Geophys. J. Int. (2008) **175,** 665–685 doi: 10.1111/j.1365-246X.2008.03923.x

Theoretical background for continental- and global-scale full-waveform inversion in the time–frequency domain

Andreas Fichtner,[1] Brian L. N. Kennett,[2] Heiner Igel[1] and Hans-Peter Bunge[1]

[1]*Department of Earth and Environmental Sciences, Geophysics Section, Ludwig-Maximilian University, Munich, Germany.*
E-mail: andreas.fichtner@geophysik.uni-muenchen.de
[2]*Research School of Earth Sciences, Australian National University, Canberra, Australia*

Accepted 2008 July 22. Received 2008 July 22; in original form 2008 May 19

SUMMARY

We propose a new approach to full seismic waveform inversion on continental and global scales. This is based on the time–frequency transform of both data and synthetic seismograms with the use of time- and frequency-dependent phase and envelope misfits. These misfits allow us to provide a complete quantification of the differences between data and synthetics while separating phase and amplitude information. The result is an efficient exploitation of waveform information that is robust and quasi-linearly related to Earth's structure. Thus, the phase and envelope misfits are usable for continental- and global-scale tomography, that is, in a scenario where the seismic wavefield is spatially undersampled and where a 3-D reference model is usually unavailable. Body waves, surface waves and interfering phases are naturally included in the analysis. We discuss and illustrate technical details of phase measurements such as the treatment of phase jumps and instability in the case of small amplitudes.

The Fréchet kernels for phase and envelope misfits can be expressed in terms of their corresponding adjoint wavefields and the forward wavefield. The adjoint wavefields are uniquely determined by their respective adjoint-source time functions. We derive the adjoint-source time functions for phase and envelope misfits. The adjoint sources can be expressed as inverse time–frequency transforms of a weighted phase difference or a weighted envelope difference.

In a comparative study, we establish connections between the phase and envelope misfits and the following widely used measures of seismic waveform differences: (1) cross-correlation time-shifts; (2) relative rms amplitude differences; (3) generalized seismological data functionals and (4) the L_2 distance between data and synthetics used in time-domain full-waveform inversion.

We illustrate the computation of Fréchet kernels for phase and envelope misfits with data from an event in the West Irian region of Indonesia, recorded on the Australian continent. The synthetic seismograms are computed for a heterogeneous 3-D velocity model of the Australian upper mantle, with a spectral-element method. The examples include *P* body waves, Rayleigh waves and *S* waves, interfering with higher-mode surface waves. All the kernels differ from the more familar kernels for cross-correlation time-shifts or relative rms amplitude differences. The differences arise from interference effects, 3-D Earth's structure and waveform dissimilarities that are due to waveform dispersion in the heterogeneous Earth.

Key words: Inverse theory; Seismic tomography; Computational seismology; Wave propagation.

1 INTRODUCTION

1.1 State of the art and summary of previous work

In recent years, developments in both theoretical seismology and numerical mathematics have led to substantial progress in seismic tomography. It is today widely accepted that wave propagation effects such as multipathing, scattering or wave front healing are important in the 3-D Earth (Williamson 1991; Williamson & Worthington 1993; Spetzler *et al.* 2001). Theories and methods that allow us to account for such phenomena in seismic tomographies and to go beyond classical ray theory, have been developed by several authors (e.g. Bamberger *et al.* 1982; Tarantola

1988; Yomogida 1992; Marquering *et al.* 1998; Dahlen *et al.* 2000; Zhao *et al.* 2000). The evidence that the results of seismic tomography do indeed improve when finite-frequency effects in 3-D media are included is steadily increasing (Yoshizawa & Kennett 2004; Zhou *et al.* 2005; Boschi 2006; Chen *et al.* 2007a). Moreover, the amount of data that we can exploit has grown immensely, thanks to efficient numerical techniques that enable us to simulate global seismic wave propagation through realistic earth models (e.g. Seriani 1998; Komatisch & Tromp 2002; de la Puente *et al.* 2007; Dumbser *et al.* 2007). We could, in principle, use complete seismic waveforms for the purpose of seismic tomography. Yet, full-waveform tomography has so far been limited to regional-scale problems (e.g. Bamberger *et al.* 1982; Igel *et al.* 1996; Dessa *et al.* 2004; Chen *et al.* 2007a). Full-waveform tomography on a global scale—as envisioned by Capdeville *et al.* (2005)—has not been achieved to date. An important part of the problem is the definition of suitable seismic waveform misfits.

In this paper, we propose an alternative approach to full-waveform inversion that is applicable on continental and global scales. It is based on the formulation of phase and envelope misfits in the time–frequency domain. We are able to provide a complete quantification of waveform differences and to separate information that is quasi-linearly related to structure (phase) from information that is non-linearly related to structure (amplitude). The phase misfit introduces a transmission tomography component that is crucial on continental and global scales, where the distribution of sources and receivers is sparse, in contrast to exploration scenarios characterized by spatial oversampling. Moreover, our approach does not rely on the isolation of particular phases, and it naturally combines body and surface wave analyses.

Full-waveform inversion, based on numerical solutions of the wave equation, was initiated in the early 1980s (Bamberger *et al.* 1982; Tarantola 1984) in the context of 1-D and 2-D seismic exploration problems. Numerical solution of the wave equation, for example, with a finite-difference scheme, automatically includes all types of waves in the synthetic seismograms, even when the earth model is complex. To exploit this wealth of information, time-domain full-waveform inversion attempts to minimize the objective functional $\sum_{r=1}^{N} \int [u(\mathbf{x}^r, t) - u^0(\mathbf{x}^r, t)]^2 \, \mathrm{d}t$, that is, the cumulative L_2 distance between the complete data u^0 and the complete synthetics u, recorded at N receivers, denoted by $\mathbf{x}^r$. The pure frequency-domain approach proposed by Pratt (1999) is analytically equivalent to the pure time-domain method but offers numerical advantages in 2-D exploration scenarios. The minimization problem needs to be solved iteratively due to the high computational costs involved in the numerical solution of the wave equation. Algorithms that have proved to be efficient include the method of steepest descent (e.g. Tarantola 1984; Gauthier *et al.* 1986) and variants of the conjugate gradient method (e.g. Mora 1987, 1988; Tape *et al.* 2007).

One of the theoretical cornerstones of full-waveform inversion is the adjoint method, originally developed in the context of optimal control theory (Lions 1968) and introduced to geophysics by Chavent *et al.* (1975). The adjoint method allows us to compute the gradient of any differentiable misfit functional by solving the forward problem (e.g. the wave equation) and its adjoint problem only once. The simplicity of calculations makes the adjoint method much more efficient than the approximation of the gradient by finite differences, which is practically infeasible when the model space comprises a large number of elements. Applications of the adjoint method are not limited to seismology (e.g. Tarantola 1988; Tromp *et al.* 2005), but can also be found in several other branches of the Earth sciences, including meteorology (e.g. Talagrand & Courtier 1987), ground water modelling (e.g. Sun 1994) or mantle dynamics (e.g. Bunge *et al.* 2003).

The advantages and disadvantages of time-domain full-waveform inversion become apparent through its close relation to diffraction tomography. In diffraction tomography (Devaney 1984; Wu & Toksöz 1987), the scattered wavefield, that is, the difference $\Delta u(t) = u(t) - u^0(t)$, is linearly related to the spectrum of the heterogeneity that caused the incident wavefield to be scattered. This relationship is reminiscent of the well-known projection slice theorem from X-ray tomography. Diffraction tomography yields accurate images, even of small-scale structural heterogeneities, but only when the following conditions are satisfied. (1) The background structure is so close to the true structure that the remaining differences are small compared with the dominant wavelength, that is, the Born approximation holds. (2) The heterogeneity is illuminated from all directions (Mora 1989).

Diffraction tomography is qualitatively equivalent to the first iteration in a time-domain full-waveform inversion—an interpretation confirmed by the comparison of numerical results (Gauthier *et al.* 1986; Wu & Toksöz 1987; Mora 1988). This suggests that time-domain full-waveform inversion works under conditions that are similar to those given above for diffraction tomography. However, meeting those conditions is problematic when the Earth is studied on continental or global scales.

A sufficiently accurate reference model is usually unavailable at length scales exceeding several tens of kilometres. This is due to the nature of the tomography problem with limited data. Even 1-D models of the whole Earth (Dziewonski & Anderson 1981; Morelli & Dziewonski 1993; Kennett *et al.* 1995) can differ by several per cents, especially in the upper mantle and near discontinuities. The 1-D density structure of the Earth has rather limited constraints (Kennett 1998). There are three immediate consequences of this dilemma. First, the remaining differences between any presently available earth model and the true Earth can often not be treated as scatterers. Second, the observed waveform residuals at periods above several seconds are mostly due to transmission and interference effects. Third, time-domain full-waveform inversion on continental or global scales is highly non-linear.

The data coverage necessary for diffraction tomography or time-domain full-waveform inversion is not achievable in a 3-D Earth. Vast regions of the Earth's surface are practically inaccessible, and sufficiently strong sources are confined to a few seismogenic zones. Strong reflectors that could in principle improve this situation (Mora 1989) are not present in the Earth's mantle. As a result, time-domain full-waveform inversion cannot work on continental or global scales, unless very long-period data are used (Capdeville *et al.* 2005).

Nevertheless, it remains desirable to extract as much waveform information as possible and to use numerical solutions of the wave equation in tomography because they allow us to correctly account for 3-D Earth's structure. A milestone towards this goal was Luo & Schuster's (1991) realization that phase information needs to be included explicitly in the objective functional. Separating phases from amplitudes is required to overcome the excessive non-linearity introduced by the objective functional $\sum_{r=1}^{N} \int [u(\mathbf{x}^r, t) - u^0(\mathbf{x}^r, t)]^2 \, \mathrm{d}t$. Luo &

Schuster's method was based on the estimation of delay times by cross-correlating data and numerically computed synthetics—a technique reminiscent of ideas expressed earlier by Dziewonski *et al.* (1972), Lerner-Lam & Jordan (1983) or Cara & Lévêque (1987) in the context of surface wave analysis. The cross-correlation approach was further formalized by Gee & Jordan (1992) and then directly applied to data by Zhou *et al.* (1995) and Chen *et al.* (2007a). It was also used for the computation of finite-frequency delay time kernels (Dahlen *et al.* 2000; Tromp *et al.* 2005; Sieminski *et al.* 2007a,b; Liu & Tromp 2008).

Waveform cross-correlation is a successful and robust technique. However, its applicability is limited to scenarios where single phases are clearly identifiable and where data and synthetics have essentially identical waveforms. In a case where two different phases interfere, the cross-correlation technique may not yield accurate delay time information for either of them. A similar effect arises when data and synthetic waveforms are not only time-shifted but distorted with respect to each other—a well-known phenomenon in the Earth where waveforms disperse due to the presence of heterogeneities. As a consequence of these limitations, potentially useful and robust information about the Earth's structure may remain unexploited by cross-correlation measurements.

1.2 Objectives and outline

The principal objective of this paper is to propose a parametrization for waveform misfit that overcomes the problems of time-domain full-waveform inversion, yet has a wider range of applicability than pure measurements of cross-correlation time-shifts. In more detail, the proposed waveform misfits are designed to fulfil the following requirements: (1) applicability in the context of continental- and global-scale waveform inversion; (2) full quantification of seismic waveform misfit in a suitable frequency range between ω_0 and ω_1, that is, data = synthetics for $\omega \in [\omega_0, \omega_1]$ in the hypothetical case of zero misfit; (3) separation of phase and amplitude information; (4) relaxation of the requirements on waveform similarity needed for the measurement of pure cross-correlation time-shifts and (5) possibility to analyse complete wave trains, including body waves, surface waves and interfering phases.

The seismic waveform misfits proposed in the following sections are based on a seismogram analysis in the time–frequency plane. This approach closely follows suggestions by Kristeková *et al.* (2006), and it is reminiscent of classical surface wave analysis (e.g. Dziewonski *et al.* 1972). Representing both data and synthetics in time–frequency space naturally introduces phase and envelope misfits. They fully characterize the time dependence of the misfit spectrum.

We start our development with the definition of the phase and envelope misfits that we will then use throughout this paper. Much emphasis will be on the technical details of phase difference measurements and the required degree of waveform similarity. Subsequently, we will derive the adjoint sources corresponding to measurements of phase and envelope misfits. This will be followed by a comparison of our results with those obtained for other widely used objective functionals, including cross-correlation time-shifts (e.g. Luo & Schuster 1991), rms amplitudes and the generalized seismological data functionals (GSDFs) by Gee & Jordan (1992). We will emphasize the conceptual similarity between the GSDFs and the time–frequency misfits. In Section 5, we will then present sensitivity kernels for phase and envelope misfits between real data and spectral-element synthetics. The examples include *P*, surface and *S* waves that interfere with higher-mode surface waves. The advantages and disadvantages of our method will be discussed in the final section of this paper. There we will also address possible inverse problem strategies involving phase and amplitude misfits. In a follow-up paper, we will demonstrate the use of our method in waveform inversion on continental scales and in particular on the Australian continent.

2 SEISMIC WAVEFORM MISFITS IN THE TIME–FREQUENCY DOMAIN

Seismic tomography allows us to infer Earth's structure from the misfit between data and synthetics. The success of a seismic tomography depends, among other factors, on the completeness of the misfit quantification, the separation of phases and envelopes and the exploitation of both time and frequency information. We, therefore, propose to characterize the temporal evolution of the frequency content of data and synthetics. This leads to naturally separated phase and envelope differences that depend on time and frequency. The resulting phase and envelope misfits are then complete descriptions of the waveform misfit between data and synthetics, which can be used in a tomographic study exploiting seismic waveforms.

2.1 The definition of phase and envelope misfits

We denote the ith component of a real seismogram recorded at the position $\mathbf{x} = \mathbf{x}^r$ by $u_i^0(\mathbf{x}^r, t)$ and the corresponding synthetic by $u_i(\mathbf{x}^r, t)$. For notational brevity, we will omit dependences on $\mathbf{x}^r$ wherever possible. We can analyse how the frequency content of the data evolves with time by computing the Fourier transforms of $u_i^0(t)$ multiplied by a sliding window function $h(t - \tau)$, centred around τ. In symbols

$$\tilde{u}_i^0(t, \omega) = F_h\left[u_i^0\right](t, \omega) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u_i^0(\tau)\, h^*(\tau - t)\, \mathrm{e}^{-\mathbf{i}\omega\tau}\, \mathrm{d}\tau. \tag{1}$$

The imaginary unit $\mathbf{i} := \sqrt{-1}$ is written in bold face to distinguish it from the index variable i. Using the complex conjugate h^* instead of h is a common convention to which we shall conform throughout this paper. Eq. (1) defines a time–frequency representation of the data. A possible choice for h is the Gaussian $(\pi\sigma)^{-1/4}\, \mathrm{e}^{-t^2/2\sigma^2}$, in which case $\tilde{u}_i^0(t, \omega)$ is referred to as the Gabor transform of $u_i^0(t)$. In analogy to (1), we define the time–frequency representation of the synthetics $u_i(t)$ as $\tilde{u}_i(t, \omega) = F_h[u_i](t, \omega)$. Both, $\tilde{u}_i^0$ and $\tilde{u}_i$ can be written in exponential

form:

$$\tilde{u}_i^0(t,\omega) = |\tilde{u}_i^0(t,\omega)|\, \mathrm{e}^{\mathrm{i}\phi_i^0(t,\omega)}, \quad \tilde{u}_i(t,\omega) = |\tilde{u}_i(t,\omega)|\, \mathrm{e}^{\mathrm{i}\phi_i(t,\omega)}. \tag{2}$$

These relations define the envelopes $|\tilde{u}_i^0(t,\omega)|$ and $|\tilde{u}_i(t,\omega)|$ and the corresponding phases $\phi_i^0(t,\omega)$ and $\phi_i(t,\omega)$, with $-\pi < \phi_i^0, \phi_i \leq \pi$. We now obtain envelope misfits E_e and phase misfits E_p in the form of L_n norms of the envelope difference $|\tilde{u}_i| - |\tilde{u}_i^0|$ and the phase difference $\phi_i - \phi_i^0$, respectively:

$$E_\mathrm{e}^n\left(u_i^0, u_i\right) := \int_{\mathbb{R}^2} W_\mathrm{e}^n(t,\omega)\left[|\tilde{u}_i(t,\omega)| - \left|\tilde{u}_i^0(t,\omega)\right|\right]^n \,\mathrm{d}t\,\mathrm{d}\omega, \tag{3a}$$

$$E_\mathrm{p}^n\left(u_i^0, u_i\right) := \int_{\mathbb{R}^2} W_\mathrm{p}^n(t,\omega)\left[\phi_i(t,\omega) - \phi_i^0(t,\omega)\right]^n \,\mathrm{d}t\,\mathrm{d}\omega. \tag{3b}$$

The symbols W_e and W_p denote positive weighting functions that we will discuss in the following sections. The envelope difference $|\tilde{u}_i^0| - |\tilde{u}_i|$ represents time- and frequency-dependent discrepancies between the amplitudes of u_i^0 and u_i. It is a more informative measure of amplitude differences than the time-domain difference $u_i^0 - u_i$ because it does not generate large misfits when u_i^0 and u_i are similar and merely slightly time-shifted. An alternative to the measurement of the envelopes themselves is the measurement of logarithmic envelopes, that is, the replacement of E_e as defined in eq. (3a) by $E_\mathrm{e}^n = \int W_\mathrm{e}^n[\log(|\tilde{u}_i|/|\tilde{u}_i^0|)]^n \,\mathrm{d}t\,\mathrm{d}\omega$. Logarithmic envelopes may be more linearly related to Earth's structure than the pure envelopes. With the exception of Section 3.5, we will, nevertheless, continue to consider the pure envelopes because this will keep the following developments more readable. The transition to logarithmic amplitudes is mostly trivial. The phase difference $\Delta\phi_i = \phi_i - \phi_i^0$ can be interpreted in terms of a time-shift Δt at frequency ω: $\Delta\phi_i = \omega\Delta t$. The quantity $\Delta\phi_i$ is physically meaningful only when it takes values between $-\pi$ and π, that is, when data and synthetics are out of phase by less than half a period. The same condition arises when arrival time differences are estimated by waveform cross-correlation. Sufficient waveform similarity can be achieved by filtering both data and synthetics or their respective time–frequency representations.

