Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements

Ebru Bozdağ,[1] Jeannot Trampert[2] and Jeroen Tromp[1,3]

[1]*Department of Geosciences, Princeton University, Princeton, NJ, USA. E-mail: bozdag@princeton.edu*
[2]*Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands*
[3]*Program in Applied & Computational Mathematics, Princeton University, Princeton, NJ, USA*

SUMMARY

Resolution in seismic tomography intimately depends on data coverage, with different parts of seismograms sensitive to different parts of Earth's structure. In classical seismic tomography, the usable amount of data is often restricted because of approximations to the wave equation. 3-D numerical simulations of wave propagation provide new opportunities for increasing the amount of usable data in seismograms by choosing appropriate misfit functions which have direct control on Fréchet derivatives. We propose new misfit functions for full waveform tomography based on instantaneous phase differences and envelope ratios between observed and synthetic seismograms. The aim is to extract as much information as possible from a single seismogram. Using the properties of the Hilbert transform, we separate phase and amplitude information in the time domain. To gain insight in the advantages and disadvantages of chosen misfit functions, we make qualitative comparisons of the corresponding finite-frequency adjoint sensitivity kernels with those from commonly used misfit functions based on cross-correlation traveltime, amplitude and waveform differences. The major advantages of our misfit functions are: (1) working in the Hilbert domain reduces non-linear behaviour of waveforms due to interaction of phase and amplitude information, and (2) we show with noise-free synthetic seismograms that it is possible to use a complete seismogram without losing information from low-amplitude phases. Complementary to instantaneous phase measurements, envelope measurements provide a way of using amplitude information of waveforms, which may also easily be extended to constrain anelastic properties. The properties of the kernels allow us to simplify the tomography problem by separating elastic and anelastic inversions. First indications are that the kernels remain well behaved in the presence of noise.

Key words: Body waves; Surface waves and free oscillations; Seismic tomography; Computational seismology; Wave propagation.

GJI Seismology

1 INTRODUCTION

Seismic tomography is moving towards full waveform inversion by taking advantage of advances in the theory of seismic wave propagation, numerical techniques and increasing computational resources. We define 'full waveform inversion' as a technique which combines 3-D numerical wave simulations as a forward theory with Fréchet kernels computed in 3-D background models, to fit complete three-component seismograms. Full waveform inversions have been performed in local and regional studies (e.g. Chen *et al.* 2007b; Fichtner *et al.* 2009; Tape *et al.* 2009), but so far it has remained a challenge for global tomography. Capdeville *et al.* (2005) proposed a global approach in a synthetic experiment based on a source-stacking technique. They focused on long-period data to reduce the non-linearity of the problem, and identified the main obstacles as being data coverage and dominance of surface-wave information. Insufficient data coverage resulting from the uneven distribution of earthquakes and stations is one of the major restrictions at the global scale. It is difficult to change this distribution, but there is scope for increasing data coverage by extracting more information from individual seismograms. The dominance of surface waves is directly linked to their chosen misfit function and can be addressed.

In seismic tomography, data often are secondary observables, such as traveltimes, phase speeds or waveforms defined as a small portion of full seismograms. An important factor controlling the usable data is the forward theory. Ray-based tomography is limited to well-isolated body-wave phases (e.g. Zhou 1996; Boschi & Dziewonski 2000) or surface waves where the fundamental and higher modes are well separated

(e.g. Trampert & Woodhouse 1995; Ekström *et al.* 1997). The advantage of ray theory is its ease of implementation and minimal computational requirements. Integration of different data sets has been used to increase resolution of ray-based tomography (e.g. Su *et al.* 1994; Masters *et al.* 1996; Ritsema *et al.* 1999; Mégnin & Romanowicz 2000; Gu *et al.* 2001). Taking into account finite-frequency effects of wave propagation is the way to potentially improve resolution of tomographic images (e.g. Montelli *et al.* 2004; Sigloch *et al.* 2008), but this has to go beyond the use of identifiable phases (Boschi *et al.* 2007). Using asymptotic finite-frequency kernels, Li & Romanowicz (1996), Mégnin & Romanowicz (2000) and Gung & Romanowicz (2004) constructed global models based on waveforms obtained by cutting seismograms into energy wave packets where each packet was weighted appropriately by its energy, highlighting the importance of weighting to retrieve information from low-amplitude parts of data.

Progress in numerical techniques and computational facilities makes it now possible to solve the wave equation numerically in realistic 3-D earth models (e.g. Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999; Capdeville *et al.* 2003). This has two major consequences for seismic tomography: (1) full non-linearity of wave propagation may be taken into account in the forward problem and (2) Fréchet kernels may be computed numerically in 3-D background models during inversion. This has provided an interesting opportunity to increase the usable data and the challenge is to compute Fréchet derivatives efficiently. Zhao *et al.* (2005) proposed to compute Green's functions in 3-D models, storing them as a function of space and time, requiring large storage facilities. Tromp *et al.* (2005) addressed this problem by combining 3-D numerical simulations with adjoint techniques (e.g. Tarantola 1984, 1988; Fink 1997; Talagrand & Courtier 1987; Crase *et al.* 1990; Pratt 1999; Akçelik *et al.* 2003). The idea is that for a chosen set of observables and misfit function, Fréchet derivatives (i.e. sensitivity of data with respect to model parameters) may be computed by two numerical simulations only; one for the forward and one for the adjoint wavefield. Chen *et al.* (2007a) discuss the advantages and disadvantages of computing and storing 3-D Green's functions for all sources and receivers as a function of space and time (also called the 'scattering integral method') and adjoint methods.

Fréchet kernels are intricately linked to the chosen misfit function. In seismic tomography, common misfit functions are based on cross-correlation traveltime measurements (e.g. Luo & Schuster 1991; Marquering *et al.* 1999; Dahlen *et al.* 2000; Zhao *et al.* 2000), relative amplitude variations (e.g. Dahlen & Baig 2002; Ritsema *et al.* 2002) or waveform differences (e.g. Tarantola 1984, 1988; Nolet 1987). Examples of adjoint sensitivity kernels based on cross-correlation traveltime measurements may be found in Liu & Tromp (2006, 2008). Cross-correlation traveltime and relative amplitude measurements provide robust estimates of time and amplitude differences, but phases have to be isolated in seismograms, because cross-correlations are only meaningful for pulses with similar shapes. Automated phase-picking algorithms have recently been developed (e.g. Maggi *et al.* 2009) particularly tailored to adjoint tomography. Tape *et al.* (2009) showed a regional example of seismic tomography based on frequency-dependent traveltime measurements, also known as multitaper measurements (e.g. Zhou *et al.* 2004), using a conjugate gradient method together with adjoint kernels. Alternatively, Gee & Jordan (1992) proposed to use generalized seismological data functionals (GSDF) for frequency-dependent measurements. When isolation of phases is not practical or easy, time-domain waveform misfit functions (e.g. Tarantola 1984; Nolet 1987), defined by differences between observed and synthetic seismograms, are used. The waveform misfit is easily applied to whole seismograms, but it favours high-amplitude phases in a wave train containing multiple phases with different amplitudes. Thus, to extract optimal information, phases should be selected as in traveltime measurements or seismograms should be cut into wave packages with appropriate weightings (e.g. Li & Romanowicz 1996). Another concern is that waveform differences can be highly non-linear with respect to the model (e.g. Gauthier *et al.* 1986; Luo & Schuster 1991). Fichtner *et al.* (2008) proposed a full waveform inversion based on time–frequency analysis, which separates phase and amplitude information. It provides an elegant way to reduce non-linearities in the framework of the Born approximation.

We propose to use instantaneous phase and envelope misfits for full waveform inversion. This extracts a maximum amount of information from a single seismogram without analysing the signal piecewise and avoids non-linear mixtures of phase and amplitude. The instantaneous phase misfit is based on squared differences between instantaneous phases of observed and synthetic seismograms, and the envelope misfit on squared logarithmic envelope ratios. Instantaneous phases are widely used in exploration seismics (e.g. Taner *et al.* 1979; Perz *et al.* 2004; Barnes 2007) to increase resolution in imaging. Working with the Hilbert transform allows us to separate phase and amplitude information, similar to Fichtner *et al.* (2008) in the time–frequency domain. However, we propose a pure time-domain approach, which requires less data processing. Using 3-D numerical simulations, adjoint techniques offer us possibilities to compute the sensitivity of an arbitrarily measured part of a signal to 3-D structure. To have better insight into advantages or disadvantages of our chosen misfit functions, we compare corresponding kernels with those from waveform and traveltime differences, and amplitude ratios that are widely used in seismic tomography. Our primary focus is on elastic simulations but, for completeness, we show how envelope measurements can easily be extended to determine anelastic sensitivity kernels.

In the following, we derive adjoint sources for instantaneous phase and envelope misfits, and give a brief summary of classical waveform, traveltime and amplitue misfits. We then present sensitivity kernels computed by adjoint techniques for several synthetic experiments. Finally, we discuss our findings in the context of global mantle tomography and give general conclusions.

2 MISFIT FUNCTIONS AND ASSOCIATED ADJOINT SOURCES

We present derivations for elastic adjoint sources of our new misfit functions based on instantaneous phase (IP) and envelope (ENV) measurements. We also give expressions for adjoint sources of commonly used misfit functions of waveform (WF), traveltime (TT) and amplitue (AMP) measurements. In addition, we show how to construct anelastic kernels, following Tromp *et al.* (2005). All calculations of adjoint sensitivity kernels in the examples are accomplished using the spectral-element method (SEM) of Komatitsch & Tromp (2002a,b).

2.1 Adjoint kernels

In seismic waveform tomography, we extract information from a set of observed seismograms on model parameters describing Earth's interior. Model parameters are updated by minimizing a chosen misfit function between observed and synthetic data. In adjoint tomography, the gradient of the misfit function can be computed through the interaction of a forward wavefield with its adjoint wavefield, which is generated by the back-propagation of measurements made on data (Tarantola 1984). The non-linear inverse problem is then solved iteratively based on a gradient method, such as a (preconditioned) conjugate gradient or steepest descent method.

