), 796–812.

Steacy, S., Nalbant, S.S., McCloskey, J., Nostro, C., Scotti, O. andBaumont, D., 2005. *Onto what planes should Coulomb stress perturbations be resolved*? Journal of Geophysical Research, *110*, B05S15, doi:10.1029/2004JB003356.

Süss, M. P., and Shaw, J. H., 2003. *P-wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California*, Journal of Geophysical Research, *108*, 2170–2188, doi:10.1029/2001JB001628

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

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. *Seismic tomography of the southern California crust based on spectral-element and adjoint methods*. Geophysical Journal International, *180*, 433–462, doi:10.1111/j.1365-246X.2009.04429.x

Tarantola, A., 1984. *Inversion of seismic reflection data in the acoustic approximation*. Geophysics Vol. *49*, No. 8, 1259–1266.

Tarantola, A., 1988. *Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation*. Pure and Applied Geophysics, Vol. *128*, Nos. 1/2, p. 365–399.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, USA, 342 pp.

Tatarskii, V. L., and United States. National Oceanic and Atmospheric Administration and National Science Foundation (U.S.), 1971. *The effects of the Turbulent Atmosphere on Wave Propagation*, Israel Program for Scientific Translations. http://books.google.com.hk/books?id=IsfiAAAAMAAJ

Thomason, D. J., 1982. *Spectrum estimation and harmonic analysis*, Proceedings of the IEEE, *70*, 1055–1096.

Tromp, J., Tape, C., and Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, **160**, 195–216.

Wiener, N., 1938. *The homogeneous chaos*, American Journal of Mathematics, vol. *60*, no. 4, pp. 897–936.

Wikle, C. K. and Berliner, L. M., 2006. *A Bayesian tutorial for data assimilation*, Physica D: Nonlinear Phenomena, Volume *230*, Issues 1-2, Data Assimilation, Pages 1–16, ISSN 0167-2789, doi:10.1016/j.physd.2006.09.017.

Williamson, P. R., 1991. *A guide to the limits of resolution imposed by scattering in ray tomography*. Geophysics, *56*, 202–207.

Williamson, P. R. and Worthington, M. H., 1993. *Resolution limits in ray tomography due to wave behavior: numerical experiments*, Geophysics, *58*(5), 727–735.

Woodward, R. L. and Masters, G., 1991. *Global upper mantle structure from long-period differential travel-times*, Journalof Geophysical Research, *96*, 6351–6377.

Wu, R.-S., 2003. *Wave propagation, scattering and imaging using dual-domain one-way and one-return propagators*. Pure and Applied Geophysics, *160*, 509–539.

Wu, R.-S. and Aki, K., 1988. *Introduction: Seismic wave scattering in three-dimensionally heterogeneous Earth*. Pure and Applied Geophysics, Vol. *28*, Nos. ½, 1–6.

Wu, R.-S. and Toksöz, M. N., 1987. *Diffraction Tomography and Multisource Holography Applied to Seismic Imaging*, Geophysics, *52*, 11–25.

Wunsch, C., 1996. *The Ocean Circulation Inverse Problem*, *Cambridge University Press*, New York, USA, 456 pp.

Xiu, D. and Karniadakis, G. E., 2003. *Modeling uncertainty in flow simulations via generalized polynomial chaos*, Journal of Computational Physics, vol. *187*, no. 1, pp. 137–167.

Yin, Z.-M., 1996. *An improved method for the determination of the tectonic stress field from focal mechanism data*, Geophysical Journal International, *125*, 841–849.

Zhao, L., Jordan, T. H. and Chapman, C.H., 2000. *Three-dimensional Fréchet differential kernels for seismic delay times*. Geophysical Journal International, *141*, 58–576.

Zhao, L., Chen, P., and Jordan, T. H., 2006. *Strain Green's Tensors, Reciprocity, and Their Applications to Seismic Source and Structure Studies*. Bulletin of the Seismological Society of America, Vol. *96*, No. 5, pp. 1753–1763, doi:10.1785/0120050253.

Zhao, L., Jordan, T. H., Olsen, K. B., and Chen, P., 2005. *Fréchet Kernels for Imaging Regional Earth Structure Based on Three-Dimensional Reference Models*. Bulletin of the Seismological Society of America, Vol. *95*, No. 6, pp. 2066–2080 doi: 10.1785/0120050081.

Zupanski, D. and Zupanski, M., 2006. *Model error estimation employing an ensemble data assimilation approach*. Monthly Weather Review, Vol *134*, pp 1337–1354.

(Received September 25, 2010, revised November 7, 2010, accepted November 9, 2010, Published online January 11, 2011)