Throughout this paper, we shall use the window function $h(t) = (\pi\sigma)^{-1/4}\,\mathrm{e}^{-t^2/2\sigma^2}$, that is, the Gabor transform. This choice is convenient and, also, advantageous from a theoretical point of view because it maximizes the time–frequency resolution (see for example Strang & Nguyen 1996, and Appendix B). We will suggest suitable values for the parameter σ in Section 2.2.3 Eq. (1) represents not the only possible characterization of data or synthetics in time–frequency space. Alternatively, one could employ a wavelet transform instead of a windowed Fourier transform (e.g. Kristeková *et al.* 2006). We defer a discussion of this issue to Section 6.

2.2 Technical details of phase measurements

Although envelope measurements are conceptually straightforward, phase and phase difference measurements are more complicated. The reasons are: (1) the phases ϕ_i and ϕ_i^0 can be discontinuous. The discontinuities map into the phase difference $\Delta\phi_i$, at least when it is computed by directly subtracting ϕ_i^0 from ϕ_i. (2) Even the smallest wavefield perturbation can in principle lead to strong variations of the phase. This effect is most pronounced when the signal amplitude is comparatively small or even zero. Therefore, a suitable weighting function W_p needs to be applied. We will address these issues in the following paragraphs.

2.2.1 Phase jumps and the practical implementation of phase difference measurements

A fundamental complication related to the measurement of the phase difference $\Delta\phi_i = \phi_i - \phi_i^0$ arises from the discontinuities in the phases ϕ_i and ϕ_i^0. For a given time t, the discontinuities of ϕ_i and ϕ_i^0 generally occur at different frequencies ω. Since the phase jumps from $-\pi$ to π, or vice versa have different locations on the frequency axis, the difference $\Delta\phi_i$ can reach values of $\pm 2\pi$, even when the signals u_i and u_i^0 are almost in phase. Unwrapping ϕ_i and ϕ_i^0 along the frequency axis for each time does not resolve the problem. This is because the number of phase jumps within the frequency interval of interest generally varies with time. Hence, while removing the phase jumps on the frequency axis, an unwrapping procedure produces jumps on the time axis.

Under the assumption that data and synthetics are indeed nearly in phase we can circumvent this obstacle as follows. First, we note that for a fixed time t, the time–frequency representations $\tilde{u}_i^0$ and $\tilde{u}_i$ are the Fourier transforms of the functions $f_t^0(\tau) =: u_i^0(\tau)h^*(\tau - t)$ and $f_t(\tau) =: u_i(\tau)h^*(\tau - t)$, respectively. Now, let the correlation function of f_t and f_t^0, denoted by $c(f_t^0, f_t)(\tau)$, be defined by

$$c\left(f_t^0, f_t\right)(\tau) := \int_{\mathbb{R}} f_t^0(t') f_t(t'+\tau)\,\mathrm{d}t'. \tag{4}$$

For the Fourier transform of $c(f_t^0, f_t)$, we then find

$$F_1[c](\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} c\left(f_t^0, f_t\right)(\tau)\,\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau = \sqrt{2\pi}\,\tilde{u}_i(t,\omega)\,\tilde{u}_i^0(t,\omega)^* = \sqrt{2\pi}|\tilde{u}_i||\tilde{u}_i^0|\,\mathrm{e}^{\mathrm{i}(\phi_i - \phi_i^0)} = |F_1[c]|\,\mathrm{e}^{\mathrm{i}(\phi_i - \phi_i^0)}. \tag{5}$$

The phase difference $\Delta\phi_i = \phi_i - \phi_i^0$ can therefore be expressed as

$$\Delta\phi_i = -\mathbf{i}\,\mathrm{Ln}\left(\frac{F_1[c]}{|F_1[c]|}\right). \tag{6}$$

This means that for a specific time t, the phase of the Fourier transformed correlation function coincides with the phase difference between the time–frequency signals $\tilde{u}_i$ and $\tilde{u}_i^0$. In those regions of time–frequency space where u_i^0 and u_i are nearly in phase, (6)

will give phase differences that range between $-\pi$ and π. Discontinuities, as produced by directly subtracting ϕ_i from ϕ_i^0, are then avoided.

When data and synthetics are, however, out of phase, discontinuities appear even when (6) is used for the computation of the phase difference $\Delta\phi_i$ (see Section 2.2.4 for an example). This indicates a level of waveform dissimilarity between data and synthetics that needs to be tamed either by filtering $u_i(t)$ and $u_i^0(t)$ or by applying a suitable weighting function W_p to the time–frequency representations $\tilde{u}_i$ and $\tilde{u}_i^0$.

2.2.2 The necessary condition for the stability of phase measurements

Phase measurements are further complicated by the fact that even the smallest wavefield perturbation can lead to strong variations of the phase unless a reasonable weighting function W_p is chosen. Choosing, for example, $W_\mathrm{p}(t, \omega) = 1$ will lead to an erratic behaviour of the weighted phase difference $W_\mathrm{p}\Delta\phi_i$ in cases where $|\tilde{u}_i^0|$ is zero or small compared with the noise.

We shall now deduce classes of phase weighting functions W_p that lead to robust measurements. Our development is based on stability arguments. Assume that the signal $u(t)$, which is either $u_i^0(t)$ or $u_i(t)$, is perturbed by $\Delta u(t)$. This small change will lead to a perturbation of the weighted phase $W_\mathrm{p}\phi(u)$ so that it becomes $W_\mathrm{p}\phi(u + \Delta u) = W_\mathrm{p}\phi(u) + W_\mathrm{p}\Delta\phi(u + \Delta u)$. Our goal is to find a weighting function W_p such that any small perturbation Δu induces a weighted phase change $W_\mathrm{p}\Delta\phi$ that is bounded by $C\,||\Delta u||_2$, where $C < \infty$ is a constant:

$$||W_\mathrm{p}\Delta\phi||_2 \leq C\,||\Delta u||_2\,, \quad \forall\,\Delta u \text{ with } ||\Delta u||_2 \ll ||u||_2. \tag{7}$$

Since $||\Delta u||_2$ is small compared with $||u||_2$ we may approximate $W_\mathrm{p}\,\Delta\phi$ with a Taylor series truncated after the linear term:

$$W_\mathrm{p}\Delta\phi \doteq W_\mathrm{p}\frac{\mathrm{d}\phi}{\mathrm{d}u}(\Delta u), \tag{8}$$

where $(\mathrm{d}\,\phi/\mathrm{d}\,u)(\Delta u)$ denotes the functional derivative of ϕ in the direction of Δu. The symbol $\doteq$ means 'correct to first order'. In what follows we will replace $\doteq$ by $=$ in the interests of simplicity. For the linear operator $\mathrm{d}\phi/\mathrm{d}u$, we find

$$\frac{\mathrm{d}\phi}{\mathrm{d}u}(\Delta u) = -\mathbf{i}\frac{\mathrm{d}}{\mathrm{d}u}\ln\left(\frac{\tilde{u}}{|\tilde{u}|}\right)(\Delta u) = -\mathbf{i}\left(\frac{1}{\tilde{u}}\frac{\mathrm{d}\tilde{u}}{\mathrm{d}u} - \frac{1}{|\tilde{u}|}\frac{\mathrm{d}|\tilde{u}|}{\mathrm{d}u}\right)(\Delta u) = \Im m\left[\frac{1}{\tilde{u}}\frac{\mathrm{d}\tilde{u}}{\mathrm{d}u}(\Delta u)\right]. \tag{9}$$

From the definition of $\tilde{u}$ (see eq. 1), we obtain

$$\frac{\mathrm{d}\tilde{u}}{\mathrm{d}u}(\Delta u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\Delta u)\,h^*(\tau - t)\mathrm{e}^{-\mathbf{i}\omega\tau}\,\mathrm{d}\tau = \Delta\tilde{u}. \tag{10}$$

Hence, $\mathrm{d}\tilde{u}/\mathrm{d}u$ is just equal to F_h. Combining eqs (8)–(10), and invoking Plancherel's formula (see Appendix B) yields an estimate for $||W_\mathrm{p}\Delta\phi||_2$:

$$||W_\mathrm{p}\Delta\phi||_2 \leq \left|\left|W_\mathrm{p}/\tilde{u}\right|\right|_2\,||\Delta\tilde{u}||_2 = ||h||_2\left|\left|W_\mathrm{p}/\tilde{u}\right|\right|_2\,||\Delta u||_2. \tag{11}$$

Based on our definition of a stable measurement, we can conclude that stability requires the quantity $||W_\mathrm{p}/\tilde{u}||_2$ to be bounded by a finite constant. This stability condition puts constraints on the properties of the weighting function W_p. The most straightforward choice is $W_\mathrm{p} = |\tilde{u}|$ because then we have $||W_\mathrm{p}\,\Delta\phi||_2 \leq ||h||_2||\Delta u||_2$ and the constant C from the stability definition (7) is equal to $||h||_2$. This window function emphasizes those parts of the seismogram that have large amplitudes and frequencies close to the dominant frequency. However, phases corresponding to low-amplitude waves, with frequencies that are not close to the dominant one, are down-weighted. This behaviour is reminiscent of the well-known amplitude effect in time-shift estimates by cross-correlation. A direct implication of the large-amplitude dominance is that different arrivals in a seismogram need to be considered separately when $W_\mathrm{p} = |\tilde{u}|$ is used as weighting function. Otherwise, phase information from lower-amplitude arrivals will be practically lost. The large-amplitude dominance of $W_\mathrm{p} = |\tilde{u}|$ can be reduced by choosing

$$W_\mathrm{p} = \log(1 + |\tilde{u}|)/\max_{\omega,t}\,\log(1 + |\tilde{u}|). \tag{12}$$

This weighting function also satisfies the criterion that $||W_\mathrm{p}/\tilde{u}||_2$ be bounded. In addition to the options proposed above, one might use W_p as a noise filter or as a means for emphasizing specific seismic phases, for example, small-amplitude core phases.

2.2.3 Suitable choices for the Gaussian window parameter σ

The choice of the parameter σ in the Gaussian window $h(t) = (\pi\sigma^2)^{-1/4}\,\mathrm{e}^{-t^2/2\sigma^2}$ influences the time–frequency representations of both data and synthetics. The technical details of the measurement process, represented by σ in our case, affect the outcome of the measurement. Optimizing this outcome has always played a central role in time–frequency and spectral analysis. In classical surface wave analysis, for example, σ is usually tuned to render measurements of group arrival times as easy as possible (Cara 1973; Nyman & Landisman 1977). Another example of measurement optimization comes from multitaper methods, which are designed to provide spectral estimates that are as free as possible from the effects of windowing functions (Thomson 1982).

In the case of full-waveform tomography, the situation is slightly different from the ones encountered in classical surface wave or multitaper analysis. The measurements, namely phase and envelope misfits, are extracted from time–frequency representations, and those time–frequency representations have, by design, a free parameter σ. This parameter makes the subjectivity inherent in any measurement

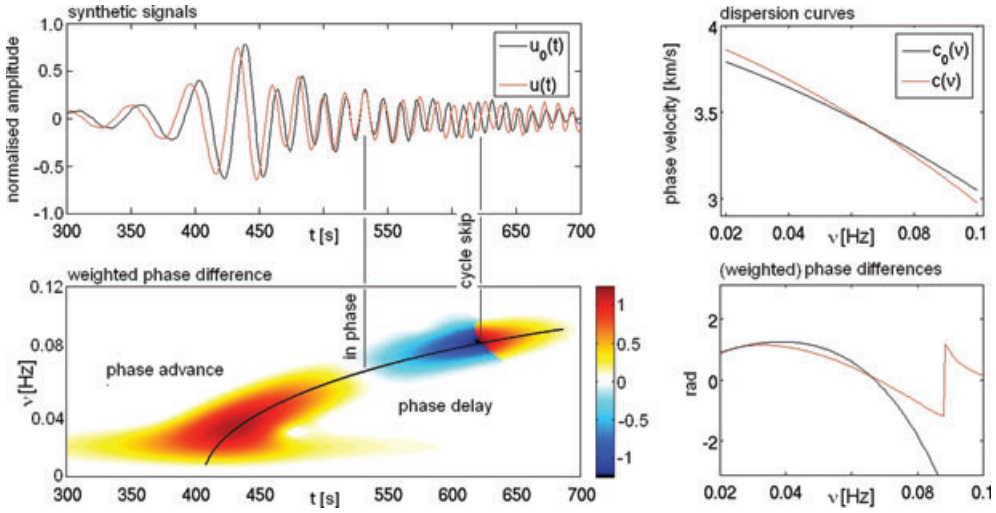

Figure 1. Upper left-hand panel: time-domain signals $u^0(t)$ plotted in black and $u(t)$ plotted in red. Upper right-hand panel: dispersion curves $c(\nu)$ and $c^0(\nu)$ in the frequency range that contributes to $u(t)$ and $u^0(t)$ (see eqs 13). Lower left-hand panel: weighted phase difference $W_p \Delta\phi$ as a function of time t and frequency ν. The weighting function is $W_p = \log(1 + |\tilde{u}^0|)/\max \log(1 + |\tilde{u}^0|)$. Vertical lines indicate the times when the signals are in phase and phase shifted by $-\pi$, respectively. The bold line traces the extremal weighted phase differences for a given frequency. Those values are plotted separately in the lower right-hand panel (red line). The weighted phase difference is compared with the phase difference $-\omega x/c(\omega) + \omega\, x/c^0(\omega)$ plotted in black.

rather explicit. In principle, we cannot exclude *a priori* a certain value for σ as long as it results in a physically meaningful measurement and as long as we interpret the results accordingly.

We can, however, tune σ such that it produces results that are in agreement with our physical intuition and experience. In this sense, we suggest choosing σ such that the mathematically defined phase difference $\Delta\phi_i$ is interpretable in terms of the intuitive meaning of a phase difference—a time-shift between two associated oscillations in the data and the synthetics. A suitable value for σ is then the dominant period of the data. Choosing σ to be several times smaller than the dominant period gives narrow Gaussian windows that cannot capture time-shifts between two cycles that span many such windows. Conversely, a value for σ that is several times larger than the dominant period leads to Gaussian windows that are so wide that the resulting phase difference can no longer be associated to a specific cycle. In the case where the dominant period varies strongly with time, one may use a time- or frequency-dependent window h. We discuss this issue in Section 6.

2.2.4 A conceptual example

This example is intended to illustrate some of the concepts arising from the measurement of phase differences and is intended to provide a compromise between realistic and reproducible results. As test signals we use the following analytically defined, dispersed wave trains:

$$u(x,t) = \int_{2\pi/50}^{2\pi/15} \omega \, \cos[\omega t - \omega x/c(\omega)] \, \mathrm{d}\omega\,, \quad c(\omega) = (4 - \omega - \omega^2)\,\mathrm{km\,s^{-1}}, \tag{13a}$$

$$u^0(x,t) = \int_{2\pi/50}^{2\pi/15} \omega \, \cos[\omega t - \omega x/c^0(\omega)] \, \mathrm{d}\omega\,, \quad c^0(\omega) = (3.91 - 0.87\omega - 0.8\omega^2)\,\mathrm{km\,s^{-1}}, \tag{13b}$$

with the epicentral distance $x = 1500$ km and $\sigma = 25$ s. Both $u(t)$ and $u^0(t)$ are shown in the upper left-hand panel of Fig. 1. The dispersion curves $c(\nu)$ and $c^0(\nu)$ with $\nu = \omega/2\pi$, are displayed in the upper right-hand panel of Fig. 1. A visual comparison of $u(t)$ and $u^0(t)$ indicates that $u(t)$ is advanced relative to $u^0(t)$, prior to $t \approx 500$ s. For later times, $u(t)$ is delayed. This delay develops into a phase shift of $-\pi$ relative to $u^0(t)$, around $t = 620$ s. All of these features map into the weighted phase misfit $W_p \Delta\phi$ (lower left-hand panel of Fig. 1), where the weighting function W_p is set to $W_p = \log(1 + |\tilde{u}^0|)/\max_{\omega,t} \log(1 + |\tilde{u}^0|)$.