Let us define a generic waveform misfit function by

$$\chi(\mathbf{m}) = \sum_{r=1}^{N} \int_0^T g(\mathbf{x}_r, t, \mathbf{m}) \, \mathrm{d}t, \tag{1}$$

where N denotes the number of receivers and $g(\mathbf{x}_r, t, \mathbf{m})$ may be any kind of misfit between observed and synthetic data at receiver r with position $\mathbf{x}_r$. Time is denoted by t which runs between 0 and T, and $\mathbf{m}$ represents the model parameters. The gradient of the misfit function is simply

$$\delta\chi = \sum_{r=1}^{N} \int_0^T \partial_{\mathbf{s}} g(\mathbf{x}_r, t, \mathbf{m}) \cdot \delta\mathbf{s}(\mathbf{x}_r, t, \mathbf{m}) \, \mathrm{d}t, \tag{2}$$

where $\delta\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})$ denotes perturbations in the synthetic displacement wavefield $\mathbf{s}(\mathbf{x}_r, t, \mathbf{m})$ due to model perturbations $\delta\mathbf{m}$. In the following, to avoid clutter, we omit the dependence of the displacement wavefield $\mathbf{s}$ on the model parameters $\mathbf{m}$. Using the Born approximation (Hudson 1977; Wu & Aki 1985), the ith component of δs may be expressed as

$$\delta s_i(\mathbf{x}, t) = -\int_0^t \int_V [\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}, \mathbf{x}'; t - t')\partial_{t'}^2 s_j(\mathbf{x}', t')$$
$$+ \delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}, \mathbf{x}'; t - t')\partial_l' s_m(\mathbf{x}', t')]d^3\mathbf{x}' \, \mathrm{d}t', \tag{3}$$

where ρ is the density, c_{jklm} is the fourth-order elastic tensor and $\delta\rho$ and δc_{jklm} are their associated perturbations. Inserting eqs (3) into (2), the gradient becomes

$$\delta\chi = -\sum_{r=1}^{N} \int_0^T \partial_{s_i} g(\mathbf{x}_r, t, \mathbf{m}) \int_0^t \int_V [\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}_r, \mathbf{x}'; t - t')\partial_{t'}^2 s_j(\mathbf{x}', t')$$
$$+ \delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}_r, \mathbf{x}'; t - t')\partial_l' s_m(\mathbf{x}', t')]d^3\mathbf{x}'\mathrm{d}t' \, \mathrm{d}t. \tag{4}$$

Using the reciprocity of the Green's function (Aki & Richards 1980; Dahlen & Tromp 1998) and reversing time, it is convenient to define the adjoint wavefield

$$s_k^\dagger(\mathbf{x}', t') = \int_0^{t'} \int_V G_{ki}(\mathbf{x}', \mathbf{x}_r; t' - t)f_i^\dagger(\mathbf{x}, t)d^3\mathbf{x} \, \mathrm{d}t, \tag{5}$$

where $f_i^\dagger$ is the adjoint source given by

$$f_i^\dagger(\mathbf{x}, t) = \sum_{r=1}^{N} \partial_{s_i} g(\mathbf{x}_r, T - t, \mathbf{m})\delta(\mathbf{x} - \mathbf{x}_r). \tag{6}$$

For a more complete derivation of these expressions, please see Tromp *et al.* (2005). A Lagrange-multiplier approach for the derivation of the adjoint wavefield may be found in Liu & Tromp (2006) and Tromp *et al.* (2008).

For an isotropic material, the gradient of a given misfit function (eq. 4) may be written in the form

$$\delta\chi = \int_V [K_\rho(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_\mu(\mathbf{x})\delta \ln \mu(\mathbf{x}) + K_\kappa(\mathbf{x})\delta \ln \kappa(\mathbf{x})] \, d^3\mathbf{x}, \tag{7}$$

where K_ρ, K_μ, K_κ are the Fréchet derivatives with respect to the model parameters density (ρ), shear modulus (μ) and bulk modulus (κ), respectively. Comparing to eq. (4), using the definition of the adjoint wavefield in eq. (5), the Fréchet derivatives become

$$K_\rho(\mathbf{x}) = -\int_0^T \rho(\mathbf{x})\mathbf{s}^\dagger(\mathbf{x}, T - t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t) \, \mathrm{d}t, \tag{8}$$

$$K_\mu(\mathbf{x}) = -\int_0^T 2\mu(\mathbf{x})\mathbf{D}^\dagger(\mathbf{x}, T - t) : \mathbf{D}(\mathbf{x}, t) \, \mathrm{d}t, \tag{9}$$

$$K_\kappa(\mathbf{x}) = -\int_0^T \kappa(\mathbf{x})[\nabla \cdot \mathbf{s}^\dagger(\mathbf{x}, T - t)][\nabla \cdot \mathbf{s}(\mathbf{x}, t)] \, \mathrm{d}t, \tag{10}$$

where $\mathbf{D}$ and $\mathbf{D}^\dagger$ are the forward and adjoint traceless strain deviators, respectively. The above expressions show that sensitivity kernels may be obtained by two numerical simulations, one for the forward and one for the adjoint wavefield. Using relations between elastic moduli, density, P-wave (α) and S-wave (β) speeds, the gradient of the misfit function may alternatively be written as

$$\delta\chi = \int_V [K_\rho'(\mathbf{x})\delta \ln \rho(\mathbf{x}) + K_\beta(\mathbf{x})\delta \ln \beta(\mathbf{x}) + K_\alpha(\mathbf{x})\delta \ln \alpha(\mathbf{x})] \, d^3\mathbf{x}, \tag{11}$$

where K'_ρ, K_β and K_α are given by

$$K'_\rho = K_\rho + K_\kappa + K_\mu, \tag{12}$$

$$K_\beta = 2 \left(K_\mu - \frac{4}{3} \frac{\mu}{\kappa} K_\kappa \right), \tag{13}$$

$$K_\alpha = 2 \left(\frac{\kappa + \frac{4}{3}\mu}{\kappa} K_\kappa \right). \tag{14}$$

Adjoint kernels depend on the adjoint wavefield, which is generated by the adjoint source. As seen in eq. (6), the adjoint source depends on the pre-defined misfit function for specific observables. Key properties of adjoint kernels can thus be assessed from analysing the corresponding adjoint sources.

2.2 Hilbert transform

A signal having no negative-frequency components is called an analytic signal $\widetilde{f}(t)$ if it is constructed from a real signal $f(t)$ and its Hilbert transform $\mathcal{H}\{f(t)\}$:

$$\widetilde{f}(t) = f(t) - i\mathcal{H}\{f(t)\}. \tag{15}$$

The Hilbert transform of a signal is defined as

$$\mathcal{H}\{f(t)\} = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(\tau)}{t - \tau} \, d\tau, \tag{16}$$

where P stands for the Cauchy principal value. The analytic signal can be written in terms of the instantaneous amplitude $E(t)$ and the instantaneous phase $\phi(t)$ as

$$\widetilde{f}(t) = E(t)e^{i\phi(t)}, \tag{17}$$

where

$$\phi(t) = \arctan \frac{\Im\left\{\tilde{f}(t)\right\}}{\Re\left\{\tilde{f}(t)\right\}}, \tag{18}$$

and

$$E(t) = \sqrt{\Re\left\{\tilde{f}(t)\right\}^2 + \Im\left\{\tilde{f}(t)\right\}^2}. \tag{19}$$

This decomposition allows us to separate phase and amplitude of a signal in the time domain rather than in the frequency domain. $E(t)$ in eq. (19) is also known as the envelope of the real signal $f(t)$. In the following, we derive adjoint sources for IP and ENV measurements.

2.3 Instantaneous phase misfits

We define the squared instantaneous phase misfit as

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} \int_0^T [\phi_r^{\text{obs}}(t) - \phi_r(t, \mathbf{m})]^2 \, dt, \tag{20}$$

where $\phi_r^{\text{obs}}(t)$ and $\phi_r(t, \mathbf{m})$ denote the instantaneous phase of a specific component of observed and synthetic seismograms, respectively, recorded at receiver r, as a function of time t and a given earth model $\mathbf{m}$. Without explicitly introducing the corresponding notation, observed and synthetic seismograms are windowed and filtered within a certain frequency band before being compared to each other. To avoid clutter, we omit the dependence of ϕ_r^{obs}, ϕ_r and s_r on t and $\mathbf{m}$. The gradient of the misfit function is then

$$\delta\chi = -\sum_{r=1}^{N} \int_0^T (\phi_r^{\text{obs}} - \phi_r)\delta\phi_r \, dt, \tag{21}$$

where $\delta\phi_r$ is the perturbation in instantaneous phase of the synthetic seismogram due to a perturbation in the model parameters $\delta\mathbf{m}$. From eqs (15) and (18), ϕ_r is defined as

$$\phi_r = \arctan \frac{\Im(\tilde{s}_r)}{\Re(\tilde{s}_r)}, \tag{22}$$

where $\tilde{s}_r$ is the analytic signal corresponding to the synthetic seismogram s_r. The perturbation in instantaneous phase then becomes

$$\delta\phi_r = \delta\left[\frac{\Im(\tilde{s}_r)}{\Re(\tilde{s}_r)}\right] \bigg/ \left\{1 + \left[\frac{\Im(\tilde{s}_r)}{\Re(\tilde{s}_r)}\right]^2\right\}, \tag{23}$$

and after a little algebra we obtain

$$\delta\phi_r = \frac{(\mathcal{H}s_r)\delta s_r - s_r\delta(\mathcal{H}s_r)}{s_r^2 + (\mathcal{H}s_r)^2} . \tag{24}$$

If we insert eqs (24) into (21), the gradient of the misfit function becomes

$$\delta\chi = -\sum_{r=1}^{N}\int_0^T (\phi_r^{\mathrm{obs}} - \phi_r)\left[\frac{(\mathcal{H}s_r)\delta s_r}{E_r^2} - \frac{s_r\delta(\mathcal{H}s_r)}{E_r^2}\right]\mathrm{d}t. \tag{25}$$
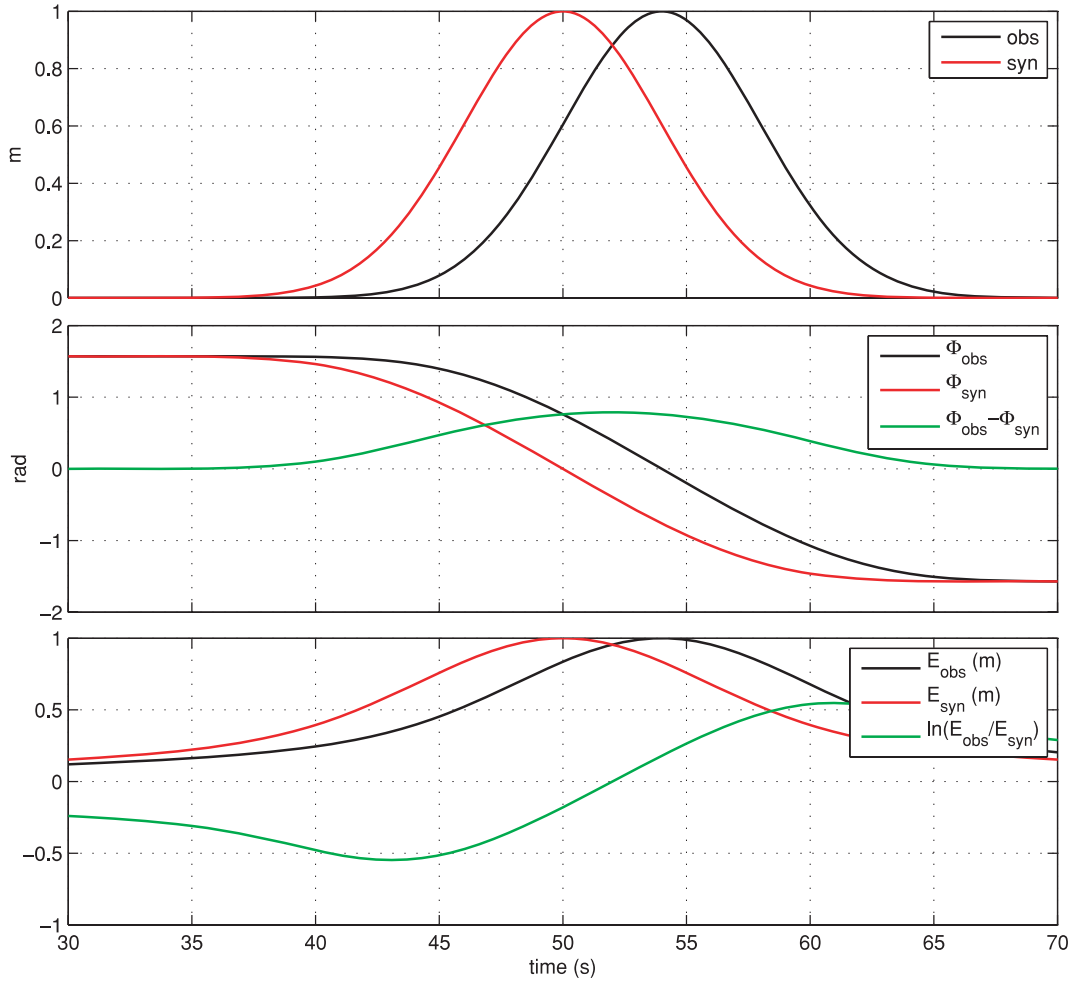
Figure 1. Instantaneous phase difference and envelope ratios for sample Gaussian signals shown in the *top plot*. In the *middle plot*, the instantaneous phase of each individual signal is shown in red and black and their difference in green. The instantaneous phase of each signal is monotonically decreasing from $\pi/2$ to $-\pi/2$. Their difference has a maximum where they intersect. In the *bottom plot*, the envelope of each signal is shown in red and black and their logarithmic ratio in green.
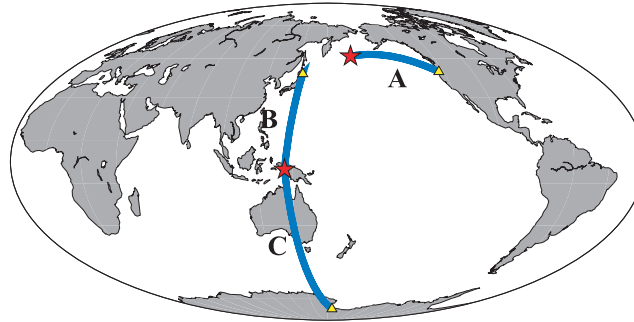
Figure 2. Great circle paths of seismograms used in this study. Path A corresponds to the Rat Island earthquake (2003 March 17, $M_w = 7.0$, depth $= 27$ km) recorded at station DBO. Note that for the synthetic experiment in Section 3.1, the CMT solution of this earthquake was modified to an explosive source. Paths B and C correspond to the Irian Jaya earthquake (2004 February 2, $M_w = 7.0$, depth $= 13$ km) recorded at stations ERM and CASY, respectively.