The phase advance of $u(t)$ is most pronounced between 400 and 450 s where the amplitudes, and therefore W_p, are largest. The phase advance disappears at $t = 520$ s, where the two signals $u(t)$ and $u^0(t)$ are exactly in phase. This is indicated by a vertical line in Fig. 1. Subsequently, $W_p \Delta\phi$ becomes negative, meaning that the phase of $u(t)$ is smaller than the phase of $u^0(t)$. This delay of $u(t)$ at later times and higher frequencies is very pronounced. Nevertheless, the values of $W_p \Delta\phi$ do not drop below -1.0 because W_p tends to suppress phase differences associated with smaller-amplitude signals. For $t \approx 620$ s, the two signals are exactly out of phase, therefore, producing a discontinuity in the weighted phase difference. This discontinuity corresponds to a cycle skip, meaning that one oscillation cycle of $u(t)$ cannot be associated uniquely with an oscillation cycle of $u^0(t)$. The cycle skip is synonymous with inadequate waveform similarity. For the weighted phase difference to be physically meaningful, the late-arriving and incoherent oscillations need to be excluded either by filtering $u(t)$ and $u^0(t)$ or by tapering in the time–frequency domain.

The effect of the phase weighting function W_p on the phase difference measurement is illustrated in Fig. 2. The unweighted phase difference $\Delta\phi$ is not well defined at frequencies higher than 1/15 Hz, due to the vanishing envelopes $|\tilde{u}|$ and $|\tilde{u}^0|$ (see the definitions of the

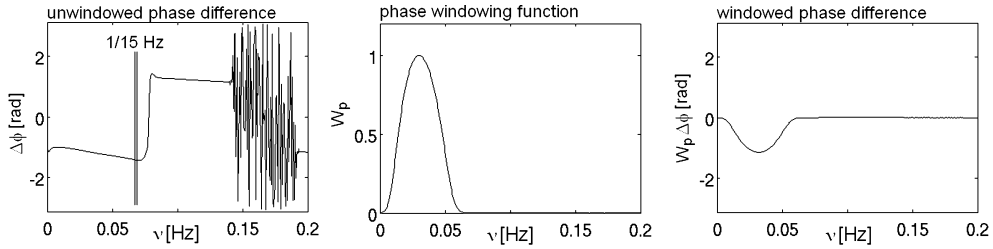

Figure 2. Left-hand panel: unweighted phase difference $\Delta\phi$ at $t = 450$ s. Central panel: phase weighting function $W_\mathrm{p} = \log(1 + |\tilde{u}^0|)/\max \log(1 + |\tilde{u}^0|)$ at $t = 450$ s. Right-hand panel: windowed phase difference $W_\mathrm{p}\Delta\phi$ at $t = 450$ s.

test signals in eqs 13). Numerical errors are the most likely reason for the rapid variation of $\Delta\phi$ beyond $1/15$ Hz. From the phase weighting function, we expect to suppress the erratic behaviour at frequencies where $|\tilde{u}|$ and $|\tilde{u}^0|$ are small, that is, in this example, beyond $1/15$ Hz. In the frequency range where the envelopes of data and synthetics are significantly different from zero, the phase weighting function should be close to unity. As already discussed in Section 2.2.2, one of many possible choices is $W_\mathrm{p} = \log(1 + |\tilde{u}^0|)/\max_{\omega,t} \log(1 + |\tilde{u}^0|)$. This function is displayed in the centre of Fig. 2 for $t = 450$ s. Applying W_p to the unweighted phase difference leads to a well-defined weighted phase difference $W_\mathrm{p}\Delta\phi$, displayed in the right-hand panel of Fig. 2.

3 SENSITIVITY KERNELS FOR ENVELOPE AND PHASE MISFITS

Seismic tomography is an optimization problem. We seek earth models that minimize a given misfit criterion. When the solution of the forward equations for a large number of models is computationally feasible, the optimization problem can be treated probabilistically (Press 1968; Trampert *et al.* 2004). Otherwise, we rely on linearizations or gradient methods for non-linear optimization. In this paper, we restrict our attention to the latter case because we assume that the number of model parameters is large and that the elastic wave equation is solved numerically. The following paragraphs are therefore devoted to the computation of sensitivity kernels for envelope and phase misfits with respect to earth model parameters. We can then obtain gradients by projecting the sensitivity kernels onto the space of model basis functions (e.g. Tarantola 2005).

3.1 A brief review of the adjoint method

We base the computation of sensitivity kernels for phase and envelope misfits on the adjoint method (e.g. Lions 1968; Chavent *et al.* 1975) because it leads to elegant expressions in a rather uncomplicated way and because its numerical implementation is straightforward. An alternative to the adjoint method is the scattering-integral method (Chen *et al.* 2007b), which can be more efficient when the number of sources is much larger than the number of receivers. To establish a consistent notation, but also in the interest of completeness, we shall re-derive or, at least, state some well-known results concerning the adjoint method in the context of elastic wave propagation. These may for example be found in Tarantola (1988) or Tromp *et al.* (2005). We shall employ to the operator formulation of the adjoint method, developed by Fichtner *et al.* (2006).

We assume that $\mathbf{u}(\mathbf{x}, t)$ is an elastic displacement field related to a set of model parameters $\mathbf{m}(\mathbf{x})$ and an external force density $\mathbf{f}(\mathbf{x}, t)$ via $\mathbf{L}(\mathbf{u}, \mathbf{m}) = \mathbf{f}$, where $\mathbf{L}$ represents the wave equation operator. More explicitly, one may write

$$\mathbf{L}(\mathbf{u}, \mathbf{m}) = \rho(\mathbf{x})\,\partial_t^2\mathbf{u}(\mathbf{x}, t) - \nabla\cdot\int_{-\infty}^{t} \boldsymbol{C}(\mathbf{x}, t-\tau) : \nabla\mathbf{u}(\mathbf{x}, \tau)\,\mathrm{d}\tau = \mathbf{f}(\mathbf{x}, t). \tag{14a}$$

The model parameters $\mathbf{m}$ comprise the mass density ρ and the rate of relaxation tensor $\boldsymbol{C}$, that is, $\mathbf{m} = (\rho, \boldsymbol{C})$. In addition to eq. (14a), the displacement field $\mathbf{u}$ is required to satisfy the initial and boundary conditions

$$\mathbf{u}|_{t=0} = \partial_t\mathbf{u}|_{t=0} = \mathbf{0} \quad \text{and} \quad \mathbf{n}\cdot\int_{-\infty}^{t} \boldsymbol{C}(\mathbf{x}, t-\tau) : \nabla\mathbf{u}(\mathbf{x}, \tau)\,\mathrm{d}\tau|_{\mathbf{x}\in\partial\oplus} = \mathbf{0}, \tag{14b}$$

where $\partial\oplus$ denotes the free surface of the earth model $\oplus$. At this point, we neglect possible internal discontinuities. They have been treated, for example, by Liu & Tromp (2008). We represent the process of measuring the wavefield $\mathbf{u}$ or extracting information from it through an objective function $E(\mathbf{u})$, which we assume to be expressible in the form of a time integral $E(\mathbf{u}) = \int_{t_0}^{t_1} \epsilon[\mathbf{u}(\mathbf{x}^r, t)]\,\mathrm{d}t$, with a suitably chosen function $\epsilon(t) = \epsilon[\mathbf{u}\,(\mathbf{x}^r, t)]$. The time t_0 represents the beginning of the measurement and t_1 is the time when the measurement ends. Given $\mathbf{u}$ as a function of time at the receiver position $\mathbf{x} = \mathbf{x}^r$, $E(\mathbf{u})$ may for example return cross-correlation time-shifts (e.g. Luo & Schuster 1991) or rms amplitudes (Dahlen & Baig 2002) of seismic phases. The adjoint method provides an expression for the Fréchet kernel or sensitivity kernel $\delta_m E$, which is the volumetric density of the derivative of E with respect to the model parameters $\mathbf{m}$. In its most general form, this expression is

$$\delta_m E = \int_{\mathbb{R}} \mathbf{u}^\dagger \cdot \partial_m \mathbf{L}(\mathbf{u}, \mathbf{m})\,\mathrm{d}t, \tag{15}$$

where $\partial_m \mathbf{L}$ denotes the partial derivative of the operator $\mathbf{L}$ with respect to the model parameters. Differentiating with respect to ρ, for example, gives $\partial_\rho \mathbf{L}(\mathbf{u}, \mathbf{m}) = \partial_t^2 \mathbf{u}$. The adjoint field $\mathrm{u}^\dagger$ is defined through the adjoint wave equation $\mathbf{L}^\dagger(\mathbf{u}^\dagger, \mathbf{m}) = -\partial_u \epsilon(t)\, \delta(\mathbf{x} - \mathbf{x}^r)$, with

$$\mathbf{L}^\dagger(\mathbf{u}^\dagger, \mathbf{p}) = \rho(\mathbf{x})\, \partial_t^2 \mathbf{u}^\dagger(\mathbf{x}, t) - \nabla \cdot \int_t^\infty \mathbf{C}(\mathbf{x}, \tau - t) : \nabla \mathbf{u}^\dagger(\mathbf{x}, \tau)\, \mathrm{d}\tau = -\partial_u \epsilon(t)\, \delta(\mathbf{x} - \mathbf{x}^r) \tag{16a}$$

and the subsidiary conditions

$$\mathbf{u}^\dagger|_{t=t_1} = \partial_t \mathbf{u}^\dagger|_{t=t_1} = 0 \quad \text{and} \quad \mathbf{n} \cdot \int_t^\infty \mathbf{C}(\mathbf{x}, \tau - t) : \nabla \mathbf{u}^\dagger(\mathbf{x}, \tau)\, \mathrm{d}\tau|_{\mathbf{x} \in \partial \oplus} = \mathbf{0}. \tag{16b}$$

Eq. (16) is still of the wave equation type, and it can therefore be solved with the same algorithms used for the solution of the regular wave equation. However, the adjoint field satisfies terminal conditions, meaning that both $\mathrm{u}^\dagger$ and $\partial_t \mathbf{u}^\dagger$ vanish at the time $t = t_1$, when the measurement ends. For this reason, the adjoint equation is usually solved backwards in time. The adjoint source $\mathbf{f}^\dagger(\mathbf{x}, t) = -\partial_u \epsilon(t)\, \delta(\mathbf{x} - \mathbf{x}^r)$ is a point source, acting at the receiver location $\mathbf{x}^r$. The conceptual simplicity of the adjoint method is due to the fact that the adjoint-source time function $s^\dagger(t) = -\partial_u \epsilon(t)$ fully determines the solution of the adjoint equation and that the adjoint operator $\mathrm{L}^\dagger$ is independent of the misfit functional. From a theoretical point of view, the computation of sensitivity kernels therefore reduces to the computation of the adjoint-source time function that corresponds to a given misfit measure. In the case of an isotropic and non-dissipative medium, described in terms of the mass density ρ and the Lamé parameters μ and λ, the three sensitivity or Fréchet kernels are

$$\delta_\rho E = -\int_{\mathbb{R}} \partial_t \mathbf{u}^\dagger \cdot \partial_t \mathbf{u}\, \mathrm{d}t, \tag{17a}$$

$$\delta_\mu E = \int_{\mathbb{R}} [(\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u}) + (\nabla \mathbf{u}^\dagger) : (\nabla \mathbf{u})^T]\, \mathrm{d}t, \tag{17b}$$

$$\delta_\lambda E = \int_{\mathbb{R}} (\nabla \cdot \mathbf{u}^\dagger)(\nabla \cdot \mathbf{u})\, \mathrm{d}t. \tag{17c}$$

Expressions for Fréchet kernels with respect to other parameters, the *S*-wave speed or the *P*-wave speed, for example, can be derived from eqs (17). A special case arises when $E(\mathbf{u})$ is equal to the ith component of the displacement field, $u_i(\mathbf{x}^r, \tau)$. We then have $\epsilon(t) = \delta(t - \tau)\, \mathbf{e}_i \cdot \mathbf{u}(\mathbf{x}^r, t)$, and the source term of the adjoint eq. (16a) becomes $-\partial_u \epsilon(t) \delta(\mathbf{x} - \mathbf{x}^r) = -\mathbf{e}_i\, \delta(t - \tau) \delta(\mathbf{x} - \mathbf{x}^r)$. This implies that the corresponding adjoint field $\mathrm{u}^\dagger$ is equal to the negative adjoint Green's function for a single force that points in the i-direction, acts at time τ and that is located at $\mathbf{x}^r$. In symbols: $\mathbf{u}^\dagger(\mathbf{x}, t) = -\mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t)$. The Fréchet kernel $\delta_m u_i(\mathbf{x}^r, \tau)$ is therefore given by

$$\delta_m u_i(\mathbf{x}^r, \tau) = -\int_{\mathbb{R}} \mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t) \cdot \partial_m \mathbf{L}[\mathbf{u}(\mathbf{x}, t), \mathbf{m}(\mathbf{x})]\, \mathrm{d}t. \tag{18}$$

With this result in mind, we can now continue with the computation of sensitivity kernels for phase and envelope misfits. As eqs (15) and (16) suggest, it is sufficient to find the adjoint-source time functions corresponding to the different misfits. They will then determine the adjoint wavefield $\mathrm{u}^\dagger$ from (16) and therefore also the sensitivity kernels from (15).

3.2 Adjoint-source time function for the envelope misfit

We start the derivation of the adjoint-source time function corresponding to measurements of the envelope misfit, by differentiating the envelope misfit E_e, as defined in 3(a), with respect to the model parameters $\mathbf{m}$:

$$D_m E_\mathrm{e} = E_\mathrm{e}^{1-n} \int_{\mathbb{R}^2} W_\mathrm{e}^n\, D_m |\tilde{u}_i|\, \left(|\tilde{u}_i| - |\tilde{u}_i^0|\right)^{n-1}\, \mathrm{d}t\, \mathrm{d}\omega. \tag{19}$$

The symbol D_m denotes the functional or Fréchet derivative in the direction of a model perturbation m', that is, $D_m E_\mathrm{e} = (\mathrm{d}E_\mathrm{e}/\mathrm{d}\mathbf{m})(\mathbf{m}')$. At this point, we keep the development as general as possible and therefore do not specify any particular model parameter. In later applications, the model parameter vector $\mathbf{m}$ may be replaced by the shear velocity, density, Q or any other quantity that enters the wave equation. We also leave the window function h that appears in the time–frequency transform (1), unspecified. In the interest of a simplified notation, we omit the dependences on $\mathbf{x}^r$, t, ω, $\mathbf{m}$ and m' whenever they are not needed to understand the formulae. For $D_m|\tilde{u}_i|$ in (19) we substitute (A5) given in Appendix A and find

$$D_m E_\mathrm{e} = E_\mathrm{e}^{1-n}\, \Re e \int_{\mathbb{R}^2} W_\mathrm{e}^n\, (|\tilde{u}_i| - |\tilde{u}_i^0|)^{n-1}\, \frac{\tilde{u}_i}{|\tilde{u}_i|}\, D_m \tilde{u}_i^*\, \mathrm{d}t\, \mathrm{d}\omega, \tag{20}$$

where $\Re e(z)$ denotes the real part of a complex-valued expression z. Writing (20) in terms of the volumetric sensitivity densities $\delta_m E_\mathrm{e}$ and $\delta_m \tilde{u}_i$ yields

$$D_m E_\mathrm{e} = \int_{\mathbb{R}^3} \mathbf{m}' \cdot \delta_m E_\mathrm{e}\, \mathrm{d}^3\mathbf{x} = E_\mathrm{e}^{1-n}\, \Re e \int_{\mathbb{R}^2} W_\mathrm{e}^n (|\tilde{u}_i| - |\tilde{u}_i^0|)^{n-1} \left(\frac{\tilde{u}_i}{|\tilde{u}_i|} \int_{\mathbb{R}^3} \mathbf{m}' \cdot \delta_m \tilde{u}_i^*\, \mathrm{d}^3\mathbf{x} \right) \mathrm{d}t\, \mathrm{d}\omega. \tag{21}$$

The volumetric sensitivity density, that is, the Fréchet kernel $\delta_m E_\mathrm{e}$ can therefore be expressed in terms of the volumetric sensitivity density $\delta_m \tilde{u}_i^*$ as follows.

$$\delta_m E_e = E_\mathrm{e}^{1-n}\, \Re e \int_{\mathbb{R}^2} W_\mathrm{e}^n (|\tilde{u}_i| - |\tilde{u}_i^0|)^{n-1}\, \frac{\tilde{u}_i}{|\tilde{u}_i|} \delta_m \tilde{u}_i^*\, \mathrm{d}t\, \mathrm{d}\omega. \tag{22}$$