By pulling out δs_r from $\delta(\mathcal{H}s_r)$ (see appendix for details), we may rewrite the gradient as

$$\delta\chi = -\sum_{r=1}^{N}\int_0^T\left[(\phi_r^{\text{obs}}-\phi_r)\frac{(\mathcal{H}s_r)}{E_r^2}\delta s_r + \mathcal{H}\left\{(\phi_r^{\text{obs}}-\phi_r)\frac{s_r}{E_r^2}\right\}\delta s_r\right]dt.$$

(26)

Then the ith component of the adjoint source becomes

$$f_i^\dagger(\mathbf{x},t) = -\sum_{r=1}^{N}\left[\left[\phi_i^{\text{obs}}(\mathbf{x}_r,T-t)-\phi_i(\mathbf{x}_r,T-t,\mathbf{m})\right]\frac{w_r(T-t)\mathcal{H}\{s_i(\mathbf{x}_r,T-t,\mathbf{m})\}}{E_i(\mathbf{x}_r,T-t,\mathbf{m})^2}\right.$$
$$\left.+\ \mathcal{H}\left\{\left[\phi_i^{\text{obs}}(\mathbf{x}_r,T-t)-\phi_i(\mathbf{x}_r,T-t,\mathbf{m})\right]\frac{w_r(T-t)s_i(\mathbf{x}_r,T-t,\mathbf{m})}{E_i(\mathbf{x}_r,T-t,\mathbf{m})^2}\right\}\right]\delta(\mathbf{x}-\mathbf{x}_r),$$

(27)

where we explicitly inserted a windowing function w_r to express the possibility of using only parts of the seismogram. E_i is the envelope of s_i defined in eq. (19). The adjoint source has two terms, each normalized with the squared envelope of the synthetic seismograms (E_i^2). The first and second terms of the adjoint source look similar, however due to the time dependence of the phase difference and the envelope, they are not quite the same. If there is no phase difference, the adjoint source will be zero. The weighting function, generically defined as $1/E_i^2$, makes the adjoint kernel independent of signal amplitude. When the synthetic signal is zero, both nominator and denominator of the weighting function are zero. In this case, care has to be taken to put w_r to zero for those parts of the signal. Alternatively, a water level may be defined by adding a small term to the denominator to avoid division by small values (e.g. Tape *et al.* 2010).

To gain some insight into the concept of instantaneous phase, let us consider two hypothetic Gaussian signals (Fig. 1). The two signals are identical except that there is 4 s time-shift between them. The instantaneous phase of each signal monotonically decreases from $\pi/2$ to $-\pi/2$. Their difference is symmetric and has its maximum where the two signals cross. The weighting function will be minimum at the maximum of the synthetic signal.
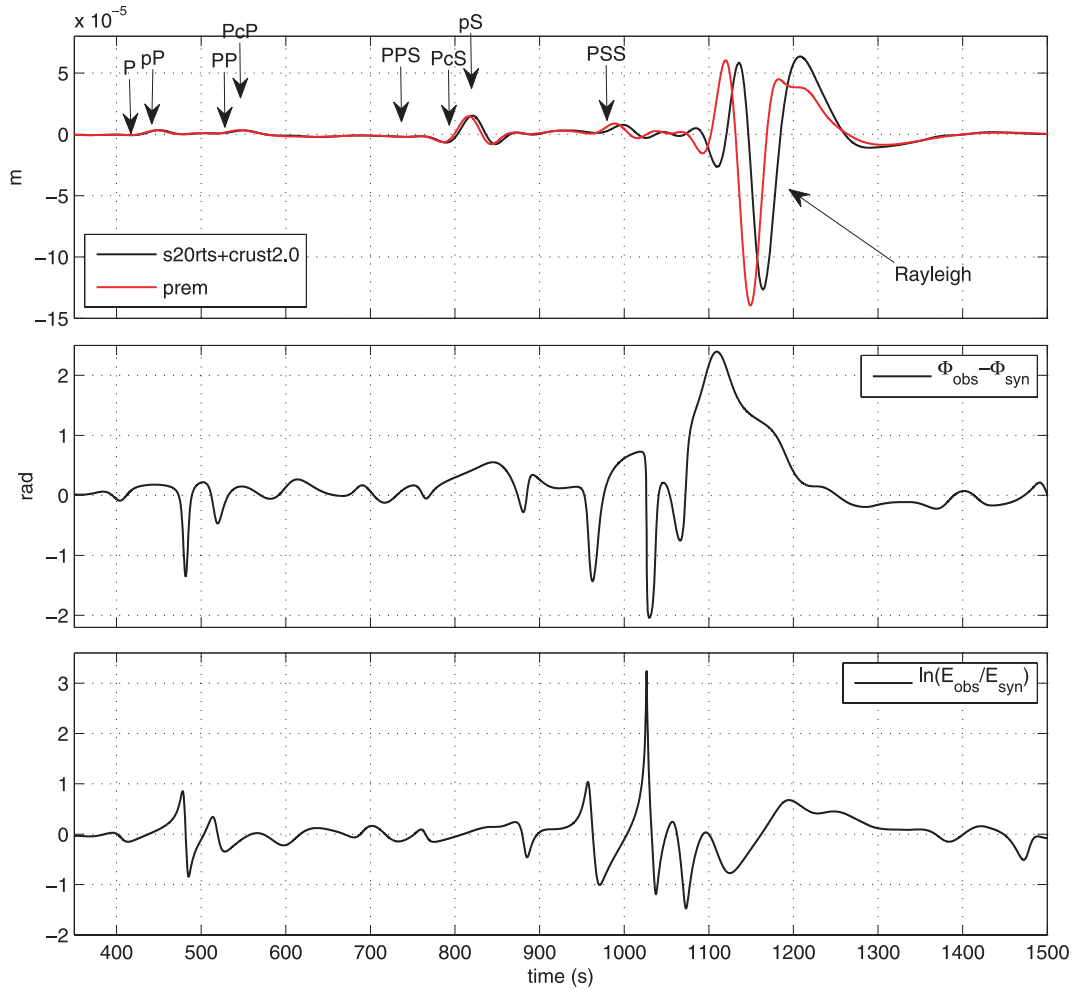

Figure 3. Vertical component SEM seismograms computed for PREM and S20RTS+Crust2.0 with seismic phases identified in the top. Note that there are only converted *S* phases due to the use of an explosive source. Seismograms are for path A shown in Fig. 2. The instantaneous phase difference is plotted in the middle and the logarithmic envelope ratio in the bottom.

2.4 Envelope misfits

We define the envelope misfit as the squared logarithmic ratio of the envelopes of observed and synthetic data

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} \int_0^T \left[\ln \frac{E_r^{\text{obs}}(t)}{E_r(t, \mathbf{m})} \right]^2 \, dt, \tag{28}$$

where $E_r(t, \mathbf{m})$ and $E_r^{\text{obs}}(t)$ are envelopes of synthetic and observed seismograms, respectively. A specific component of the signal is recorded at receiver r. To avoid clutter, we omit the dependence of E_r^{obs}, E_r and s_r on t and $\mathbf{m}$. The gradient of the envelope misfit is then

$$\delta\chi = -\sum_{r=1}^{N} \int_0^T \ln\left(\frac{E_r^{\text{obs}}}{E_r}\right) \frac{1}{E_r} \delta E_r \, dt, \tag{29}$$

where δE_r is the perturbation in the envelope of the synthetic seismogram due to a perturbation in the model parameters $\delta\mathbf{m}$. From eqs (15) and (19), the envelope of a component of the synthetic seismogram s_r is

$$E_r = \sqrt{\Re(\tilde{s}_r)^2 + \Im(\tilde{s}_r)^2}, \tag{30}$$

thus its perturbations become

$$\delta E_r = \frac{s_r \delta s_r + (\mathcal{H} s_r)\delta(\mathcal{H} s_r)}{\sqrt{s_r^2 + (\mathcal{H} s_r)^2}}. \tag{31}$$

If we substitute eqs (31) into (29), we obtain

$$\delta\chi = -\sum_{r=1}^{N} \int_0^T \ln\left(\frac{E_r^{\text{obs}}}{E_r}\right) \left[\frac{s_r \delta s_r}{E_r^2} + \frac{(\mathcal{H} s_r)\delta(\mathcal{H} s_r)}{E_r^2} \right] dt. \tag{32}$$

If we extract δs_r from $\delta(\mathcal{H} s_r)$ (see appendix for details),

$$\delta\chi = -\sum_{r=1}^{N} \int_0^T \left[\ln\left(\frac{E_r^{\text{obs}}}{E_r}\right) \frac{s_r}{E_r^2} \delta s_r - \mathcal{H}\left\{ \ln\left(\frac{E_r^{\text{obs}}}{E_r}\right) \frac{(\mathcal{H} s_r)}{E_r^2} \right\} \delta s_r \right] dt. \tag{33}$$
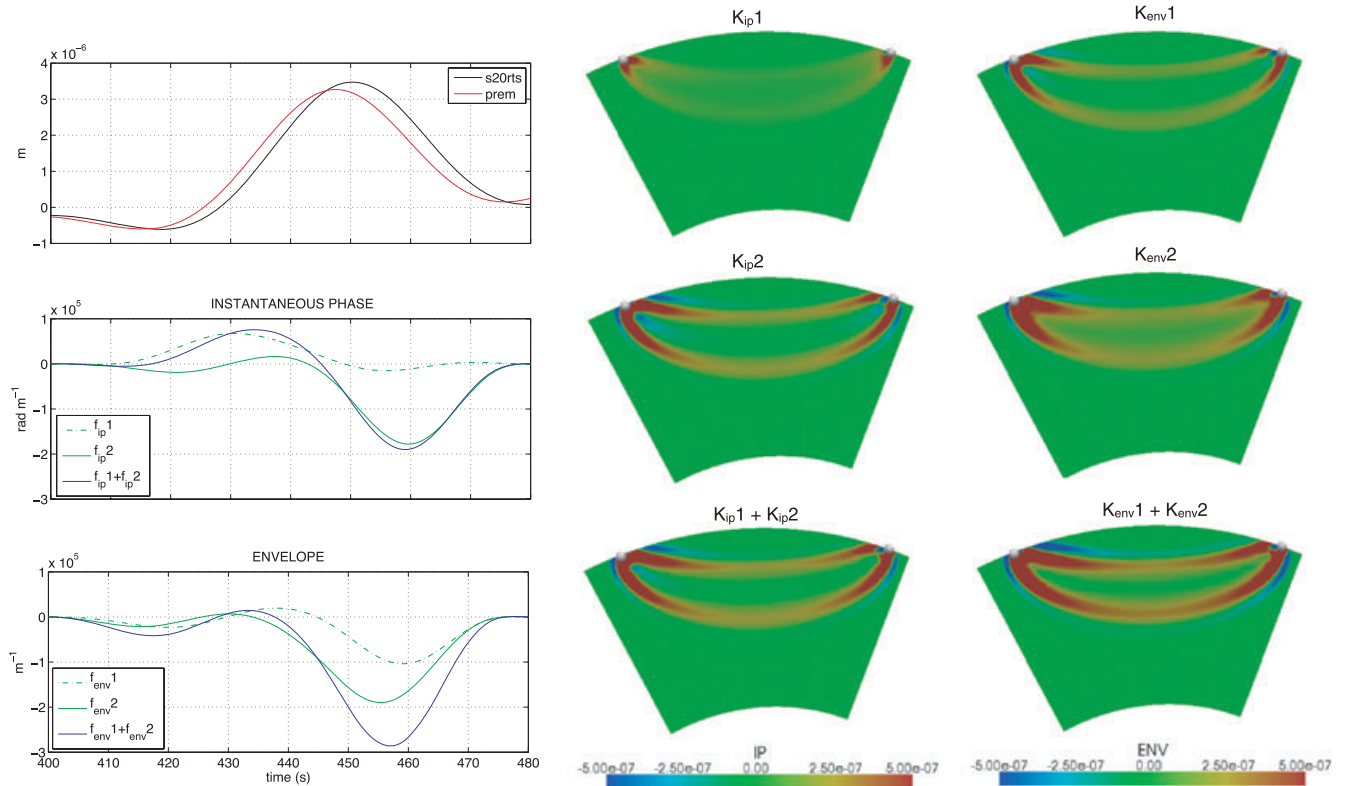

Figure 4. Construction of IP and ENV kernels for the first arriving P wave isolated from the seismograms shown in Fig. 3. Left: Adjoint sources for IP (middle plot) and ENV (bottom plot) measurements are shown for the first and the second terms in their expressions (eqs 27 and 34, respectively), separately (green lines). The total adjoint sources are simply the sum of these two terms (blue lines). P-wave seismograms are shown in the top plot. Middle: IP adjoint kernels from the first (upper plot) and the second (middle plot) terms of its adjoint source (eq. 27). Their sum gives the total adjoint kernel for IP measurements (bottom plot). Right: ENV adjoint kernels from the first (upper plot) and the second (middle plot) term of its adjoint source (eq. 34). Their sum gives the total adjoint kernel for ENV measurements (bottom plot). Units of IP and ENV kernels are rad s m^{-3} and s m^{-3}, respectively.

Then we can write the adjoint source for envelope measurements as

$$
f_i^\dagger(\mathbf{x}, t) = - \sum_{r=1}^{N} \left[\ln \left[\frac{E_r^{obs}(\mathbf{x}_r, t)}{E_r(\mathbf{x}_r, t, \mathbf{m})} \right] \frac{w_r(t) s_i(\mathbf{x}_r, T - t, \mathbf{m})}{E_i(\mathbf{x}_r, T - t, \mathbf{m})^2} \right.
$$
$$
\left. - \mathcal{H} \left\{ \ln \left[\frac{E_r^{obs}(\mathbf{x}_r, t)}{E_r(\mathbf{x}_r, t, \mathbf{m})} \right] \frac{w_r(t) \mathcal{H}\{s_i(\mathbf{x}_r, T - t, \mathbf{m})\}}{E_i(\mathbf{x}_r, T - t, \mathbf{m})^2} \right\} \right] \delta(\mathbf{x} - \mathbf{x}_r), \tag{34}
$$

where again we explicitly inserted the windowing function w_r. The adjoint source for the ENV misfit function, similar to that of the IP misfit, consists of two parts weighted by the same generic weighting function. When envelopes of observed and synthetic seismograms are the same, the adjoint source becomes zero, which happens when observed and synthetic seismograms are identical both in phase and in amplitude.

In Fig. 1, together with the instantaneous phase difference, the envelopes of two Gaussian signals and their logarithmic ratio are shown. Negative values of the logarithmic ratio occur when the envelope of the synthetic seismogram is larger than that of the observed seismogram. The weighting function, as for the instantaneous phase adjoint source, will be minimum at the maximum of the synthetic signal due to the normalization with the squared envelope.
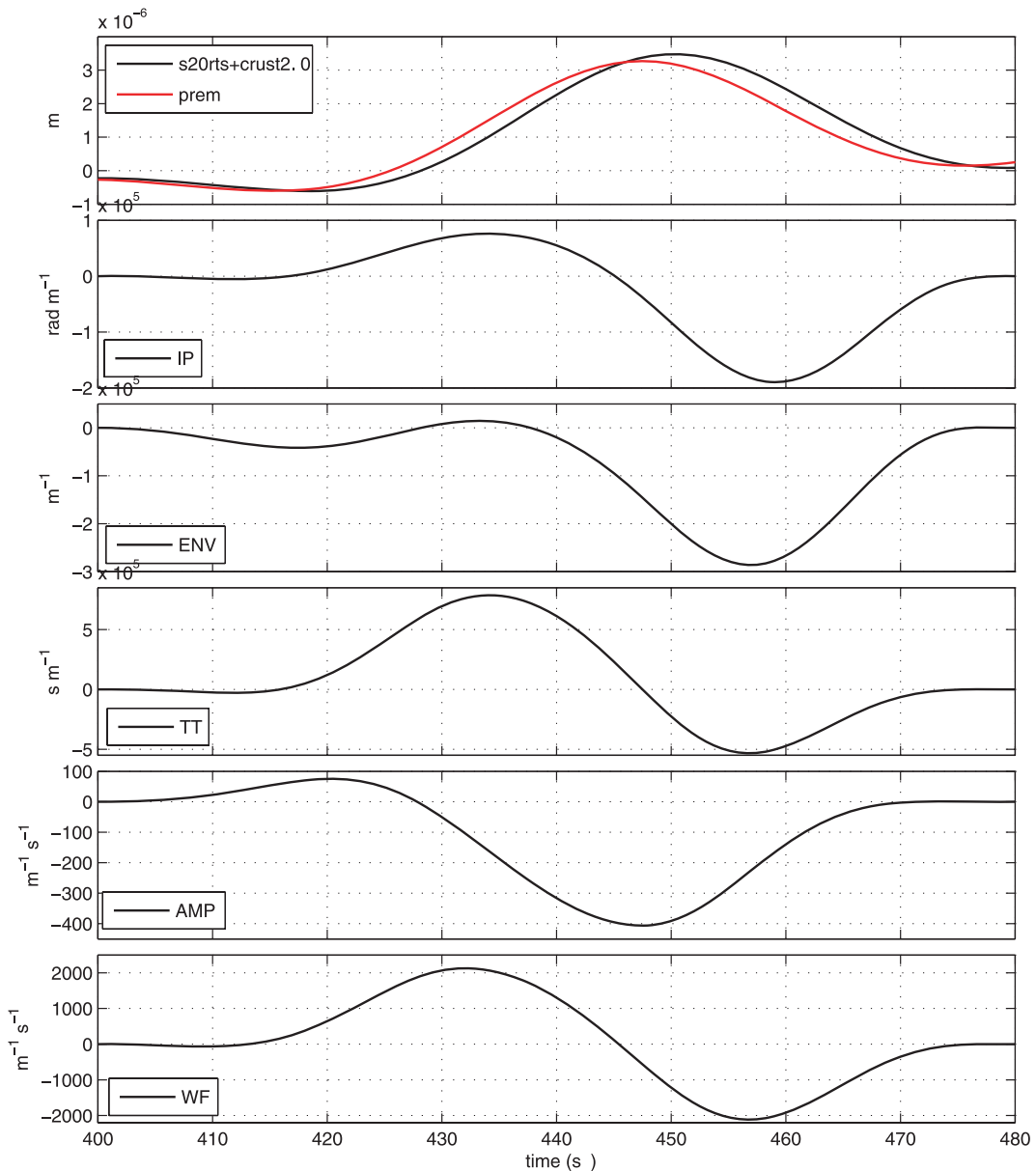

Figure 5. Adjoint sources for the first arriving *P* wave (top) cut from seismograms in Fig. 3 using IP, ENV, TT, AMP and normalized WF measurements. Adjoint sources are computed for full seismograms, then a cosine taper is applied to extract the adjoint source for the time window of the phase of interest.

2.5 Waveform misfits

This classical misfit function is based on differences between observed and synthetic waveforms combining amplitude together with phase information, and is defined as (e.g. Tarantola 1984, 1987, 1988; Nolet 1987)

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} \int_0^T \|\mathbf{d}(\mathbf{x}_r, t) - \mathbf{s}(\mathbf{x}_r, t, \mathbf{m})\|^2 \, dt, \tag{35}$$
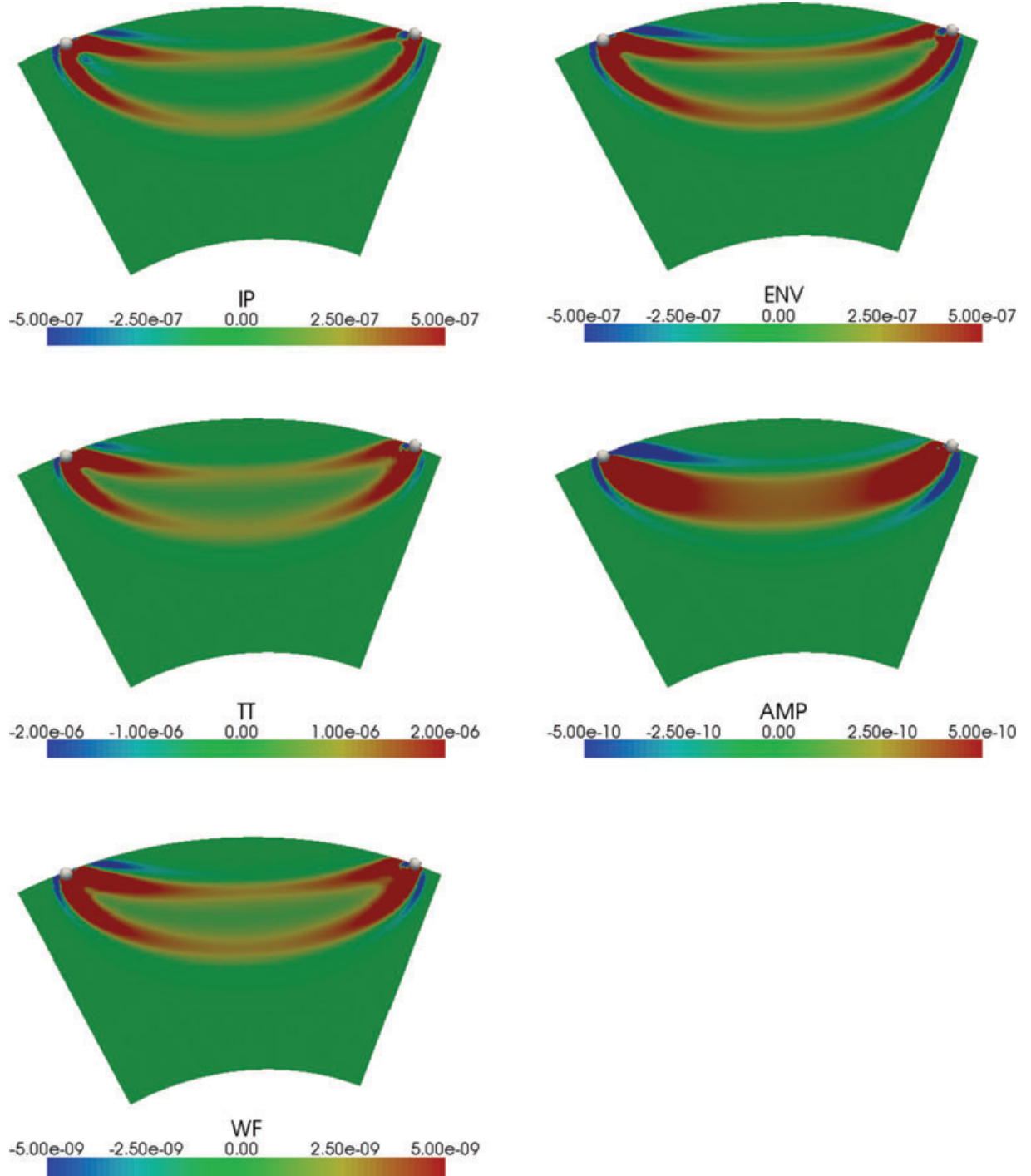
Figure 6. Source–receiver cross-sections of K_α adjoint kernels for P waves computed based on the adjoint sources in Fig. 5. Units of the kernels are rad s m^{-3} for instantaneous phase misfit (IP), s m^{-3} for envelope misfit, s^2 m^{-3} for traveltime misfit (TT), m^{-3} for amplitude misfit (AMP), m^{-3} for normalized waveform misfit (WF).