The quantity $\delta_m \tilde{u}_i^*$ is the Fréchet kernel of the time–frequency representation $\tilde{u}_i^*$ and can be written in terms of the Fréchet kernel $\delta_m u_i$. The latter is then expressible through the adjoint Green's function $\mathbf{g}_i^\dagger$, as suggested by (18). This sequence of substitutions is summarized in Appendix A and leads to the following expression for $\delta_m E_\mathrm{e}$.

$$\delta_m E_\mathrm{e} = -\frac{E_\mathrm{e}^{1-n}}{\sqrt{2\pi}}\,\Re e \int_{\mathbb{R}^2} W_\mathrm{e}^n(t,\omega)\left[|\tilde{u}_i(t,\omega)| - |\tilde{u}_i^0(t,\omega)|\right]^{n-1}\left\{\frac{\tilde{u}_i(t,\omega)}{|\tilde{u}_i(t,\omega)|}\int_{\mathbb{R}^2}\mathbf{g}_i^\dagger(\mathbf{x}^r,\tau;\mathbf{x},t')\cdot\partial_m\mathbf{L}[\mathbf{u}(t')]\,h(\tau-t)\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau\,\mathrm{d}t'\right\}\mathrm{d}t\,\mathrm{d}\omega. \tag{23}$$

To generate the canonical form $\delta_m E_\mathrm{e} = \int \mathbf{u}_{\mathrm{e},i}^\dagger \cdot \partial_m \mathbf{L}[\mathbf{u}]\,\mathrm{d}t'$, already introduced in eq. (15), we define the adjoint field for envelope measurements as follows:

$$\mathbf{u}_{\mathrm{e},i}^\dagger(\mathbf{x},t') = -\frac{E_\mathrm{e}^{1-n}}{\sqrt{2\pi}}\,\Re e \int_{\mathbb{R}}\mathbf{g}_i^\dagger(\mathbf{x}^r,\tau;\mathbf{x},t')\int_{\mathbb{R}^2} W_\mathrm{e}^n(t,\omega)\left[|\tilde{u}_i(t,\omega)| - |\tilde{u}_i^0(t,\omega)|\right]^{n-1}\left[\frac{\tilde{u}_i(t,\omega)}{|\tilde{u}_i(t,\omega)|}\right]h(\tau-t)\,\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}t\,\mathrm{d}\omega\,\mathrm{d}\tau. \tag{24}$$

A closer look at the integrals in (24) reveals that the adjoint-source time function $s^\dagger{}_{\mathrm{e},i}$ corresponding to the adjoint wavefield $\mathbf{u}_{\mathrm{e},i}^\dagger$ is

$$s_{\mathrm{e},i}^\dagger(\tau) = -\frac{E_\mathrm{e}^{1-n}}{\sqrt{2\pi}}\,\Re e \int_{\mathbb{R}^2} W_\mathrm{e}^n(t,\omega)\left[|\tilde{u}_i(t,\omega)| - |\tilde{u}_i^0(t,\omega)|\right]^{n-1}\left[\frac{\tilde{u}_i(t,\omega)}{|\tilde{u}_i(t,\omega)|}\,h(\tau-t)\mathrm{e}^{\mathrm{i}\omega\tau}\right]\mathrm{d}t\,\mathrm{d}\omega. \tag{25}$$

The subscripts in $s_{\mathrm{e},i}^\dagger$ indicate that the envelope measurement is made on the ith component of the displacement field; they do not symbolize vector components. Eq. (25) implies that the adjoint wavefield is formally given by

$$\mathbf{u}_{\mathrm{e},i}^\dagger(\mathbf{x},t') = \int_{\mathbb{R}} s_{\mathrm{e},i}^\dagger(\tau)\mathbf{g}_i^\dagger(\mathbf{x}^r,\tau;\mathbf{x},t')\,\mathrm{d}\tau. \tag{26}$$

In practice, eq. (26) is not used because it is difficult to compute numerical Green's functions and to convolve them at every point in space with the source time function. Instead, we compute the adjoint wavefield $\mathbf{u}_{\mathrm{e},i}^\dagger$ numerically, by solving the adjoint wave eq. (16) with the adjoint-source time function $-\partial_u \epsilon(t)$, equal to $s_{\mathrm{e,i}}^\dagger(t)$. We can simplify eq. (25) considerably by writing it in terms of the inverse time–frequency transform (see Appendix B):

$$s_{\mathrm{e},i}^\dagger(\tau) = -E_\mathrm{e}^{1-n}\Re e\, F_h^{-1}\left[W_\mathrm{e}^n\left(|\tilde{u}_i| - |\tilde{u}_i^0|\right)^{n-1}\frac{\tilde{u}_i}{|\tilde{u}_i|}\right](\tau). \tag{27}$$

Expression (27) states that the adjoint-source time function for envelope misfits is simply the inverse time–frequency transform of the $(n-1)$ th power of the envelope difference $(|\tilde{u}_i| - |\tilde{u}_i^0|)^{n-1}$ times the weighting function $W_\mathrm{e}^n \tilde{u}_i/|\tilde{u}_i|$.

3.3 Adjoint-source time function for the phase misfit

Our strategy for the derivation of the adjoint-source time function for the phase misfit is identical to the one used in the previous paragraph. First, we differentiate the phase misfit E_p, defined in eq. (3b), with respect to the model parameters $\mathbf{m}$:

$$D_m E_\mathrm{p} = E_\mathrm{p}^{1-n}\int_{\mathbb{R}^2} W_\mathrm{p}^n\, D_m\phi_i\,\left(\phi_i - \phi_i^0\right)^{n-1}\,\mathrm{d}t\,\mathrm{d}\omega. \tag{28}$$

For $D_m\phi_i$ we substitute

$$D_m\phi_i = -\mathbf{i}\,D_m\,\ln\left(\frac{\tilde{u}_i}{|\tilde{u}_i|}\right) = \mathbf{i}\left(\frac{1}{|\tilde{u}_i|}\,D_m|\tilde{u}_i| - \frac{1}{\tilde{u}_i}\,D_m\tilde{u}_i\right) \tag{29}$$

and replace $D_m|\tilde{u}_i|$ by (A5), given in Appendix A. Reordering the terms yields

$$\begin{aligned} D_m E_\mathrm{p} &= \mathbf{i}\,E_\mathrm{p}^{1-n}\int_{\mathbb{R}^2} W_\mathrm{p}^n(\phi_i-\phi_i^0)^{n-1}\left(\frac{\tilde{u}_i}{2|\tilde{u}_i|^2}D_m\tilde{u}_i^* - \frac{\tilde{u}_i^*}{2|\tilde{u}_i|^2}D_m\tilde{u}_i\right)\mathrm{d}t\,\mathrm{d}\omega \\ &= -E_\mathrm{p}^{1-n}\,\Im m\int_{\mathbb{R}^2} W_\mathrm{p}^n(\phi-\phi_i^0)^{n-1}\left(\frac{\tilde{u}_i}{|\tilde{u}_i|^2}\,D_m\tilde{u}_i^*\right)\mathrm{d}t\,\mathrm{d}\omega. \end{aligned} \tag{30}$$

The symbol $\Im m(z)$ denotes the imaginary part of a complex valued quantity z. We now make the transition from the derivatives $D_m E_\mathrm{p}$ and $D_m\tilde{u}_i$ to their respective volumetric densities:

$$D_m E_\mathrm{p} = \int_{\mathbb{R}^3}\mathbf{m}'\cdot\delta_m E_\mathrm{p}\,\mathrm{d}^3\mathbf{x} = -E_\mathrm{p}^{1-n}\,\Im m\int_{\mathbb{R}^2} W_\mathrm{p}^n(\phi_i-\phi_i^0)^{n-1}\left(\frac{\tilde{u}_i}{|\tilde{u}_i|^2}\int_{\mathbb{R}^3}\mathbf{m}'\cdot\delta_m\tilde{u}_i^*\,\mathrm{d}^3\mathbf{x}\right)\mathrm{d}t\,\mathrm{d}\omega. \tag{31}$$

Changing the order of the integration provides an expression for the sensitivity density $\delta_m E_\mathrm{p}$ in terms of the sensitivity density $\delta_m\tilde{u}_i$:

$$\delta_m E_\mathrm{p} = -E_\mathrm{p}^{1-n}\,\Im m\int_{\mathbb{R}^2} W_\mathrm{p}^n(\phi_i-\phi_i^0)^{n-1}\left(\frac{\tilde{u}_i}{|\tilde{u}_i|^2}\,\delta_m\tilde{u}_i^*\right)\mathrm{d}t\,\mathrm{d}\omega. \tag{32}$$

As demonstrated in Appendix A, we can write $\delta_m\tilde{u}_i$, and therefore its complex conjugate, in terms of the i th adjoint Green's function $\mathbf{g}^\dagger{}_i(\mathbf{x}^r,\tau;\mathbf{x},t')$:

$$\delta_m E_\mathrm{p} = \frac{E_\mathrm{p}^{1-n}}{\sqrt{2\pi}}\,\Im m\int_{\mathbb{R}^2} W_\mathrm{p}^n(t,\omega)\left[\phi_i(t,\omega)-\phi_i^0(t,\omega)\right]^{n-1}\left[\frac{\tilde{u}_i(t,\omega)}{|\tilde{u}_i(t,\omega)|^2}\int_{\mathbb{R}^2}\mathbf{g}_i^\dagger(\mathbf{x}^r,\tau;\mathbf{x},t')\cdot\partial_m\mathbf{L}[\mathbf{u}(t')]\,h(\tau-t)\,\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau\,\mathrm{d}t'\right]\mathrm{d}t\,\mathrm{d}\omega. \tag{33}$$

This expression reduces to the canonical form $\delta_m E_\mathrm{p} = \int \mathbf{u}_{\mathrm{p},i}^\dagger\cdot\partial_m\mathbf{L}[\mathbf{u}]\,\mathrm{d}t'$ by defining the adjoint field for the phase measurement as follows:

$$\mathbf{u}_{\mathrm{d},i}^\dagger(\mathbf{x},t') = \frac{E_\mathrm{d}^{1-n}}{\sqrt{2\pi}}\int_{\mathbb{R}^3} W_\mathrm{d}^n(t,\omega)\,\left[\phi_i(t,\omega)-\phi_i^0(t,\omega)\right]^{n-1}\left[\frac{\tilde{u}_i(t,\omega)}{|\tilde{u}_i(t,\omega)|^2}\,\mathbf{g}_i^\dagger(\mathbf{x}^r,\tau;\mathbf{x},t')\,h(\tau-t)\,e^{\mathrm{i}\omega\tau}\right]\mathrm{d}\tau\,\mathrm{d}t\,\mathrm{d}\omega. \tag{34}$$

From eq. (34), we can deduce the adjoint-source time function $s^\dagger_{\mathrm{p},i}(\tau)$ that generates the adjoint wavefield $\mathbf{u}^\dagger_{\mathrm{p},i}(\mathbf{x}, t')$. It is given by

$$s^\dagger_{\mathrm{p},i}(\tau) = \frac{E_\mathrm{p}^{1-n}}{\sqrt{2\pi}}\, \Im m \int_{\mathbb{R}^2} W_\mathrm{p}^n(t,\omega)[\phi_i(t,\omega) - \phi_i^0(t,\omega)]^{n-1} \left[\frac{\tilde{u}_i(t,\omega)}{|\tilde{u}_i(t,\omega)|^2}\, h(\tau - t)\, \mathrm{e}^{\mathrm{i}\omega\tau}\right] \mathrm{d}t\, \mathrm{d}\omega. \tag{35}$$

By using the inverse time–frequency transform F_h^{-1} (Appendix B), we can condense (35) to the following expression:

$$s^\dagger_{\mathrm{p},i}(\tau) = E_\mathrm{p}^{1-n}\, \Im m\, F_h^{-1}\left[W_\mathrm{p}^n(\phi_i - \phi_i^0)^{n-1}\frac{\tilde{u}_i}{|\tilde{u}_i|^2}\right](\tau)\,. \tag{36}$$

The eq. (36) closely resembles (27) which provides the adjoint-source time function for envelope-misfit measurements, with the phase difference $(\phi_i - \phi_i^0)^{n-1}$ playing the role of the envelope difference $(|\tilde{u}_i| - |\tilde{u}_i^0|)^{n-1}$. Note that the term in square brackets is well defined despite the factor $|\tilde{u}_i|^{-2}$: since $n \geq 2$, the term $W_\mathrm{p}^n/|\tilde{u}_i|^2$ is bounded because of the stability requirement for phase measurements.

3.4 Adjoint-source time functions for envelope and phase misfits based on velocity or acceleration seismograms

Often, seismic data come in the form of velocity or acceleration seismograms that cannot be integrated, due to the presence of long-period seismic noise, for example. It can therefore be convenient to measure phase and envelope misfits directly for velocities or accelerations. In those cases, the adjoint-source time functions that we already found for displacement seismograms need to be modified. Since the derivation of the adjoint-source time functions for velocity and acceleration measurements closely follows the scheme introduced in the previous paragraphs, we merely state the results. First, assuming that all envelopes, phases and corresponding misfits are measured from the velocities $v_i = \dot{u}_i$, $v_i^0 = \dot{u}_i^0$, we have

$$s^\dagger_{\mathrm{e},i}(\tau) = E_\mathrm{e}^{1-n}\partial_\tau\, \Re e\, F_h^{-1}\left[W_\mathrm{e}^n\,(|\tilde{v}_i| - |\tilde{v}_i^0|)^{n-1}\frac{\tilde{v}_i}{|\tilde{v}_i|}\right](\tau)\,, \qquad s^\dagger_{\mathrm{p},i}(\tau) = -E_\mathrm{p}^{1-n}\,\partial_\tau\, \Im m\, F_h^{-1}\left[W_\mathrm{p}^n(\phi_i - \phi_i^0)^{n-1}\frac{\tilde{v}_i}{|\tilde{v}_i|^2}\right](\tau). \tag{37}$$

For measurements made from the accelerations $a_i = \ddot{u}_i$, $a_i^0 = \ddot{u}_i^0$, we find in a similar way

$$s^\dagger_{\mathrm{e},i}(\tau) = -E_\mathrm{e}^{1-n}\partial_\tau^2\, \Re e\, F_h^{-1}\left[W_\mathrm{e}^n\,(|\tilde{a}_i| - |\tilde{a}_i^0|)^{n-1}\frac{\tilde{a}_i}{|\tilde{a}_i|}\right](\tau)\,, \qquad s^\dagger_{\mathrm{p},i}(\tau) = E_\mathrm{p}^{1-n}\,\partial_\tau^2\, \Im m\, F_h^{-1}\left[W_\mathrm{p}^n(\phi_i - \phi_i^0)^{n-1}\frac{\tilde{a}_i}{|\tilde{a}_i|^2}\right](\tau). \tag{38}$$

It is important to note that despite being potentially convenient from a purely observational point of view, velocity and acceleration measurements tend to pose numerical problems. The adjoint-source time functions in eqs (37) and (38) are proportional to the second and fourth derivatives of u_i, respectively. The resulting adjoint fields will therefore have higher dominant frequencies than the adjoint fields based on displacement measurements. This is, from a numerical point of view, disadvantageous because higher frequencies require a finer discretization and higher computational costs.

3.5 Adjoint-source time function for measurements of logarithmic envelopes

We mentioned in Section 2.1 that it can be advantageous to measure logarithmic rather than pure envelopes, that is, to use an envelope misfit defined through $E_\mathrm{e}^n = \int W_\mathrm{e}^n[\log(|\tilde{u}_i|/|\tilde{u}_i^0|)]^n\, \mathrm{d}t\, \mathrm{d}\omega$. For completeness, we therefore give the adjoint-source time function corresponding to this objective functional, without derivation,

$$s^\dagger_{\mathrm{e},i}(\tau) = -E_\mathrm{e}^{1-n}\Re e\, F_h^{-1}\left[W_\mathrm{e}^n\left(\log|\tilde{u}_i|/|\tilde{u}_i^0|\right)^{n-1}\frac{|\tilde{u}_i^0|\tilde{u}_i}{|\tilde{u}_i|^2}\right](\tau). \tag{39}$$

4 INTERRELATIONS BETWEEN OBJECTIVE FUNCTIONALS AND THEIR ASSOCIATED ADJOINT SOURCES

When the signals u_i and u_i^0, the window h and the weighting functions W_p and W_e fulfil certain requirements, then the time–frequency misfits yield the same numerical results as other objective functionals that, already, found wide-spread use in seismology. Moreover, the associated adjoint sources and Fréchet kernels are then identical. In the following paragraphs, we will establish relations between the time–frequency misfits and (1) measurements of cross-correlation time-shifts, (2) measurements of rms amplitude differences, (3) GSDFs (Gee & Jordan 1992) and (4) time-domain full-waveform inversion. The comparisons will highlight similarities that occur under well-defined circumstances and point out when significant differences between objective functionals are to be expected. The comparisons will also facilitate the interpretation of the Fréchet kernels shown in Section 5. The following analyses assume that the phase and envelope misfits are the L_2 norms of the corresponding time- and frequency-dependent phase and envelope differences ($n = 2$).