where **d** and **s** denote observed and synthetic waveforms, respectively. We quote the results from Tromp *et al.* (2005). The gradient of the misfit function is

$$\delta\chi = -\sum_{r=1}^{N}\int_{0}^{T}[\mathbf{d}(\mathbf{x}_r,t)-\mathbf{s}(\mathbf{x}_r,t,\mathbf{m})]\,\delta\mathbf{s}(\mathbf{x}_r,t,\mathbf{m})\,\mathrm{d}t, \tag{36}$$

where $\delta\mathbf{s}$ is the perturbation in the synthetic displacement field **s** due to perturbations in model parameters $\delta\mathbf{m}$. The *i*th component of the adjoint source is recognized to be

$$f_i^{\dagger}(\mathbf{x},t) = -\sum_{r=1}^{N}\frac{1}{M_r}[d_i(\mathbf{x}_r,T-t)-s_i(\mathbf{x}_r,T-t,\mathbf{m})]w_r(T-t)\delta(\mathbf{x}-\mathbf{x}_r), \tag{37}$$

where the time window w_r expresses the possibility of using only part of the waveform. We also inserted a normalization term M_r, defined as the energy of data, $M_r = \int_0^T w_r(t)d_i(x_i,t)^2\,\mathrm{d}t$, which equalizes amplitudes of different time windows in the inversion.
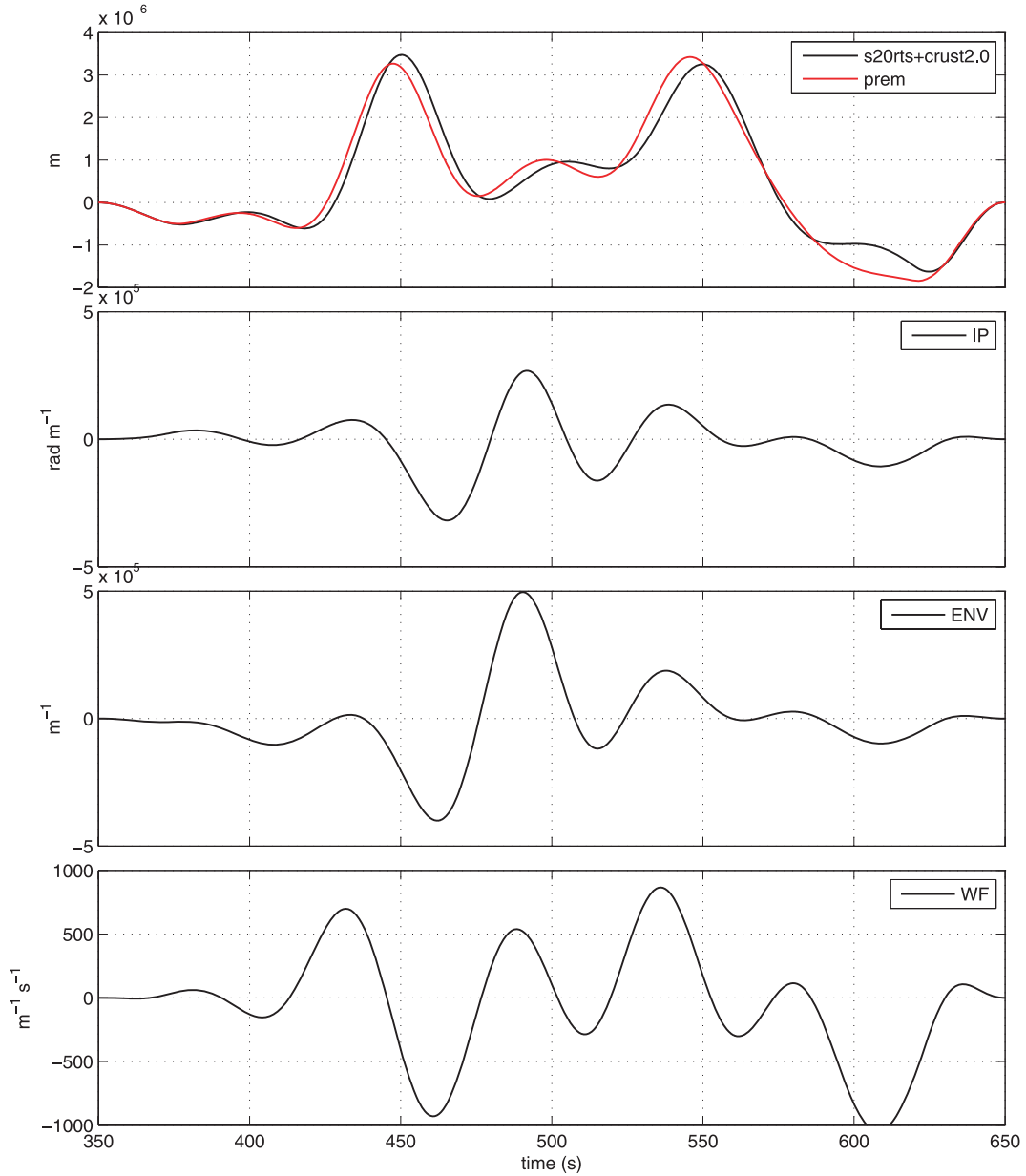

Figure 7. Adjoint sources for multiple *P* arrivals (*P*, *pP*, *PP*, *PcP*) (top) cut from the seismograms shown in Fig. 4 using IP, ENV, TT, AMP and normalized WF measurements. Adjoint sources are computed for full seismograms, then a cosine taper is applied to extract the adjoint source for the time window of the phase of interest.

2.6 Traveltime misfits

The misfit based on squared traveltime differences is widely used in seismic tomography and is defined as

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} [T_r^{\text{obs}} - T_r(\mathbf{m})]^2, \tag{38}$$

where T_r^{obs} is the observed traveltime of a selected phase at receiver r and $T_r(\mathbf{m})$ is the corresponding predicted traveltime. In this misfit function there is no explicit integration over time and traveltime difference is usually measured as the maximum cross-correlation between predicted and observed wavelets. Following the notation of Tromp *et al.* (2005), the gradient of the traveltime misfit function is

$$\delta\chi = -\sum_{r=1}^{N} [T_r^{\text{obs}} - T_r(\mathbf{m})]\delta T_r, \tag{39}$$

where δT_r is the perturbation in traveltime due to model perturbations $\delta\mathbf{m}$. If traveltime differences are measured by cross-correlation, it may be expressed as (e.g. Luo & Schuster 1991; Tanimoto 1995; Marquering *et al.* 1999; Dahlen *et al.* 2000)

$$\delta T_r = \frac{1}{N_r} \int_0^T w_r(t)\partial_t s_i(\mathbf{x}_r, t, \mathbf{m})\delta s_i(\mathbf{x}_r, t, \mathbf{m})\,dt, \tag{40}$$

where N_r is a normalization factor given by

$$N_r = \int_0^N w_r(t)s_i(\mathbf{x}_r, t, \mathbf{m})\partial_t^2 s_i(\mathbf{x}_r, t, \mathbf{m})\,dt, \tag{41}$$

and $w_r(t)$ is an appropriate window which isolates a specific phase. Inserting eqs (40) into (39), the *i*th component of the corresponding adjoint source is

$$f_i^{\dagger}(\mathbf{x}, t) = -\sum_{r=1}^{N} [T_r^{\text{obs}} - T_r(\mathbf{m})]\frac{1}{N_r}w_r(T - t)\partial_t s_i(\mathbf{x}_r, T - t, \mathbf{m})\delta(\mathbf{x} - \mathbf{x}_r). \tag{42}$$
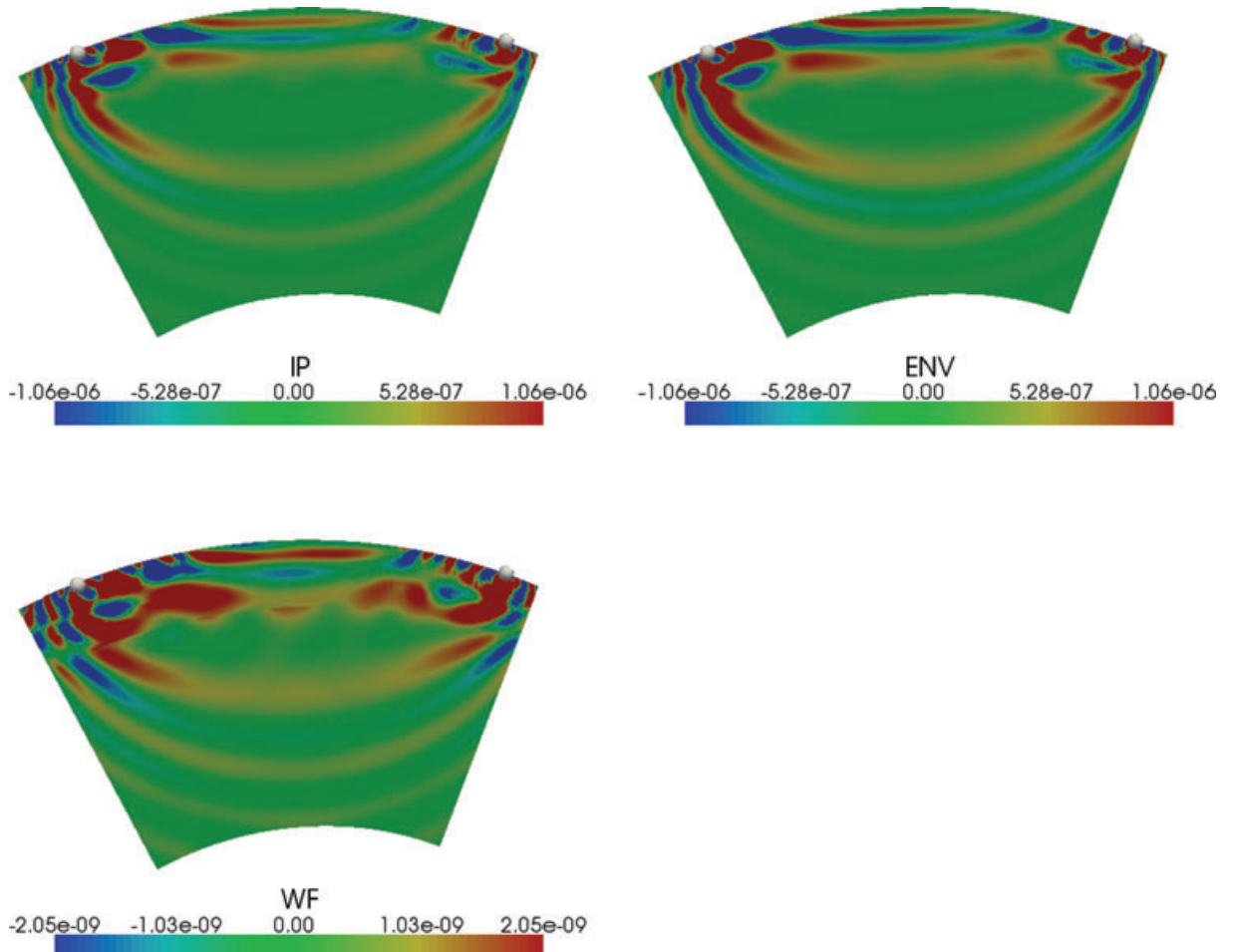