4.1 Phase misfits and cross-correlation time-shifts

The estimation of time-shifts between data and synthetics via cross-correlation is a classical tool in surface wave analysis (Dziewonski *et al.* 1972), which was translated to waveform inversion by Luo & Schuster (1991) to bridge the gap between the highly non-linear time-domain

full-waveform inversion (e.g. Gauthier *et al.* 1986) and ray-based traveltime tomography. The concept has been directly applied to data (e.g. Zhou *et al.* 1995) and is used to derive finite-frequency delay time kernels (e.g. Dahlen *et al.* 2000; Tromp *et al.* 2005).

Cross-correlation time-shifts and phase misfits yield identical values when data and synthetics are exactly time-shifted. This means that for a given datum $u_i^0(t)$, the synthetic is $u_i(t) = u_i^0(t - \Delta t)$. Furthermore, the window function h is required to vary much more slowly than the data and synthetics, and the phase weighting function W_p must be equal to $|\tilde{u}_i^0|/||v_i^0||$. For the time–frequency representation of u_i, we then find

$$\tilde{u}_i(t,\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u_i(\tau) h^*(\tau - t)\, \mathrm{e}^{-\mathbf{i}\omega\tau}\, \mathrm{d}\tau = \frac{\mathrm{e}^{-\mathbf{i}\omega\Delta t}}{\sqrt{2\pi}} \int_{\mathbb{R}} u_i^0(\tau) h^*(\tau - t + \Delta t)\, \mathrm{e}^{-\mathbf{i}\omega\tau}\, \mathrm{d}\tau \approx \mathrm{e}^{-\mathbf{i}\omega\Delta t}\, \tilde{u}_i^0(t,\omega). \tag{40}$$

The phase difference therefore is $\Delta\phi = \phi - \phi_0 = -\omega\Delta t$. Invoking Plancherel's formula allows us to estimate the phase misfit E_p:

$$E_\mathrm{p}^2 = \int_{\mathbb{R}^2} W_\mathrm{p}^2 \omega^2 \Delta t^2\, \mathrm{d}\omega\, \mathrm{d}t = ||v_i^0||^{-2}\, \Delta t^2 \int_{\mathbb{R}^2} |\omega\tilde{u}_i^0|^2\, \mathrm{d}\omega\, \mathrm{d}t \approx ||v_i^0||^{-2}\, \Delta t^2 \int_{\mathbb{R}^2} |\tilde{v}_i^0|^2\, \mathrm{d}\omega\, \mathrm{d}t = ||v_i^0||^{-2}\, \Delta t^2 \int_{\mathbb{R}} |v_i^0|^2\, \mathrm{d}t = \Delta t^2. \tag{41}$$

Hence, under these specific assumptions, the phase misfit is equal to the time-shift Δt, which could also be measured by cross-correlation. The adjoint-source time function for phase misfits also reflects this similarity to cross-correlation time-shift measurements. Introducing $E_\mathrm{p} = \Delta t$ and $W_\mathrm{p} = |\tilde{u}_i^0|/||v_i^0||$ into (36) yields

$$\begin{aligned} s_{\mathrm{p},i}^\dagger(t) &= E_\mathrm{p}^{-1}\, \Im m\, G^{-1}\left[W_\mathrm{p}^2 (\phi_i - \phi_i^0) \frac{\tilde{u}_i}{|\tilde{u}_i|^2} \right](t) = -||v_i^0||_2^{-2} \Im m\, G^{-1}\left[\omega |\tilde{u}_i^0|^2 \frac{\tilde{u}_i}{|\tilde{u}_i|^2} \right](t) \approx -||v_i^0||_2^{-2} \Im m\, G^{-1}\left[\omega\tilde{u}_i\right](t) \\ &\approx \frac{||v_i^0||_2^{-2}}{\sqrt{2\pi}||h||_2} \Im m\, \partial_t \int_{\mathbb{R}^2} \mathbf{i}\tilde{u}_i(\tau,\omega) \mathrm{e}^{\mathbf{i}\omega t}\, h(t-\tau)\, \mathrm{d}\omega\, \mathrm{d}\tau = ||v_i||_2^{-2} \dot{u}_i(t). \end{aligned} \tag{42}$$

A comparison with (C7) shows that $s^\dagger{}_{\mathrm{p},i}$ is identical to the adjoint-source time function $s^\dagger{}_{\mathrm{cc},i}$ corresponding to the measurement of time-shifts, via cross-correlation of data and synthetics. To obtain this, result we needed to assume that the data u_i^0 and the synthetics u_i are related through $u_i(t) = u_i^0(t - \Delta t)$, which means that they need to have identical waveforms. In practice, this condition is rarely met due to dispersion in the heterogeneous Earth with resulting waveform distortions. Consequently, the Fréchet kernels for cross-correlation and phase-misfit measurements will generally be different.

4.2 Envelope misfits and rms amplitude differences

The relative rms amplitude misfit between real and synthetic waveforms is defined through the equations

$$E_\mathrm{rms}^2 = \frac{(A - A^0)^2}{(A^0)^2} = \frac{\Delta(A^0)^2}{A^2}, \quad A = \left(\int_{\mathbb{R}} u_i^2\, \mathrm{d}t \right)^{1/2} = ||u_i||_2, \quad A^0 = \left[\int_{\mathbb{R}} (u_i^0)^2\, \mathrm{d}t \right]^{1/2} = ||u_i^0||_2, \tag{43}$$

where u_i and u_i^0 are assumed to be already windowed to a particular phase or a specific part of the seismogram. Tibuleac *et al.* (2003) demonstrated that E_rms for P waves contains information about the Earth's structure. The corresponding sensitivity kernels have been derived by Dahlen & Baig (2002). The envelope and the rms amplitude misfits are identical under the unrealistic condition that the data u_i^0 and the synthetics u_i are exactly scaled, that is, $u_i = \gamma\, u_i^0$, with a real number γ. Using the envelope weighting function $W_\mathrm{e} = ||u_i^0||_2^{-1}$ then yields $E_\mathrm{rms}^2 = E_\mathrm{e}^2 = \Delta A^2/(A^0)^2$. Consequently, the corresponding adjoint-source time functions $s_{\mathrm{e},i}^\dagger$ and $s_{\mathrm{rms},i}^\dagger$ are also identical:

$$s_{\mathrm{e},i}^\dagger(t) = s_{\mathrm{rms},i}^\dagger(t) = -\frac{\operatorname{sign}\Delta A}{||u_i^0||_2 ||u_i||_2}\, u_i(t). \tag{44}$$

In practice, the assumptions under which E_rms equals E_e are rarely satisfied because of the time dependence of amplitude variations (see Section 5 for examples involving real data). Thus, the corresponding Fréchet kernels will mostly not be identical.

4.3 Generalized seismological data functionals

GSDFs were introduced by Gee & Jordan (1992) as a means of extracting and separating phase and amplitude information from seismic waveforms. They have recently been used by Chen *et al.* (2007a) for the imaging of crustal structure in the Los Angeles region. GSDFs are the formalization of ideas expressed earlier by Lerner-Lam & Jordan (1983) and Cara & Lévêque (1987). The development is based on the concept of an isolation filter $f(t)$, which is a processed version of the synthetic waveform $u(t)$, for example, a windowed and filtered synthetic P wave. The correlation of the isolation filter with the data $u^0(t)$, denoted by $c(u^0, f)$, is first localized in time by windowing and then localized in the frequency domain by narrow-band filtering around a chosen frequency ω_i. The modified correlation function $F_i Wc(u^0, f)$ may then be approximated by a Gaussian wavelet with half-width σ_s^{-1} and centre frequency ω_s:

$$F_i Wc(u^0, f)(t) \approx g(t) = A\, \mathrm{e}^{-\sigma_s^2 (t - t_\mathrm{g}^0)^2/2}\, \cos[\omega_\mathrm{s}(t - t_\mathrm{p}^0)], \tag{45}$$

where the quantities t_p^0 and t_g^0 are interpreted as phase and group delays, respectively. Repeating the same procedure for the correlation of the isolation filter $f(t)$ with the synthetic seismogram $u(t)$ defines the phase and group delays t_p and t_g. Two of the four GSDFs are then defined through

$$\Delta t_\mathrm{p} = t_\mathrm{p} - t_\mathrm{p}^0 \quad \text{and} \quad \Delta t_\mathrm{g} = t_\mathrm{g} - t_\mathrm{g}^0. \tag{46}$$

These are the differential phase and group delays, respectively. We do not consider the remaining two GSDFs. We can relate the phase difference $\Delta\phi$ to the GSDFs $\Delta t_{\rm p}$ and $\Delta t_{\rm g}$ when the isolation filter $f(t)$ is a well-separated phase that arrives at some time $t = T$ and that does not interfere with other phases. The correlation of $f(t)$ with $u(t)$ is then effectively an autocorrelation. This implies $t_{\rm g} = t_{\rm p} = 0$, $\Delta t_{\rm p} = t_{\rm p}^0$ and $\Delta t_{\rm g} = t_{\rm g}^0$. If we furthermore assume that $Wc(u^0, f)$ does not need to be filtered severely to be representable by a Gaussian wavelet, then we have

$$c(f_T^0, f_T)(t) \approx Wc(u^0, f)(t) \approx A\, {\rm e}^{-\sigma_{\rm s}^2 (t - t_{\rm g}^0)^2/2}\, \cos[\omega_{\rm s}(t - t_{\rm p}^0)], \tag{47}$$

where $c(f_T^0, f_T)$ is the correlation of the data and synthetics windowed by $h(t - T)$ (see eq. 4). The phase of $c(f_T^0, f_T)$ is equal to the phase difference $\Delta\phi(T, \omega)$. Since the Fourier transform of the Gaussian wavelet in (45) and (47) is

$$\tilde{g}(\omega) = \frac{\sqrt{2\pi}}{2\sigma_s}\, {\rm e}^{-(\omega_{\rm s} - \omega)/2\sigma_{\rm s}^2}\, {\rm e}^{{\rm i}\omega_{\rm s} t_{\rm p}^0 - {\rm i} t_{\rm g}^0 (\omega_{\rm s} - \omega)}, \tag{48}$$

we can infer the following relation:

$$\Delta\phi(T, \omega) \approx \omega_{\rm s}\, \Delta t_{\rm p} + (\omega - \omega_{\rm s})\, \Delta t_{\rm g}. \tag{49}$$

The phase difference $\Delta\phi$ mixes phase and group delay information. It nevertheless completely quantifies both through $\Delta t_{\rm p} = \Delta\phi(T, \omega = \omega_{\rm s})/\omega_{\rm s}$ and $\Delta t_{\rm g} = \partial\, \Delta\phi(T, \omega)/\partial\omega$.

Conceptually, the time–frequency misfits and the GSDFs are closely related. When measured at a series of frequencies,the GSDFs quantify frequency-dependent phase and amplitude discrepancies. Thus, minimizing the complete collection of GSDFs is similar to jointly minimizing the phase and the envelope misfits. When the data and synthetics are too dissimilar, the processed correlation function $F_i Wc(u^0, f)$ may not resemble a Gaussian wavelet. It is then possible to expand $F_i Wc(u^0, f)$ into a Gram–Charlier series with, at least in theory, infinitely many parameters (Gee & Jordan 1992). In this sense, the GSDFs can be infinite-dimensional and complete along the time axis, just as the time–frequency misfits $E_{\rm p}$ and $E_{\rm e}$.

4.4 The relation of envelope and phase misfits to time-domain full-waveform inversion

Luo & Schuster (1991) established a link between time-domain full-waveform inversion and wave-equation traveltime inversion based on cross-correlation time-shifts (Section 4.1), assuming that small velocity perturbations lead to small time-shifts, time-domain full-waveform inversion and wave-equation traveltime inversion are essentially identical. This suggests that an iterative minimization based on cross-correlation time-shifts might combine the advantages of transmission and diffraction tomography.

We can show a similar link between the phase misfit and the L_2 misfit $||\mathbf{u} - \mathbf{u}^0||_2^2$, used in time-domain full-waveform inversion. By analogy with the approach taken by Luo & Schuster (1991), we assume that the differences between the real Earth and the model earth are so small that $\tilde{u}_i$ and $\tilde{u}_i^0$ only differ by a small phase shift $\Delta\phi_i$. For the phase weighting function $W_{\rm p}$, we choose $W_{\rm p} = |\tilde{u}_i^0|$. A Taylor expansion truncated after the linear term then gives

$$\Delta\tilde{u}_i = \tilde{u}_i - \tilde{u}_i^0 = |\tilde{u}_i^0|\, {\rm e}^{{\rm i}(\phi_i^0 + \Delta\phi_i)} - |\tilde{u}_i^0|\, {\rm e}^{{\rm i}\phi_i^0} = {\rm i}\Delta\phi_i\, \tilde{u}_i^0. \tag{50}$$

By invoking Plancherel's relation, we find that the phase misfit $E_{\rm p}$ is then equal to the L_2 norm of $\Delta u_i = u_i - u_i^0$:

$$E_{\rm p}^2 = \int_{\mathbb{R}^2} W_{\rm p}^2\, \Delta\phi_i^2\, {\rm d}\omega\, {\rm d}t = \int_{\mathbb{R}^2} |\Delta\tilde{u}_i|^2\, {\rm d}\omega\, {\rm d}t = \int_{\mathbb{R}} \Delta u_i^2\, {\rm d}t = ||\Delta u_i||_2^2. \tag{51}$$

Since the misfit measures are equal, the corresponding adjoint sources are also identical. The term 'small' is problematic when it is used to quantify differences between earth models. In fact, when the model differences are 'small', it is equally justifiable that the phase differences vanish and that only the amplitudes of data and synthetics vary by a factor of say $\gamma > 0$, that is, $\tilde{u}_i = \gamma\tilde{u}_i^0$. Choosing $W_{\rm e} = 1$, we find in that case that the envelope misfit $E_{\rm e}$ is equal to the L_2 norm of Δu_i:

$$E_{\rm e}^2 = \int_{\mathbb{R}^2} W_{\rm e}^2 (\gamma - 1)^2 |\tilde{u}_i^0|^2\, {\rm d}\omega\, {\rm d}t = \int_{\mathbb{R}^2} W_{\rm e}^2 |\Delta\tilde{u}_i|^2\, {\rm d}\omega\, {\rm d}t = \int_{\mathbb{R}^2} |\Delta\tilde{u}_i|^2\, {\rm d}\omega\, {\rm d}t = \int_{\mathbb{R}} \Delta u_i^2\, {\rm d}t = ||\Delta u_i||_2^2. \tag{52}$$

In practice, both amplitude and phase differences will be observed even when the model earth is close to the real Earth. It is therefore generally not possible to equate time-domain full-waveform inversion with waveform inversion in the time–frequency domain, when the model differences are deemed small.

5 DATA EXAMPLES

5.1 Earth model and data

To illustrate the computation of sensitivity kernels for time–frequency domain misfits with a real data example, we choose an event that occurred on 1993 June 12 in the West Irian region of Indonesia (see Fig. 3). The earthquake location is latitude: -4.37°; longitude: 135.12°; depth: 15 km. The CMT (Central Moment Tensor) solution is visualized in Fig. 3. Unprocessed velocity seismograms, recorded at station CTAO (latitude: -20.09°; longitude: 146.25°; $\Delta = 19.02^\circ$) are plotted in the left column panel of Fig. 3. Instead of integrating both data and synthetics, we decide to work with velocities rather than with displacements.

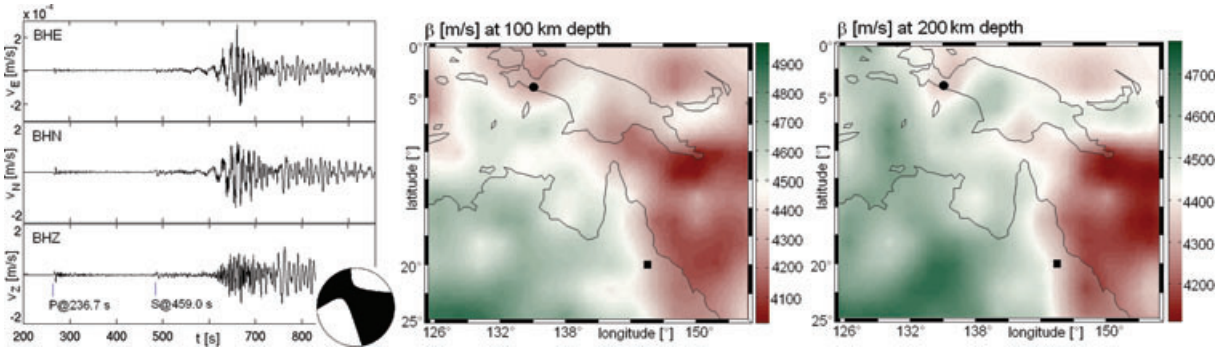

Figure 3. Left-hand panel: unprocessed velocity seismograms of the West Irian event (1993 June 12) recorded at the permanent station CTAO, located in NW Australia. The CMT solution is visualized in the lower right-hand panel of the BHZ channel recording. Central panel: model of the *S*-wave velocity β at the depth of 100 km. The maximum lateral variations $\Delta\beta$ reach 10 per cent of the background value. The source and receiver locations are plotted as a circle (•) and a square (****), respectively. Right-hand panel: the same as in the central one but at the depth of 200 km. The lateral variations are smaller than at 100 km depth and finally vanish below 350 km.