Figure 8. Source-receiver cross-sections of K_α adjoint kernels for multiple *P* arrivals (*P*, *PP*) computed based on the adjoint sources in Fig. 7.

2.7 Amplitude misfits

If $A_r^{obs}/A_r(\mathbf{m})$ denotes the ratio between observed and predicted amplitudes at station r, we can define the amplitude misfit as

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{N} \left[\ln \frac{A_r^{obs}}{A_r(\mathbf{m})} \right]^2. \tag{43}$$

We prefer a logarithmic ratio to the more commonly used relative amplitude variation (Dahlen & Baig 2002; Tromp *et al.* 2005) and again there is no integration over time because the amplitudes represent rms averages of isolated wavelets (Dahlen & Baig 2002). The gradient of the amplitude misfit function then becomes

$$\delta\chi = -\sum_{r=1}^{N} \ln \left[\frac{A_r^{obs}}{A_r(\mathbf{m})} \right] \delta \ln A_r, \tag{44}$$

where $\delta\ln A_r$ is given by (see Dahlen & Baig 2002)

$$\delta \ln A_r = \frac{1}{M_r} \int_0^T w_r(t) s_i(\mathbf{x}_r, t, \mathbf{m}) \delta s_i(\mathbf{x}_r, t, \mathbf{m}) \, dt. \tag{45}$$
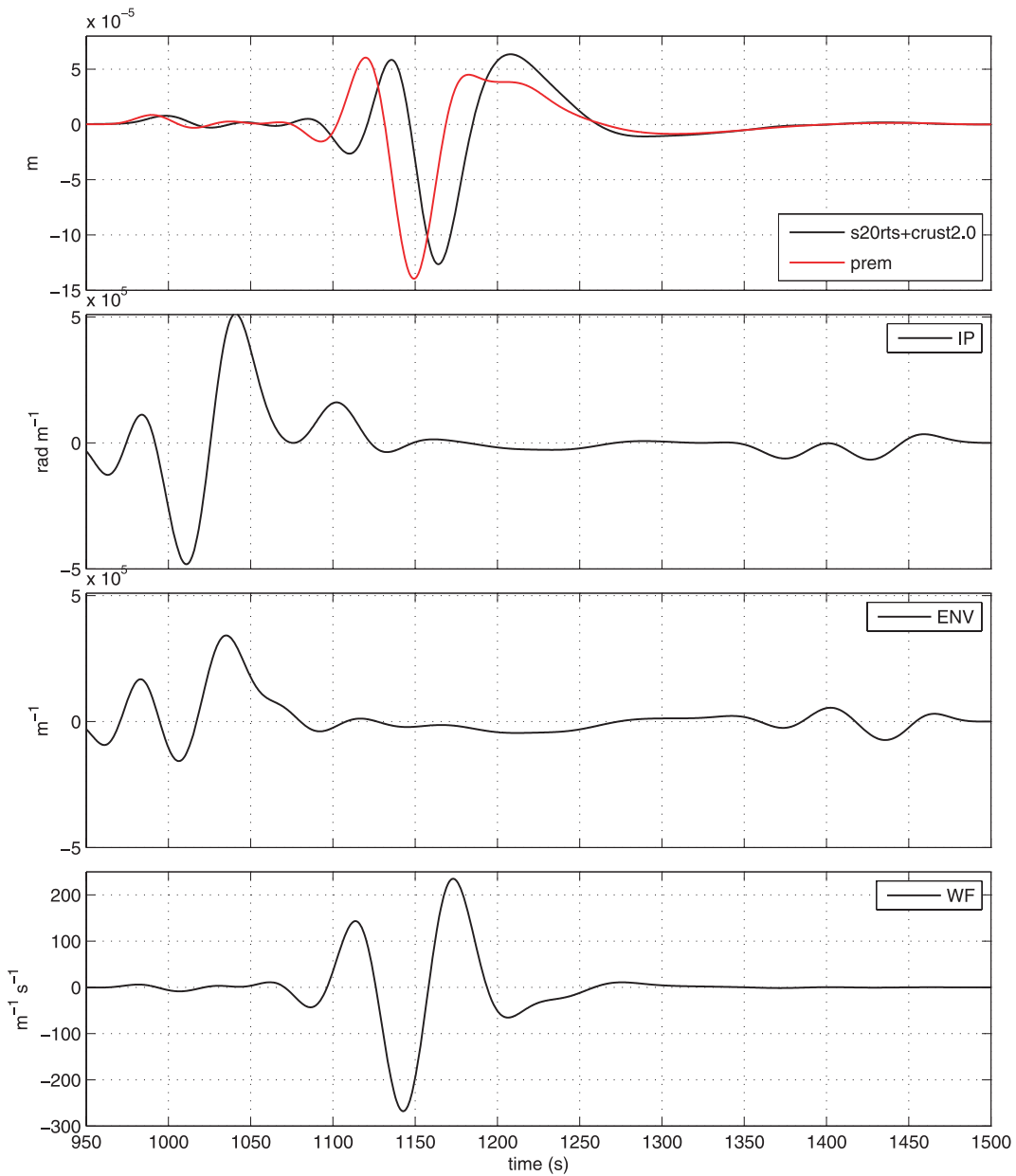

Figure 9. Same as in Fig. 7 but for surface waves.

The time window is $w_r(t)$ and δs_i denotes perturbations in the displacement wavefield due to changes in model parameters $\delta \mathbf{m}$. M_r is the normalization factor

$$M_r = \int_0^T w_r(t) s_i^2(\mathbf{x}_r, t, \mathbf{m}) \, \mathrm{d}t. \tag{46}$$

If we substitute eqs (45) into (44), the gradient may be written as

$$\delta \chi = -\sum_{r=1}^N \ln \left[\frac{A_r^{\mathrm{obs}}}{A_r(\mathbf{m})} \right] \frac{1}{M_r} \int_0^T w_r(t) s_i(\mathbf{x}_r, t) \delta s_i(\mathbf{x}_r, t) \, \mathrm{d}t, \tag{47}$$
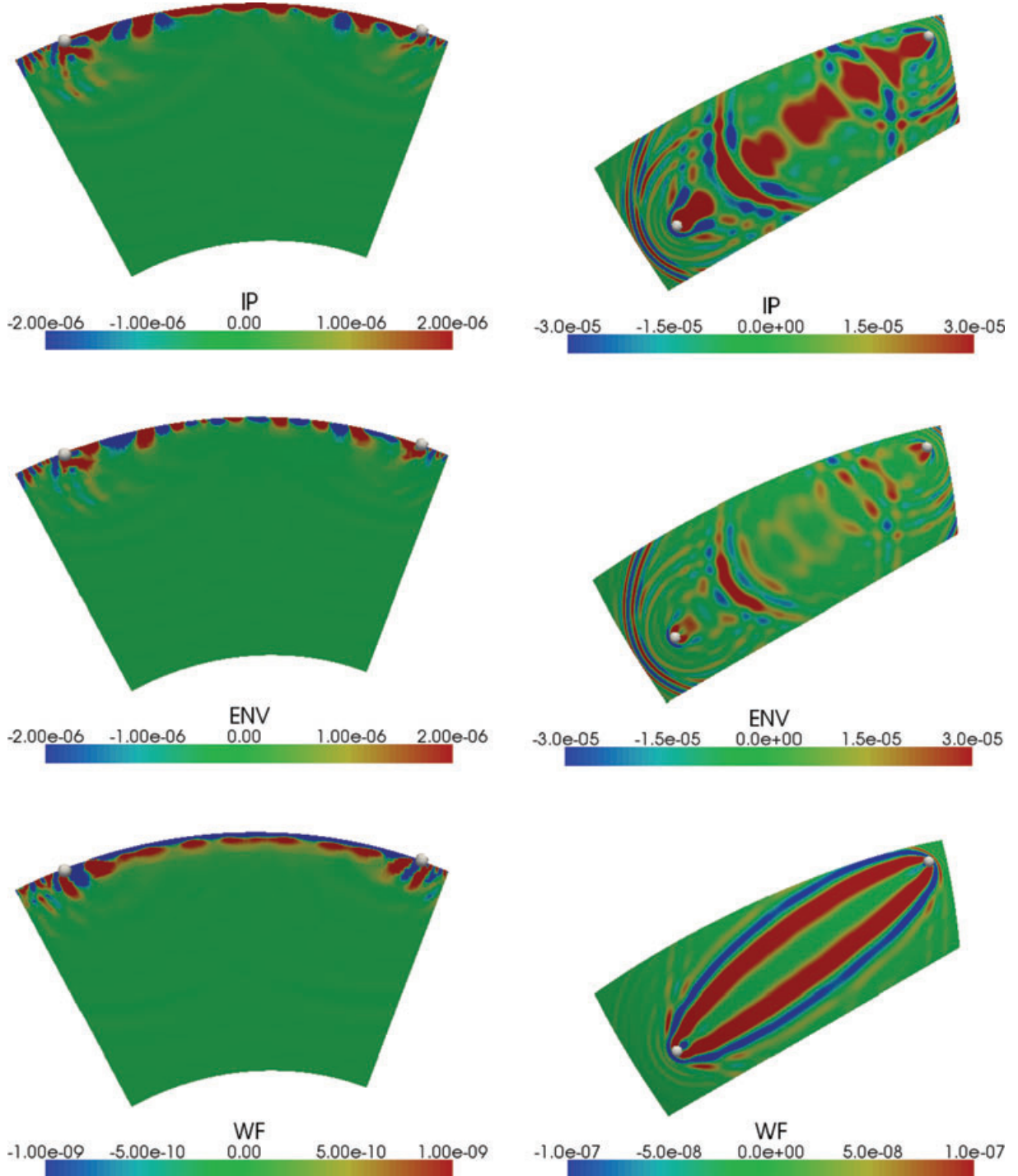