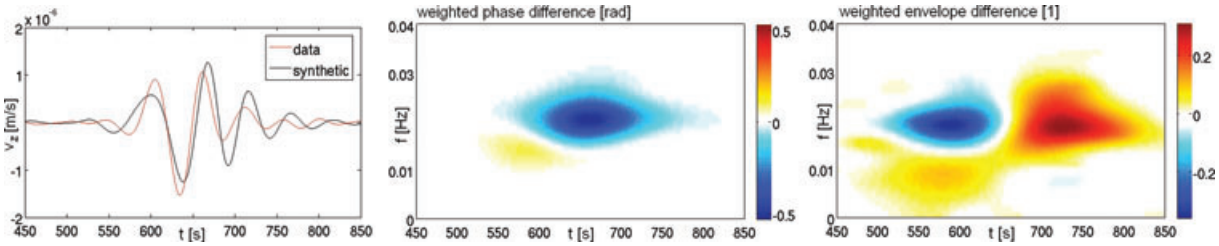

Figure 4. Left-hand panel: comparison of vertical-component surface wave trains, low-pass filtered with a cut-off frequency of 0.02 Hz (50 s). The synthetic is plotted in black and the data in red. Centre panel: weighted phase difference in time–frequency space. Both positive and negative phase differences are observable. Right-hand panel: weighted envelope difference in time–frequency space.

The earth model used to compute the synthetic seismograms is an oversmoothed version of the *S*-velocity model, derived by Fishwick *et al.* (2005) on the basis of a surface wave tomography (Fig. 3). The model shows lateral velocity variations down to 350 km. The 1-D background model is ak135 (Kennett *et al.* 1995). Based on the analysis of refracted waves (Kaiho & Kennett 2000), we set the lateral *P*-wave speed variations equal to 2/3 of the *S*-wave speed variations and disregard any lateral variations in density. This model is intended to be an initial model for a waveform tomography. It reproduces the data sufficiently well and therefore justifies the use of the time–frequency domain misfits E_{e} and E_{p}. Instead of classical tomographies one may alternatively use initial models that are based on geodynamic modelling (Schuberth *et al.* 2008). We computed all synthetic seismograms with a spectral-element method described in Fichtner & Igel (2008) and implemented on a cluster with 160 processors (Oeser *et al.* 2006).

5.2 Surface waves

In the interest of simplicity, we restrict our attention to the vertical component, even though all components should ideally be used in a waveform tomography. The surface wave data and synthetics are low-pass filtered with a cut-off frequency of 0.02 Hz (50 s). A comparison of the data and the synthetic surface wave train are shown on the left-hand panel of Fig. 4.

For the time–frequency analysis, we choose $\sigma = 50$ s, that is, the cut-off frequency. The weighting function for the phase difference is $W_{\mathrm{p}} = \log(1 + |\tilde{v}_z^0|)/\max \log(1 + |\tilde{v}_z^0|)$, and for the envelope difference, we choose $W_{\mathrm{e}} = ||v_z^0||_2^{-1}$. The weighted phase misfit, shown in the centre of Fig. 4, is dominated by negative values that are due to the overall phase delay of the synthetic relative to the data wave train. However, there are also slightly positive values of the phase difference near the onset of the wave train. The weighted envelope difference is more complex than the phase difference, as one can see on the right-hand panel of Fig. 4. For frequencies below 0.01 Hz, the envelope difference is generally positive, meaning that the synthetic is larger than the data. Around the dominant frequency of 0.02 Hz, the envelope misfit agrees well with the visual analysis—smaller amplitudes of the synthetic prior to 650 s followed by larger amplitudes of the synthetic that persist until around 800 s.

The phase and envelope differences translate to adjoint-source time functions (eq. 37) that we use to compute sensitivity kernels for E_{p} and E_{e} with respect to the shear wave speed β. The results are shown in the left columns of Figs 5 and 6, respectively. Effectively, the kernels contain contributions from the entire passband up to 50 s. The complexity of both kernels is due to (1) the complexity of the phase and envelope differences in time–frequency space, (2) the 3-D structure of the earth model and (3) the radiation pattern of the source. While both kernels attain comparatively large values in the vicinity of the ray path, pronounced side lobes also appear.

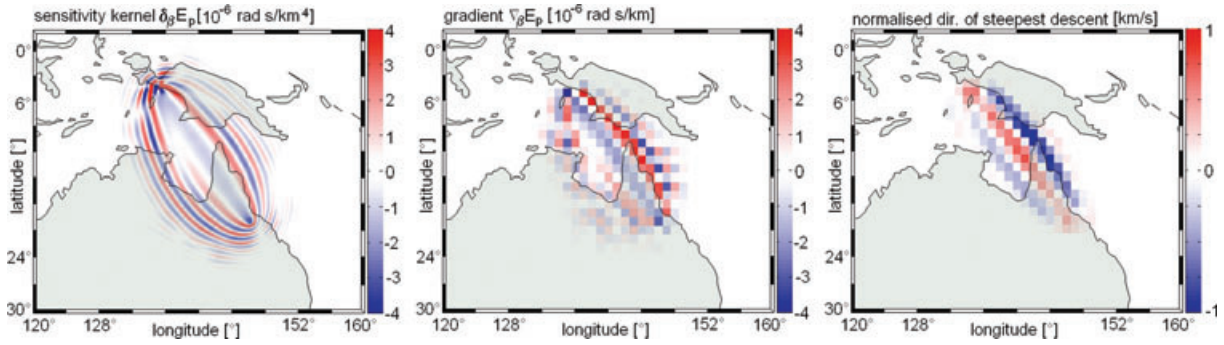

Figure 5. Left-hand panel: horizontal slice through the sensitivity kernel $\delta_\beta\ E_\mathrm{p}$ at the surface. The kernel corresponds to the phase difference shown in the centre of Fig. 4. It, therefore, contains contributions from the entire passband up to 50 s. Central panel: horizontal slice through the gradient of the phase misfit E_p with respect to the shear wave speed. The model basis functions are blocks that are $1^\circ \times 1^\circ$ wide and 10 km deep. Right-hand panel: normalized steepest descent direction obtained by multiplying the negative gradient by a covariance matrix. The covariance matrix introduces a horizontal correlation length of 100 km and a vertical correlation length of 20 km.

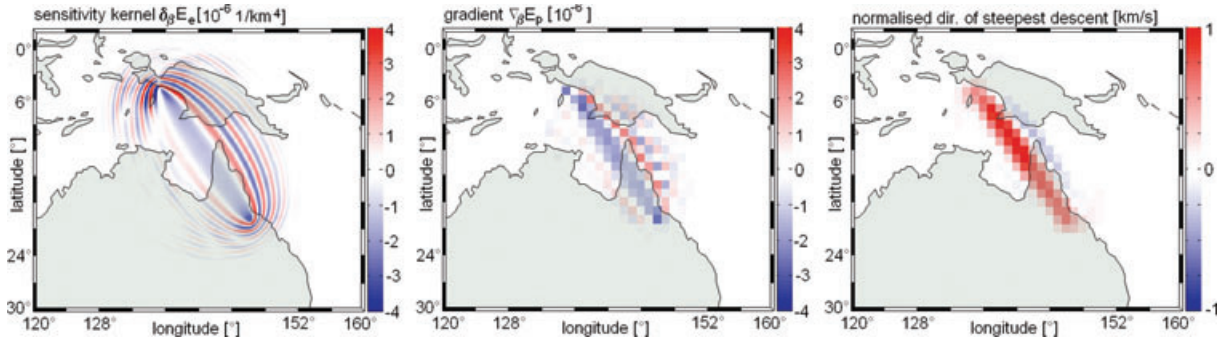

Figure 6. The same as Fig. 5 but for the envelope misfit E_e.

The significance of the side lobes in the context of a gradient-method-based misfit minimization can be evaluated by computing the gradients and the steepest descent directions that correspond to the different kernels. The gradient is the projection of the kernel onto the space of model basis functions, and the direction of steepest descent equals the negative gradient multiplied by a covariance matrix. Choosing, for the purpose of illustration, the model basis functions to be blocks that are $1^\circ \times 1^\circ$ wide and 10 km deep produces the gradients shown in the centres of Figs 5 and 6. The side lobes with an oscillation period of 1° and less disappear so that the gradients are dominated by the two central lobes. Multiplying the negative gradients by a covariance matrix that introduces a horizontal correlation length of 100 km and a vertical correlation length of 20 km yields the steepest descent directions shown in the right columns of Figs 5 and 6. The covariance matrix acts as a smoothing operator and therefore removes oscillations at length scales smaller than the correlation length. In this particular example, only the two central lobes of the sensitivity kernels are relevant for an iterative misfit minimization based on gradient methods. This is, however, not a general statement because the characteristics of the steepest descent direction depend strongly on the actual waveform misfits, the set of basis functions and the regularization via the covariance matrix.

The Fréchet kernels displayed in Figs 5 and 6 cannot be compared directly to those derived, for example, by Marquering *et al.* (1998), Friederich (1999), Zhou *et al.* (2004) or Yoshizawa & Kennett (2005), on the basis of semi-analytic solutions of the elastic wave equation. The reason for the lack of comparability arises mostly from the different measurement techniques, different notions of phase and amplitude and the intrinsic data dependence of our kernels. Nevertheless, the nature of wave propagation ensures that all kernels are qualitatively similar in shape, regardless of the measurement details—slowly oscillating structural sensitivity around the geometrical ray path corresponding to the first Fresnel zone and rapidly oscillating off-path sensitivity corresponding to the higher Fresnel zones.

5.3 Body waves

We consider, as an example, vertical-component *P* and *S* waveforms. The unfiltered data are shown in the centres of Figs 9 and 10. At the epicentral distance of 19°, the *P* waveform is composed of various phases associated with reflection and refraction from the 410 and 660 km discontinuities. The surface-reflected phase *PnPn* does not yet have a clearly separate identity but appears as a lower-frequency tail of the composite *P*-wave train. Matters are more complicated in the case of the *S*-wave train. In addition to upper-mantle reflections and refractions, there are also higher-mode surface waves interacting with the direct *S* wave.

For our analysis, we have applied a low-pass filter to the waveforms with a cut-off frequency of 0.04 Hz (25 s). The time-domain *P* and *S* waveforms, together with their respective phase and envelope misfits, are displayed in Figs 7 and 8. The weighting functions W_e and W_p

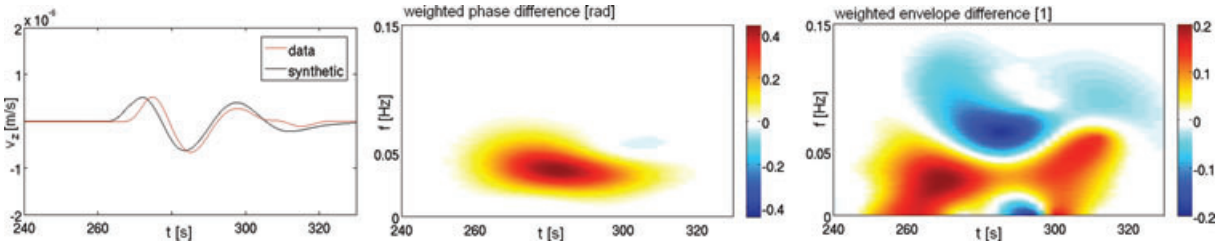

Figure 7. Left-hand panel: comparison of vertical-component *P*-wave trains, low-pass filtered with a cut-off frequency of 0.04 Hz (25 s). The synthetic is plotted in black and the data in red. Central panel: weighted phase difference in time–frequency space. Right-hand panel: weighted envelope difference in time–frequency space.

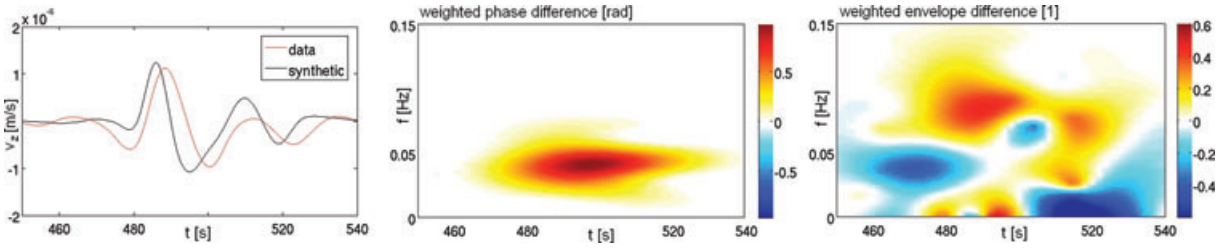

Figure 8. The same as Fig. 7 but for the *S*-wave train.

are the same as in the section on surface waves. In both cases, the data are delayed with respect to the synthetics for all times and frequencies, therefore producing generally positive phase differences (centres of Figs 7 and 8). This is in agreement with the visual impression. The characteristics of the envelope misfits, shown on the right-hand panel of Figs 7 and 8, are more complicated.

5.3.1 P-wave Fréchet kernels

The Fréchet kernels for the *P* wave with respect to the *P*-wave speed α can be seen in Fig. 9. They contain contributions from the complete passband up to 25 s. Both the envelope misfit and the phase-misfit kernels show a broad central zone surrounding the geometrical ray path. The outer lobes are not as pronounced as in the surface wave case. The kernels are comparatively simple because the *P* waveform is a well-pronounced single peak, despite the interference of several phases inside the analysed time window. The minimum along the ray path is reminiscent of the seemingly paradoxical zero observed in other types of phase delay kernels (Woodward 1992; Yomogida 1992; Marquering *et al.* 1999). It can be interpreted in terms of in-phase scattering on the ray path.

The filtered data and synthetic *P* waveforms shown on the left-hand panel of Fig. 8 are similar, though not identical. In consequence, from our comparison of misfit functionals in Section 4, the phase-misfit kernel should be qualitatively similar to cross-correlation time delay kernels (Dahlen *et al.* 2000; Liu & Tromp 2008). Similarly, the envelope misfit kernel should qualitatively resemble the rms amplitude kernels of Dahlen & Baig (2002). Fig. 9 confirms this conjecture.

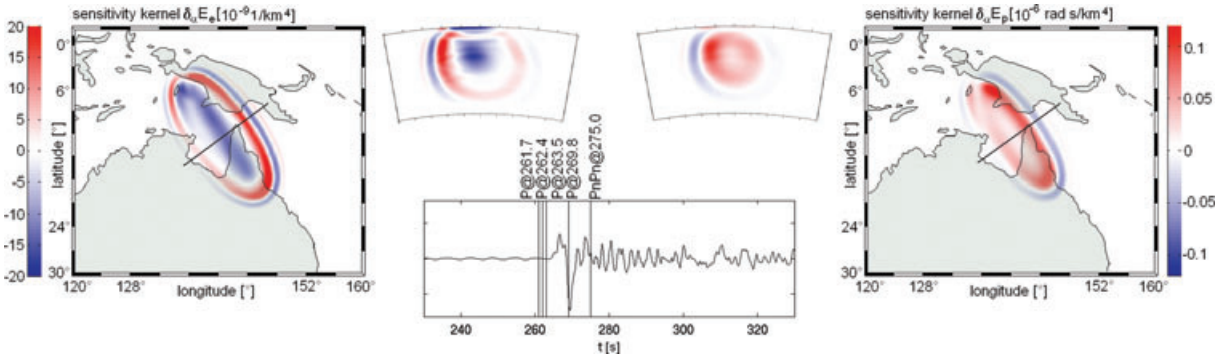

Figure 9. Left-/central left-hand panel: horizontal slice at 371 km depth and vertical slice through the envelope-misfit kernel for the *P*-wave train. Centre panel: unprocessed vertical-component velocity seismogram comprising the *P* arrival. Right/centre right-hand panel: horizontal slice at 371 km depth and vertical slice through the phase-misfit kernel for the *P*-wave train. The kernels in this figure correspond to the envelope and phase differences shown in Fig. 7. They contain contributions from the entire passband up to 25 s.

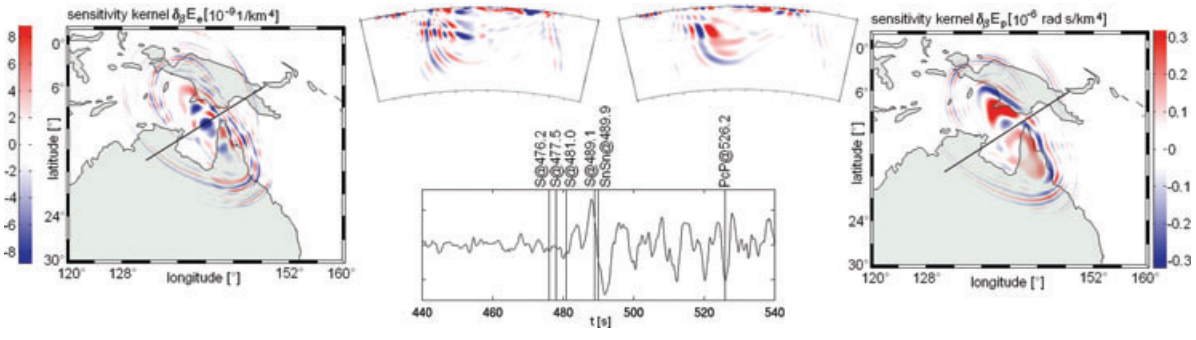

Figure 10. The same as Fig. 10 but for the *S*-wave train.

5.3.2 S-wave Fréchet kernels

Fig. 10 displays the Fréchet kernels corresponding to the *S* waveform. The kernels for both the phase and the envelope misfit differ largely from the *P*-wave kernels. The complexity of the *S*-wave kernels results from the complexity of the *S*-wave train (see Figs 8 and 10) inside the considered time window. The earth model ak135 (Kennett *et al.* 1995) predicts four distinct *S* phases, the *SnSn* phase and the *PcP* phase to arrive between 470 and 530 s (Knapmeyer 2004). The situation is further complicated by the presence of higher-mode surface waves. At periods around 25 s, these phases interfere with the *S*-wave train and cannot be separately distinguished. Consequently the sensitivity kernels become a superposition of the kernels corresponding to each single phase plus the higher-mode surface waves. The *P*-wave train between 260 and 280 s is also composed of several phases including *PnPn*. Nevertheless, the *P* waveform (see Figs 7 and 9) is clearly dominated by one single peak, which leads to a comparatively simple sensitivity kernel.

Both the complex interferences and the dissimilarity of data and synthetics add to the fact that the phase-misfit kernel shown in Fig. 10 does not resemble cross-correlation time delay kernels (Dahlen *et al.* 2000; Liu & Tromp 2008). Also, the envelope-misfit kernel differs from rms amplitude kernels (Dahlen & Baig 2002). This reflects the fact that different measurement techniques yield different kernels.

We abstain from a more detailed analysis of the sensitivity kernels because most statements would not be general due to the strong dependence of the kernels on the actual data, the model earth and the source characteristics.

6 DISCUSSION

6.1 Advantages of the time–frequency domain misfits in the context of structural inverse problems

The principal advantage of the time–frequency domain misfits is their ability to extract the maximum amount of waveform information and to separate this information into two parts: the phase which is quasi-linearly related to Earth's structure and the envelope which can be highly non-linearly related to Earth's structure. This makes the time–frequency misfits usable in continental- and global-scale tomography, that is, in a scenario where the seismic wavefield is spatially undersampled and where the background structure is insufficiently well known.

The computation of the time–frequency misfits does not require the isolation of particular phases or the extraction of any secondary observables. It is applicable to body waves, surface waves and interfering wave trains. Seismic phases that are of specific interest, for example, small-amplitude core phases, can be emphasized through the weighting functions W_p and W_e.

The phase and envelope misfits are physically meaningful quantifications of seismic waveform differences. This is not always the case when the time-domain L_2 norm $\sum_{r=1}^{N} \int [u(\mathbf{x}^r, t) - u^0(\mathbf{x}^r, t)]^2 \, \mathrm{d}t$ is used instead. The latter can be interpreted as the energy of the scattered wavefield, however, only when the remaining structural heterogeneities can indeed be treated as scatterers. The time-domain L_2 norm can be large even when data and synthetics are similar but slightly time-shifted. This contributes to the excessive non-linearity of time-domain full-waveform inversion. Phase and envelope misfits avoid this deficiency.

The phase and envelope misfits together fully quantify waveform differences, that is, data = synthetics in the hypothetical case of zero phase and zero envelope misfit. In this sense, the maximum amount of waveform information is extracted. A structural inverse problem based on both time–frequency misfits is a full-waveform inversion.

6.2 Disadvantages

The time–frequency domain misfits E_p and E_e share one disadvantage with all other measures of seismic waveform differences; they are meaningful only when data and synthetics are similar to some degree. There is currently no technique that allows us to compare largely dissimilar waveforms in a physically meaningful way. This may be due to the practical need to linearize tomographic problems, but also to our insufficient understanding of wave propagation in complex media and our rather inflexible physical intuition (In fact, linearizability and physical intuition often seem to coincide). Our intuition favours a comparison of waveforms by associating oscillation cycles in the data with oscillation cycles in the synthetics. This is problematic when data and synthetics are too dissimilar, that is, when there are no cycles to be associated. In the context of a structural inverse problem, this difficulty can be circumvented by inverting for lower frequencies first

and then by including successively higher frequencies. The general success of this strategy seems obvious but, strictly speaking, is based on conjecture. When data and synthetic waveforms are too dissimilar only Monte Carlo methods will be generally successful. The choice of the misfit functional is then likely to be of lesser importance.

As other measures of full-waveform differences, the phase and envelope misfits are relatively sensitive to noise and modelling errors. The influence of noise can be reduced by using the weighting functions W_{p} and W_{e} as filters that suppress frequency bands where the noise is particularly high. Still, full-waveform inversion, in any of its variants, is a tomographic method that relies on high-quality data and that is unlikely to be efficient when the quality of the recordings is too low. Modelling errors can be minimized by using sophisticated numerical methods that correctly account for 3-D Earth's structure and the physics of wave propagation.

6.3 The wavelet transform and time-adaptive window functions

A disadvantage of the time–frequency transforms (Gabor transforms) as defined in (1) is that the width of the sliding window h does not depend on time or frequency. In some applications, such dependence might be desirable. In the case of a strongly dispersed surface wave train, for example, one may wish to sample the longer-period part with a broader window and the late-arriving part with a narrow window. More generally, h should be broad for low frequencies and comparatively narrow for high frequencies.

This well-known drawback of the Gabor transform is the principal motivation for its replacement by a continuous wavelet transform (*CWT*), studied for example by Kristeková *et al.* (2006). Indeed, *CWT* with the Morlet wavelet ψ at reference frequency ω_0, defined through

$$CWT[u(\mathbf{x}_i^r)](\omega, t) := \frac{1}{\sqrt{|\omega_0/\omega|}} \int_{\mathbb{R}} u(\mathbf{x}^r, t)\, \psi^* \left(\frac{\tau - t}{\omega_0/\omega} \right) \mathrm{d}\tau\,, \qquad \psi(t) := \pi^{-1/4}\, \mathrm{e}^{\mathrm{i}\omega_0 t}\, \mathrm{e}^{-t^2/2}, \tag{53}$$

would solve this problem while being conceptually very close to the windowed Fourier transform. Due to this similarity, most of the results derived in the previous sections remain almost unchanged when F_h is replaced by *CWT*. Hence, if a particular data set requires a more sophisticated time–frequency transform, pre-existing data analysis codes can be modified with ease.

In our case, where the data are body and surface wave trains, recorded at epicentral distances beyond 15°, we found the Gabor transform to be sufficient. The reason for this is that we can compare data and synthetics only in a frequency range where they are close, that is, phase shifted by less than $\pm\pi$. This is, unfortunately, true for comparatively low frequencies only, even when a 3-D earth model is used for the computation of the synthetics. Therefore, despite improving the quality of the time–frequency representations, a *CWT* does not necessarily improve the quantification of the misfits that one can use in a seismic waveform tomography.

6.4 Computational aspects

Eq. (15) implies that the regular wavefield $\mathbf{u}\,(\mathbf{x}, t)$ and the adjoint wavefield $\mathbf{u}^\dagger(\mathbf{x}, t)$ need to be known simultaneously to compute the sensitivity kernel via the time integral $\delta_m E = \int_{\mathbb{R}} \mathbf{u}^\dagger \cdot \partial_m \mathbf{L}(\mathbf{u}, \mathbf{m})\, \mathrm{d}t$. In practice $\delta_m E$ is computed during the solution of the adjoint equation, which runs backwards in time due to the terminal conditions (16b). The regular wavefield is then made available through one of the following three approaches. (1) The final state of the regular wavefield, $\mathbf{u}\,(\mathbf{x}, t_1)$ is stored and then marched backwards in time, together with the adjoint wavefield. This is possible when the medium is non-dissipative. (2) The regular wavefield is stored at sufficiently many time steps during the solution of the regular wave equation and then loaded during the solution of the adjoint equation. This method is applicable in the case of dissipative media. It is time-efficient but has very high storage requirements. (3) Checkpointing algorithms provide a balance between storage and time efficiency (e.g. Griewank & Walther 2000; Charpentier 2001). The regular wavefield is stored at a smaller number of time steps, called checkpoints, and solved from there until the current time of the adjoint calculation is reached. For the sensitivity kernels presented in this paper, we used approach (2), that is, the storage of the complete regular wavefield.

Although the computation of the sensitivity kernels is conceptually simple, a full-waveform tomography based on the adjoint method is still a major challenge. Sensitivity kernels need to be computed for a large number of data and several iterations may be necessary to achieve a satisfactory misfit reduction. The computational costs can be reduced by computing composite kernels for all data corresponding to one source (Tape *et al.* 2007). When the number of sources outnumbers the number of receivers, the scattering-integral formulation by Chen *et al.* (2007a,b) can be more efficient than the adjoint method.

6.5 A strategy for full-waveform inversion using time–frequency misfits

A time–frequency domain full-waveform inversion is likely to be most efficient when the phase and envelope misfits are minimized iteratively, for example, with a conjugate gradient method. During the first iterations, the emphasis should be on the phase misfit of longer-period waveforms to keep the problem as linear as possible. The envelope misfit can then be upweighted successively, depending on the already achieved waveform similarity and the data quality. With increasing number of iterations, higher frequencies can then be taken into account. Under which circumstances the envelope misfit may be used to discern variations of Q from elastic structure, still requires further investigations.

ACKNOWLEDGMENTS

The authors would like to thank the Alexander von Humboldt Foundation, the German Academic Exchange Service (DAAD) and the Bavarian Elite Network (ENB) for their support. Discussions with Yann Capdeville, Jeannot Trampert, Peter Moczo, Jozef Kristek, Miriam Kristeková, Qinya Liu and Li Zhao helped us to improve the manuscript. Not a single numerical calculation would have been possible without the help of Jens Oeser.

REFERENCES

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

Boschi, L., 2006. Global multiresolution models of surface wave propagation: comparing equivalently regularized Born and ray theoretical solutions, *Geophys. J. Int.,* **167,** 238–252.

Bunge, H.-P., Hagelberg, C.R. & Travis, B.J., 2003. Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.,* **152,** 280–301.

Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.,* **162,** 541–554.

Cara, M., 1973. Filtering of dispersed wave trains, *Geophys. J. R. astr. Soc.,* **33,** 65–80.

Cara, M. & Lévêque, J.J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.,* **14**(10), 1046–1049.

Charpentier, I., 2001. Checkpointing schemes for adjoint codes: application to the meteorological model Meso-NH, *SIAM J. Sci. Comput.,* **22**(6), 2135–2151.

Chavent, G., Dupuy, M. & Lemonnier, P., 1975. History matching by use of optimal theory, *J. Soc. Petrol. Eng.,* **15**(1), 74–86.

Chen, P., Zhao, L. & Jordan, T.H., 2007a. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97**(4), 1094–1120.

Chen, P., Jordan, T.H. & Zhao, L., 2007b. Full three-dimensional tomography: a comparison between the scattering-integral and the adjoint-wavefield methods, *Geophys. J. Int.,* **170,** 175–181.

Dahlen, F.A. & Baig, A.M., 2002. Fréchet kernels for body-wave amplitudes, *Geophys. J. Int.,* **150,** 440–466.

Dahlen, F.A., Hung, S.H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes, I: theory, *Geophys. J. Int.,* **141**(1), 157–174.

de la Puente, J., Käser, M., Dumbser, M. & Igel, H., 2007. An arbitrary high order discontinuous method for elastic waves on unstructured meshes, IV: anisotropy, *Geophys. J. Int.,* **169**(3), 1210–1228.

Dessa, J.-X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J. & Kaneda, Y., 2004. Multiscale seismic imaging of the eastern Nankai trough by full waveform inversion, *Geophys. Res. Lett.,* **31,** doi:10.1029/2004GL020453.

Devaney, A.J., 1984. Geophysical diffraction tomography, *IEEE Trans. Geosci. Remote Sens.,* **22**(1), 3–13.

Dumbser, M., Käser, M. & de la Puente, J., 2007. Arbitrary high-order finite volume schemes for seismic wave propagation on unstructured meshes in 2D and 3D, *Geophys. J. Int.,* **171**(2), 665–694.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Int.,* **25,** 297–356.

Dziewonski, A.M., Mills, J. & Bloch, S., 1972. Residual dispersion measurement—a new method of surface wave analysis, *Bull. seism. Soc. Am.,* **62**(1), 129–139.

Fichtner, A. & Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models—a technique based on non-linear dispersion curve matching (DCM), *Geophys. J. Int.,* **57**(6), 475–482.

Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology, I: theory, *Phys. Earth planet. Int.,* **157,** 86–104.

Fishwick, S., Kennett, B.L.N. & Reading, A.M., 2005. Contrasts in lithospheric structure within the Australian craton—insights from surface wave tomography, *Earth planet. Sci. Lett.,* **231,** 163–176.

Friederich, W., 1999. Propagation of seismic shear and surface waves in a laterally heterogeneous mantle by multiple forward scattering, *Geophys. J. Int.,* **136,** 180–204.

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

Gee, L.S. & Jordan, T.H., 1992. Generalized seismological data functionals, *Geophys. J. Int.,* **111,** 363–390.

Griewank, A. & Walther, A., 2000. Revolve: an implementation of checkpointing for the reverse or adjoint mode of computational differentiation, *Trans. Math. Software,* **26,** 19–45.

Igel, H., Djikpesse, H. & Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.,* **124**(2), 363–371.

Kaiho, Y. & Kennett, B.L.N., 2000. Three-dimensional seismic structure beneath the Australasian region from refracted wave observations, *Geophys. J. Int.,* **142,** 651–668.

Kennett, B.L.N., 1998. On the density distribution within the Earth, *Geophys. J. Int.,* **132**(2), 374–382.

Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.,* **122,** 108–124.

Knapmeyer, M., 2004. TTBOX: a Matlab tool for the computation of 1D teleseismic traveltimes, *Seism. Res. Lett.,* **75**(6), 726–733.

Komatisch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation, I: validation, *Geophys. J. Int.,* **149**(2), 390–412.

Kristeková, M., Kristek, J., Moczo, P. & Day, S.M., 2006. Misfit criteria for quantitative comparison of seismograms, *Bull. seism. Soc. Am.,* **96**(5), 1836–1850.

Lerner-Lam, A.L. & Jordan, T.H., 1983. Earth structure from fundamental and higher-mode waveform analysis, *Geophys. J. R. astr. Soc.,* **75,** 759–797.

Lions, J.L., 1968. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, Dunod Gauthier-Villars.

Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.,* **174**(1), 265–286.

Luo, Y. & Schuster, G.T., 1991. Wave equation traveltime inversion, *Geophysics,* **56**(5), 645–653.

Marquering, H., Nolet, G. & Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.,* **132,** 521–534.

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

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics,* **52,** 1211–1228.

Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics,* **53**(6), 750–759.

Mora, P., 1989. Inversion=migration+tomography, *Geophysics,* **54,** 1575–1586.

Morelli, A. & Dziewonski, A.M., 1993. Body wave traveltimes and a spherically symmetric P- and S-wave velocity model, *Geophys. J. Int.,* **122,** 178–194.

Nyman, D.C. & Landisman, M., 1977. The display-equalized filter for frequency-time analysis, *Bull. seism. Soc. Am.,* **67**(2), 393–404.

Oeser, J., Bunge, H.-P. & Mohr, M., 2006. Cluster design in the Earth sciences: Tethys, *High Perform. Comput. Commun.,* Proceedings, **4208,** 31–40.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model, *Geophysics,* **64**(3), 888–901.

Press, F., 1968. Earth models obtained by Monte Carlo inversion, *J. geophys. Res.,* **73**(16), 5223–5234.

Schuberth, B., Bunge, H.-P., Steinle-Neumann, G., Moder, Ch. & Oeser, J., 2008. Thermal vs. elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: a case for strong core heating, *Geochem. Geophys. Geosyst.,* submitted.

Seriani, G., 1998. 3-D large-scale wave propagation modeling by spectral element method on CRAY T3E multiprocessor, *Comput. Methods Appl. Mech. Eng.*, **164**(1–2), 235–247.

Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007a. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.,* **168**(3), 1153–1174.

Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007b. Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods, *Geophys. J. Int.,* **171**(1), 368–389.

Spetzler, J., Trampert, J. & Snieder, R., 2001. Are we exceeding the limits of the great circle approximation in global surface wave tomography?, *Geophys. Res. Lett.,* **28**(12), 2341–2344.

Strang, G. & Nguyen, T., 1996. *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, USA.