Figure 10. Source–receiver cross-sections (left column) and horizontal slices at the surface (right column) of K_α adjoint kernels for surface waves computed from the adjoint sources shown in Fig. 9.

where the ith component of the adjoint source is recognized to be

$$f_i^{\dagger}(\mathbf{x}, t) = -\sum_{r=1}^{N} \ln\left[\frac{A_r^{\text{obs}}}{A_r(\mathbf{m})}\right] \frac{1}{M_r} w_r(T-t) s_i(\mathbf{x}_r, T-t, \mathbf{m}) \delta(\mathbf{x} - \mathbf{x}_r). \tag{48}$$

2.8 Attenuation kernels

Envelope and amplitude measurements presented in Sections 2.4 and 2.7 are used to infer Earth's elastic properties. It is, however, well known that amplitudes or envelopes of seismograms are also very sensitive to variations in anelastic structure (e.g. Romanowicz 1995; Bhattacharyya *et al.* 1996; Reid *et al.* 2001; Dalton & Ekström 2006; Dalton *et al.* 2008). We may express the gradient of the misfit function in terms of perturbations in anelastic structure δQ^{-1}:

$$\delta\chi = \int_V K_\mu^Q(\mathbf{x}) \delta Q_\mu^{-1}(\mathbf{x}) \, d^3\mathbf{x}, \tag{49}$$
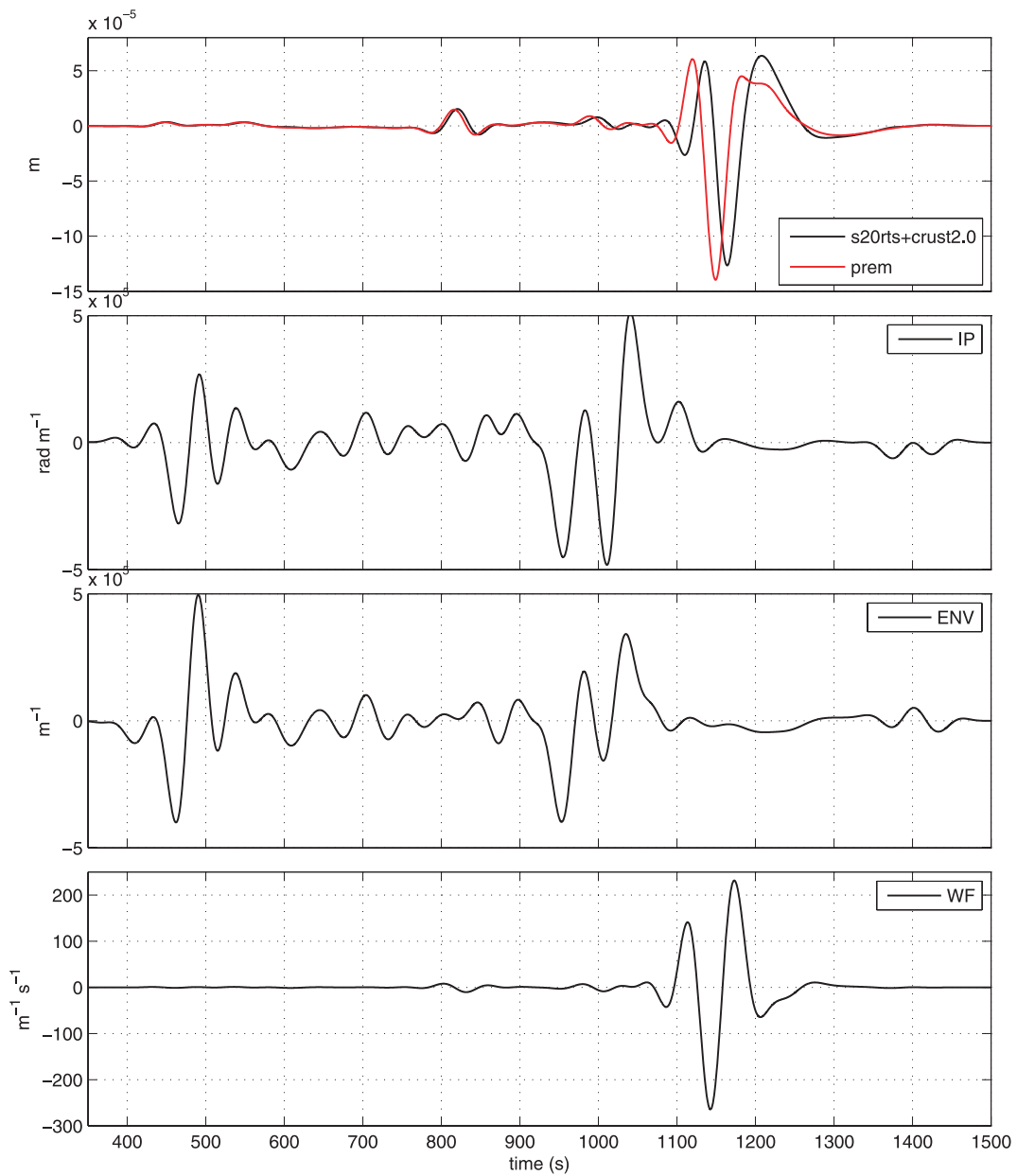

Figure 11. Same as in Fig. 7 but for complete seismograms.

where we ignore Q_κ^{-1} since intrinsic attenuation is dominated by shear attenuation Q_μ^{-1}. Assuming that Q^{-1} is constant over the seismic frequency band, the frequency-dependent shear modulus may be written as (Liu *et al.* 1976)

$$\mu(\omega) = \mu(\omega_0)[1 + (2/\pi)Q_\mu^{-1}\ln(|\omega|/\omega_0) - \mathrm{isgn}(\omega)Q_\mu^{-1}], \tag{50}$$

where ω_0 is a reference angular frequency. Following Tromp *et al.* (2005), the change in shear modulus $\delta\mu$ due to a change in shear attenuation δQ_μ^{-1} then becomes

$$\delta\mu(\omega) = \mu(\omega_0)[(2/\pi)\ln(|\omega|/\omega_0) - \mathrm{isgn}(\omega)]\delta Q_\mu^{-1}, \tag{51}$$

where $[(2/\pi)\ln(|\omega|/\omega_0)]$ captures the physical dispersion around the reference frequency. If we insert eq. (51) into the Fourier transformed Born approximation given in eq. (3), we may define the anelastic adjoint source which generates the anelastic adjoint wavefield as

$$\tilde{f}_i^\dagger(\mathbf{x}, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}[(2/\pi)\ln(|\omega|/\omega_0) - \mathrm{isgn}(\omega)]f_i^\dagger(\mathbf{x}, \omega)\exp(i\omega t)\,d\omega, \tag{52}$$

where $f_i^\dagger(\mathbf{x}, \omega)$ is the Fourier transform of the regular adjoint source defined for the elastic case which may be, for instance, the envelope or amplitude adjoint source given in eqs (34) and (48), respectively. Note that, the anelastic adjoint source has two parts: the first term is responsible for physical dispersion whereas the second part, which is simply the Hilbert transform of the regular adjoint source ($f_i^\dagger$), is responsible for amplitude variations. More information on the computation of anelastic kernels by a spectral-element method may be found in a forthcoming paper by Zhou *et al.* (in preparation).

3 RESULTS

To understand the properties of the newly introduced instantaneous phase (IP) and envelope (ENV) misfits, we compare them to the more commonly used waveform (WF), traveltime (TT) and amplitude (AMP) misfits. The latter two should be used for isolated phases and waveforms of similar shape. They cannot easily be generalized to whole seismograms, and figure only in examples where isolated phases are concerned. We compute 3-D sensitivity kernels for all misfit functions using adjoint simulations based on the SEM of Komatitsch & Tromp (2002a,b). To gain insight into how the selected misfit functions behave, we perform purely synthetic experiments. We first calculate adjoint
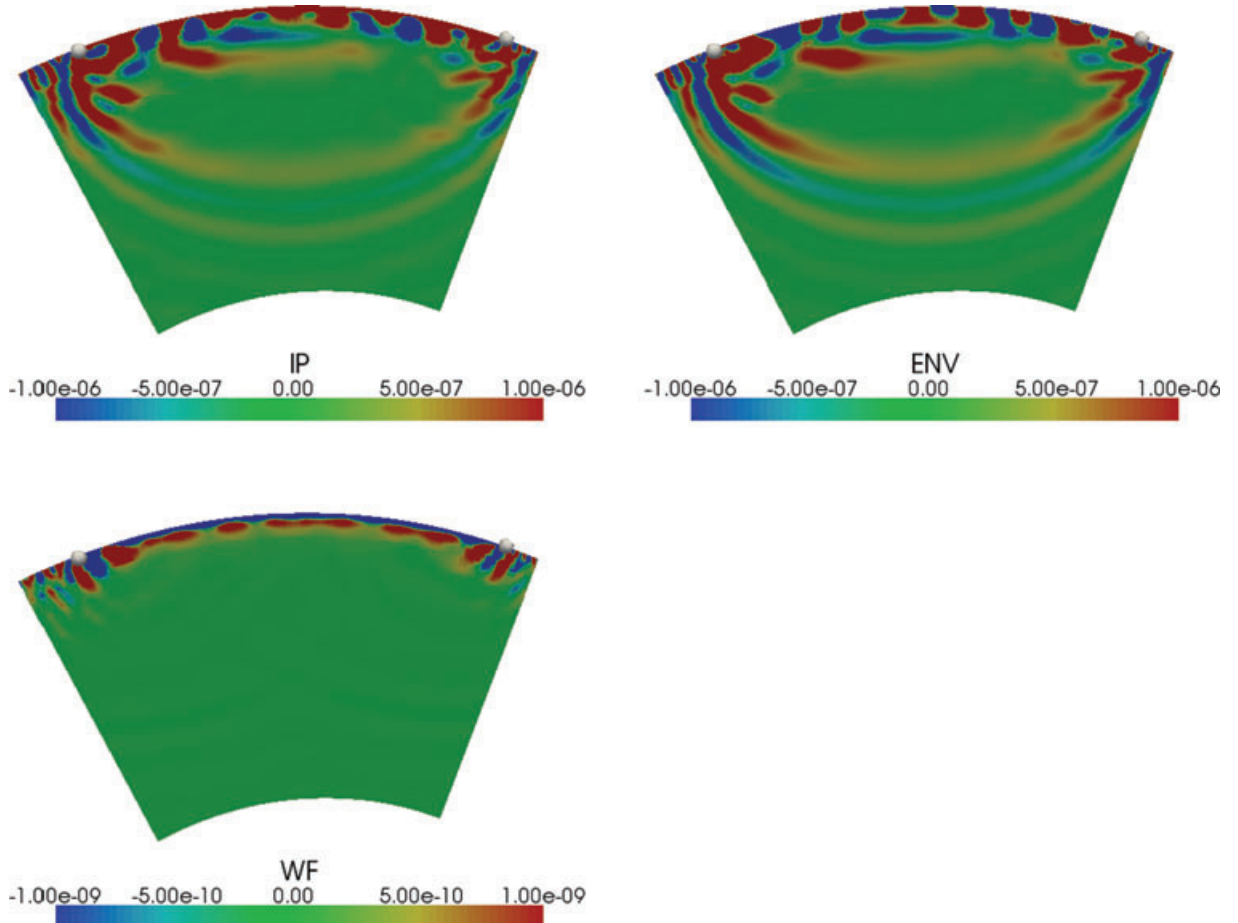

Figure 12. Source–receiver cross-sections of K_α adjoint kernels for complete seismograms computed from the adjoint sources shown in Fig. 11.

kernels in a 1-D background model, and use isotropic PREM (Dziewonski & Anderson 1981) and 3-D mantle model S20RTS (Ritsema *et al.* 1999) together with 3-D crustal model Crust2.0 (Bassin *et al.* 2000) (S20RTS+Crust2.0) to calculate seismograms corresponding to our *synthetic* and *observed* data, respectively. We use an explosive source to simplify kernels and present results for the vertical component. In the second example, we calculate kernels in a 3-D background model. We use the crust and uppermantle from S20RTS+Crust2.0 and fix the lower mantle to PREM (S20RTS$_{um}$+Crust2.0) to construct *synthetic* reference data. Whole 3-D mantle model S20RTS with Crust2.0 (S20RTS+Crust2.0) is again used to calculate the *observed* seismograms. In this case, we use the original CMT solution of a real earthquake and present results for the transverse component. In the last part, we illustrate anelastic kernels. We use isotropic PREM with 1-D PREM Q and 3-D Q (Dalton *et al.* 2008) models to compute *synthetic* and *observed* seismograms, respectively, and show shear attenuation kernels for S waves based on envelope and amplitude measurements.