Sun, N.-Z., 1994. *Inverse Problems in Ground Water Modelling*, Kluwer Academic Publishers, Dordrecht, the Netherlands.

Talagrand, O. & Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation, I: theory, *Q. J. R. Meteorol. Soc.,* **113,** 1311–1328.

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

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

Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.,* **128,** 365–399.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.

Thomson, D.J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE,* **70,** 1055–1096.

Tibuleac, I.M., Nolet, G., Michaelson, C. & Koulakov, I., 2003. P wave amplitudes in a 3-D earth, *Geophys. J. Int.,* **155**(1), 1–10.

Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science,* **306,** 853–856.

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

Williamson, P.R., 1991. A guide to the limits of resolution imposed by scattering in ray tomography, *Geophysics,* **56**(2), 202–207.

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

Woodward, M.J., 1992. Wave-equation tomography, *Geophysics,* **57**(1), 15–26.

Wu, R.-S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics,* **52**(1), 11–25.

Yomogida, K., 1992. Fresnel zone inversion for lateral heterogeneities in the Earth, *Pure appl. Geophys.,* **138**(3), 391–406.

Yoshizawa, K. & Kennett, B.L.N., 2004. Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects, *J. geophys. Res.,* **109,** doi:10.1029/2002JB002254.

Yoshizawa, K. & Kennett, B.L.N., 2005. Sensitivity kernels for finite-frequency surface waves, *Geophys. J. Int.,* **162,** 910–926.

Zhou, C.X., Cai, W.L., Luo, Y., Schuster, G.T. & Hassanzadeh, S., 1995. Acoustic wave-equation travel-time and wave-form inversion of crosshole seismic data, *Geophysics,* **60**(3), 765–773.

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

Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.,* **158,** 142–168.

Zhou, Y., Dahlen, F.A., Nolet, G. & Laske, G., 2005. Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.,* **163,** 1087–1111.

APPENDIX A: THE RELATIONS FROM SECTION 3

We demonstrate the relations needed for the derivations of the adjoint source functions in Section 3. First, we consider the sensitivity density of the time–frequency domain signal $\tilde{u}_i$ with respect to the model parameters, $\delta_p \tilde{u}_i$. For the sensitivity density of the time-domain signal $u_i(\tau)$, we have

$$\delta_p u_i(\mathbf{x}^r, \tau) = -\int_{\mathbb{R}} \mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t') \cdot \partial_p \mathbf{L}[\mathbf{u}(t')]\, \mathrm{d}t', \tag{A1}$$

which is essentially a repetition of eq. (15). Then based on the definition of $\tilde{u}_i$, eq. (1), we find

$$\delta_p \tilde{u}_i(\mathbf{x}^r, t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \delta_p u_i(\mathbf{x}^r, \tau) h^*(\tau - t)\, \mathrm{e}^{-\mathbf{i}\omega\tau}\, \mathrm{d}\tau. \tag{A2}$$

Introducing (A1) into (A2) yields the desired result

$$\delta_p \tilde{u}_i(\mathbf{x}^r, t, \omega) = -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^2} \mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t') \cdot \partial_p \mathbf{L}[\mathbf{u}(t')]\, h^*(\tau - t)\, \mathrm{e}^{-\mathbf{i}\omega\tau}\, \mathrm{d}\tau\, \mathrm{d}t'. \tag{A3}$$

Now, we try to find an expression for $D_p|\tilde{u}_i|$ in which we can make use of (A3). For this expression, we differentiate $|\tilde{u}_i|^2$ with respect to the model parameters p, omitting in the notation the differentiation direction, as usual.

$$D_p|\tilde{u}_i|^2 = D_p(\tilde{u}_i \tilde{u}_i^*) = 2|\tilde{u}_i|\, D_p|\tilde{u}_i| = \tilde{u}_i\, D_p \tilde{u}_i^* + \tilde{u}_i^*\, D_p \tilde{u}_i. \tag{A4}$$

Solving for $D_p|\tilde{u}_i|$ yields

$$D_p|\tilde{u}_i| = \frac{1}{2}\left(\frac{\tilde{u}_i}{|\tilde{u}_i|}\, D_p \tilde{u}_i^* + \frac{\tilde{u}_i^*}{|\tilde{u}_i|}\, D_p \tilde{u}_i\right) = \Re e\left(\frac{\tilde{u}_i}{|\tilde{u}_i|}\, D_p \tilde{u}_i^*\right). \tag{A5}$$

APPENDIX B: SOME RESULTS FROM TIME–FREQUENCY ANALYSIS

We will briefly review some of the principal definitions and results of time–frequency analysis. Most of the proofs are not entirely mathematically rigorous to keep the treatment readable. Throughout the text we work with the following definition of the Fourier transform of a function f

$$\tilde{f}(\omega) = F[f](\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}t. \tag{B1}$$

The corresponding inverse Fourier transform is

$$f(t) = F^{-1}[\tilde{f}](t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \tilde{f}(\omega) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}\omega. \tag{B2}$$

For two functions f and g, we obtain Parseval's relation by combining eqs (B1) and (B2):

$$\begin{aligned}(\tilde{f}, \tilde{g}) &= \int_{\mathbb{R}} \tilde{f}(\omega) \tilde{g}^*(\omega) \, \mathrm{d}\omega = \frac{1}{2\pi} \int_{\mathbb{R}} \left[\int_{\mathbb{R}} f(t) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}t \right] \left[\int_{\mathbb{R}} g^*(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} \right] = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} f(t) g^*(\tau) \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\omega(\tau - t)} \, \mathrm{d}\omega \, \mathrm{d}t \, \mathrm{d}\tau \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(t) g^*(\tau) \, \delta(\tau - t) \, \mathrm{d}\tau \, \mathrm{d}t = \int_{\mathbb{R}} f(t) g^*(t) \, \mathrm{d}t = (f, g).\end{aligned} \tag{B3}$$

Plancherel's formula follows immediately by setting $f = g$: $||\tilde{f}||_2 = ||f||_2$. Now, we define the windowed Fourier transform of a function f as the regular Fourier transform of $f(y)h^*(y - t)$, where h is a sliding window. In symbols

$$\tilde{f}_h(t, \omega) = F_h[f](t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\tau) h^*(\tau - t) \mathrm{e}^{-\mathrm{i}\omega\tau} \, \mathrm{d}\tau. \tag{B4}$$

By defining a time-shift operator T_t through $T_t h(\tau) = h(\tau - t)$, we can express F_h in terms of F, $F_h[f](t, \omega) = F[f T_t h^*](\omega)$. Making use of Parseval's relation for the Fourier transform (B3), we can then derive a similar result for the windowed Fourier transform:

$$\begin{aligned}||h||_2^{-2}(\tilde{f}_h, \tilde{g}_h) &= ||h||_2^{-2} \int_{\mathbb{R}^2} F_h[f](t, \omega) F_h^*[g](t, \omega) \, \mathrm{d}t \, \mathrm{d}\omega = ||h||_2^{-2} \int_{\mathbb{R}^2} F[f T_t h^*](\omega) F^*[g T_t h^*](\omega) \, \mathrm{d}t \, \mathrm{d}\omega \\ &= ||h||_2^{-2} \int_{\mathbb{R}^2} f(\tau) h^*(\tau - t) g^*(\tau) h(\tau - t) \, \mathrm{d}\tau, \mathrm{d}t \\ &= ||h||_2^{-2} \int_{\mathbb{R}^2} f(\tau) g^*(\tau) |h(\tau - t)|^2 \, \mathrm{d}t \, \mathrm{d}\tau = \int_{\mathbb{R}} f(\tau) g^*(\tau) \, \mathrm{d}\tau = (f, g).\end{aligned} \tag{B5}$$

Setting $f = g$ gives an analogue of Plancherel's formula: $||h||_2^{-1} ||\tilde{f}_h||_2 = ||f||_2$. From (B5), we can derive an expression for the inverse of the windowed Fourier transform. For this, we write

$$\begin{aligned}\textstyle\int_{\mathbb{R}} f(t) g^*(t) \, \mathrm{d}t &= ||h||_2^{-2} \int_{\mathbb{R}^2} F_h[f](\tau, \omega) F_h^*[g](\tau, \omega) \, \mathrm{d}\tau \, \mathrm{d}\omega \\ &= \frac{1}{\sqrt{2\pi}||h||_2} \int_{\mathbb{R}^2} F_h[f](\tau, \omega) \int_{\mathbb{R}} g^*(t) h(t - \tau) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t \, \mathrm{d}\tau \, \mathrm{d}\omega \\ &= \frac{1}{\sqrt{2\pi}||h||_2} \int_{\mathbb{R}} \left[\int_{\mathbb{R}^2} F_h[f](\tau, \omega) h(t - \tau) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}\tau \, \mathrm{d}\omega \right] g^*(t) \, \mathrm{d}t.\end{aligned} \tag{B6}$$

Since g can be any function, we deduce

$$f(t) = F_h^{-1}[\tilde{f}_h](t) = \frac{1}{\sqrt{2\pi}||h||_2} \int_{\mathbb{R}^2} F_h[f](\tau, \omega) h(t - \tau) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}\omega \, \mathrm{d}\tau. \tag{B7}$$

Another interesting result can be derived by invoking Parseval's relation. Defining $g_{\omega,t}(\tau) = h(\tau - t) \, \mathrm{e}^{i\omega\tau}$ gives

$$F_h[f](t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\tau) g^*_{\omega,t}(\tau) \, \mathrm{d}\tau = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \tilde{f}(\nu) \tilde{g}^*_{\omega,t}(\nu) \, \mathrm{d}\nu. \tag{B8}$$

For the interpretation of (B8), we consider a fixed point (ω, t) in the time–frequency space. The time–frequency representation $F_h[f]$ can now be generated in two complimentary ways: (1) by integrating over the time representation $f(\tau)$ multiplied by the time window $g^*_{\omega,t}(\tau)$, shifted by t or (2) by integrating over the frequency representation $\tilde{f}(\nu)$ multiplied by the frequency window $\tilde{g}^*_{\omega,t}(\nu)$ shifted by ω. A narrow time window, that is, a high time resolution, will usually lead to a broad frequency window and therefore to a low frequency resolution. Analogously, a low time resolution will result in a high frequency resolution. This trade-off depends strongly on the choice of the sliding time window h in the definition of F_h (B4). The trade-off can be quantified with the well-known uncertainty principle, which we state here without proof:

$$\Delta_{\tilde{g}} \Delta_g \geq \frac{1}{2} ||g_{\omega,t}||_2 = \frac{1}{2} ||h||_2, \tag{B9}$$

The symbols $\Delta_{\tilde{g}}$ and Δ_g denote the variances of the frequency window $\tilde{g}_{\omega,t}$ and the time window $g_{\omega,t}$, respectively. An effective window function h should allow us to increase the time resolution (reduce Δ_g) while reducing the frequency resolution (increase $\Delta_{\tilde{g}}$) as little as possible. The optimal choice for h is the Gaussian

$$h_\sigma(t) = (\pi\sigma^2)^{-1/4} \mathrm{e}^{-x^2/2\sigma^2}, \tag{B10}$$

which generates an equal sign in the uncertainty principle (B9), and its L_2 norm is 1, that is, $||h_\sigma||_2 = 1$, $\forall\,\sigma$. We omit the lengthly but straightforward derivations of these results. For $h = h_\sigma$, the windowed Fourier transform is termed the Gabor transform. To simplify the notation, we introduce the symbolizms $F_{h_\sigma} = G$ and $G[f](t,\omega) = \tilde{f}(t,\omega)$. The dependence of G and $\tilde{f}(t,\omega)$ on σ is implicit.

We close this short review with the expressions for the Gabor transform pair because it is of outstanding importance for the analysis in Section 3:

$$\tilde{f}(t,\omega) = G[f](t,\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\tau) h^*_\sigma(\tau - t) \mathrm{e}^{-\mathrm{i}\omega\tau}\, \mathrm{d}\tau, \tag{B11a}$$

$$f(t) = G^{-1}[\tilde{f}](t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^2} \tilde{f}(\tau,\omega) h_\sigma(t - \tau) \mathrm{e}^{\mathrm{i}\omega t}\, \mathrm{d}\omega\, \mathrm{d}\tau. \tag{B11b}$$

APPENDIX C: ADJOINT-SOURCE TIME FUNCTIONS FOR MEASUREMENTS OF CROSS-CORRELATION TIME-SHIFTS AND RMS AMPLITUDE DIFFERENCES

C1 Measurements of cross-correlation time-shifts

We closely follow the concept introduced by Luo & Schuster (1991). The cross-correlation time-shift Δt is defined as the time where the cross-correlation function

$$c(u_i^0, u_i)(\tau) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u_i^0(t) u_i(t + \tau)\, \mathrm{d}t, \tag{C1}$$

attains its global maximum. This implies that Δt satisfies the necessary condition

$$\partial_\tau c(u_i^0, u_i)(\tau)|_{\tau=\Delta t} = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} u_i^0(t) \dot{u}_i(t + \Delta t)\, \mathrm{d}t = 0. \tag{C2}$$

Eq. (C2) implicity defines Δt, at least when there is only one maximum. Invoking the implicit function differentiation yields the sensitivity density $\delta_p \Delta t$:

$$\delta_p \Delta t = -\frac{\delta_p \dot{c}(\Delta t)}{\ddot{c}(\Delta t)} = -\frac{\int_{\mathbb{R}} u_i^0(\tau - \Delta t)\, \delta_p \dot{u}_i(\tau)\, \mathrm{d}\tau}{\int_{\mathbb{R}} u_i^0(\tau - \Delta t)\, \ddot{u}_i(\tau)\, \mathrm{d}\tau}. \tag{C3}$$

Under the assumption that u_i^0 and u_i are purely time-shifted and not otherwise distorted with respect to each other, we can interchange $u_i^0(\tau - \Delta t)$ and $u_i(\tau)$ in (C3). For the sensitivity kernel $\delta_p\, \Delta t$, we then have

$$\delta_p \Delta t = -||v_i^0||_2^{-2} \int_{\mathbb{R}} \dot{u}_i(\tau)\, \delta_p u_i(\tau) \mathrm{d}\tau\,. \tag{C4}$$

The sensitivity kernel $\delta_p\, u_i(\tau)$ can be expressed in terms of the adjoint Green's function (18):

$$\delta_p \Delta t = ||v_i^0||_2^{-2} \int_{\mathbb{R}^2} \dot{u}_i(\tau) \left[\mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t) \cdot \partial_p \mathbf{L}(\mathbf{u}(\mathbf{x}, t)) \right] \mathrm{d}t\, \mathrm{d}\tau\,. \tag{C5}$$

Eq. (C5) defines the adjoint field $\mathbf{u}^\dagger{}_{\mathrm{cc},t}$ for measurements of cross-correlation time-shifts on the ith component of the data and synthetics:

$$\mathbf{u}^\dagger_{\mathrm{cc},i}(\mathbf{x}, t) = ||v_i^0||_2^{-2} \int_{\mathbb{R}} \dot{u}_i(\tau)\, \mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t)\, \mathrm{d}\tau \tag{C6}$$

and the corresponding adjoint-source time function:

$$s^\dagger_{\mathrm{cc},i}(\tau) = \frac{\dot{u}_i(\tau)}{||v_i^0||_2^2}. \tag{C7}$$

C2 Measurements of rms amplitude differences

In the case of the rms amplitude misfit $E^2_{\mathrm{rms}} = \Delta A^2/(A^0)^2$ with $\Delta A = A - A^0$ (see eq. 43), we follow, essentially, the same steps as in the previous paragraph on cross-correlation time-shifts. For the Fréchet kernel of E_{rms} we have

$$\delta_p E_{\mathrm{rms}} = \delta_p \sqrt{\frac{\Delta A^2}{(A^0)^2}} = \frac{\mathrm{sign}\,\Delta A}{A^0}\, \delta_p A = \frac{\mathrm{sign}\,\Delta A}{A\, A^0} \int_{\mathbb{R}} u_i(\tau)\, \delta_p u_i(\tau)\, \mathrm{d}\tau. \tag{C8}$$

Substituting (18) into (C8) yields

$$\delta_p E_{\mathrm{rms}} = \int_{\mathbb{R}} \mathbf{u}^\dagger_{\mathrm{rms},i}(\mathbf{x}, t)\, \partial_p \mathbf{L}[\mathbf{u}(\mathbf{x}, t)]\, \mathrm{d}t\,, \quad \mathbf{u}^\dagger_{\mathrm{rms},i}(\mathbf{x}, t) = -\frac{\mathrm{sign}\,\Delta A}{A\, A^0} \int_{\mathbb{R}} u_i(\tau)\, \mathbf{g}_i^\dagger(\mathbf{x}^r, \tau; \mathbf{x}, t)\, \mathrm{d}\tau. \tag{C9}$$

This defines the corresponding adjoint-source time function $s_{\mathrm{rms},i}$:

$$s_{\mathrm{rms},i}(\tau) = -\frac{\mathrm{sign}\,\Delta A}{||u_i||_2 ||u_i^0||_2}\, u_i(\tau). \tag{C10}$$