3.1 Kernels in 1-D background models

We compute seismograms in isotropic PREM and in S20RTS+Crust2.0 which play the role of *synthetic* and *observed* data, respectively. We modified the CMT solution of the Rat Islands earthquake (2003 March 17, $M_w = 7.0$) to an explosive source in order to eliminate complexity due to the radiation pattern and to simplify kernels. We present results for the vertical component of the wavefield recorded at station DBO ($\Delta = 40°$) (path A in Fig. 2). Before computing the adjoint sources, we bandpass filter the SEM seismograms between 40 and 500 s. Seismograms together with their instantaneous phase difference and their logarithmic envelope ratios are shown in Fig. 3. The largest phase shifts and envelope differences are observed for surface waves and multiply reflected body waves.

We compute adjoint sources and their associated adjoint kernels for windowed sections of data. We use a cosine window to extract the part of interest from the seismograms. We qualitatively compare computed finite-frequency adjoint kernels for different misfit functions. We filter adjoint sources before computing sensitivity kernels with the same bandpass filter used for filtering the seismograms to avoid high-frequency noise in kernels.
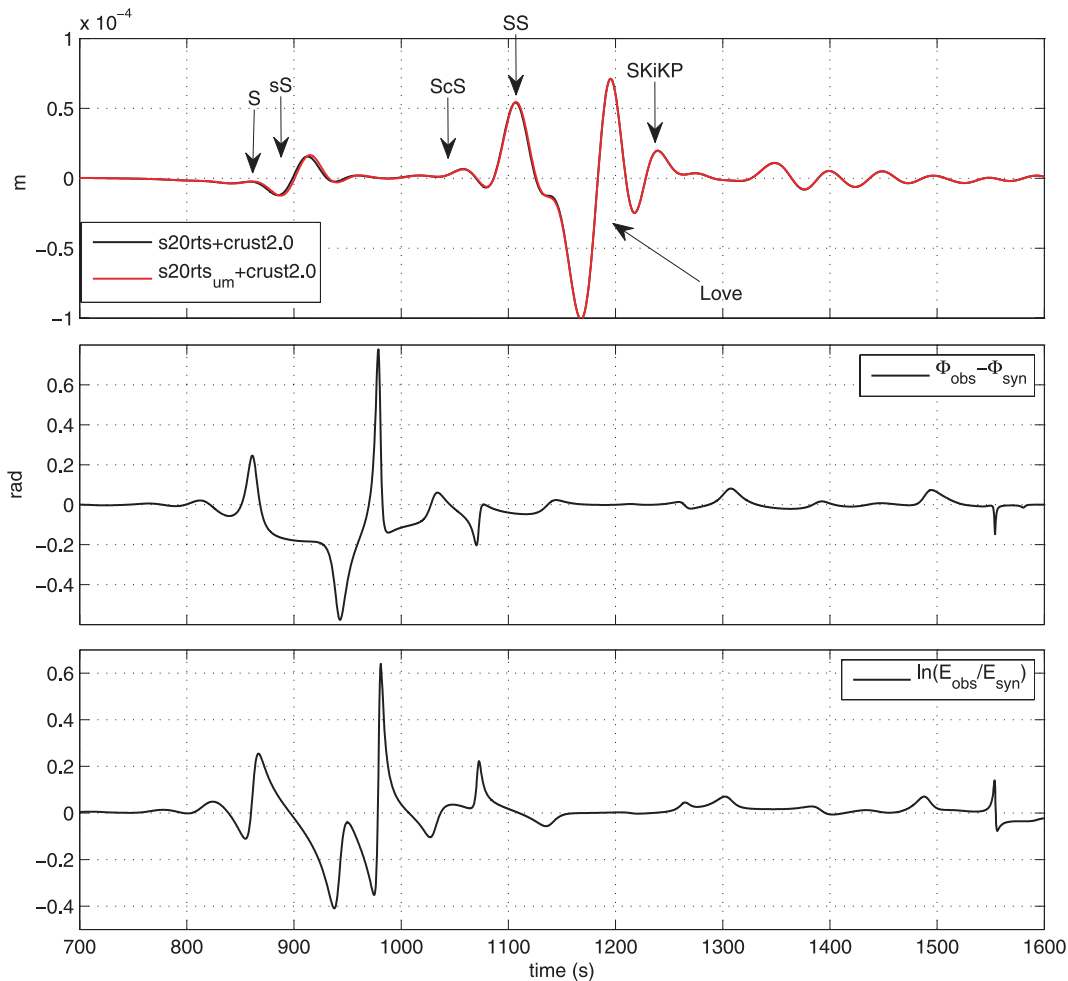

Figure 13. Transverse component SEM seismograms computed for S20RTS$_{um}$+Crust2.0 and S20RTS+Crust2.0, with seismic phases identified in the top. Models for both seismograms are fixed in the uppermantle, and only differs in the lower mantle to induce body-wave misfit only. Seismograms are for path B shown in Fig. 2. Instantaneous phase difference is plotted in the middle and the logarithmic envelope ratio in the bottom.

3.1.1 First arriving P wave

We first present the construction of IP and ENV kernels from the first and the second terms of the associated adjoint sources using the *P*-wave example. We show adjoint sources for the first and the second terms in eqs (27) and (34) and their associated kernels in Fig. 4. The total adjoint sources and kernels for IP and ENV measurements are simply sums of those two terms. Because of time-dependent normalization factors, the two parts of the adjoint sources, and hence the kernels, are not the same, although they appear similar in shape. For *P* waves, the total IP kernel is broader than the ENV kernel, whereas the amplitude of the ENV kernel is larger. However, overall they have similar attributes.

In Fig. 5, we compare adjoint sources computed for IP, ENV, TT, AMP and WF measurements for the same *P* waves considered earlier. There is approximately a 2.5 s difference between observed and synthetic *P* waves. The adjoint sources are not identical, reflecting different properties of their corresponding misfit functions. Being both phase measurements, IP and TT adjoint sources, for a single phase, are similar. Since the amplitudes of the *P* waves do not differ much in this case, WF also has a similar adjoint source. The AMP adjoint source differs the most from the others, whereas the ENV adjoint source is similar to the others. We see the characteristics of the adjoint sources in their adjoint kernels as well (Fig. 6). The TT adjoint kernel for *P* waves shows a typical banana-doughnut shape. The IP kernel is similar, but has a broader hole. Because of the time dependence of the ENV misfit, it behaves similarly to the WF misfit, thus their kernels are similar.
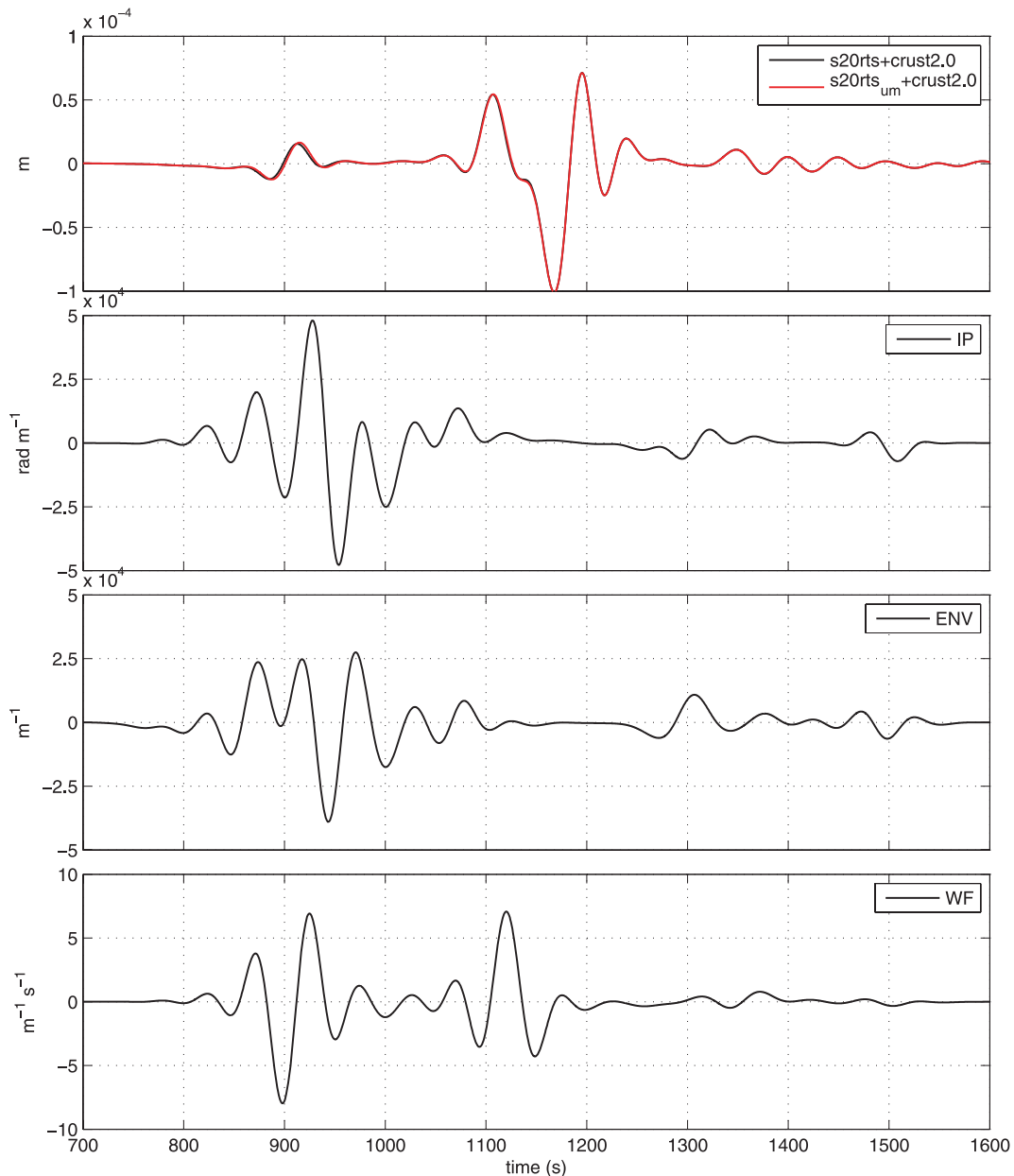

Figure 14. Adjoint sources computed for complete seismograms (top, also shown in Fig. 13) for IP, ENV and normalized WF measurements.

The AMP measurement produces a classical amplitude kernel. If the phases are aligned before measurement, then ENV should give similar sensitivity to AMP.

3.1.2 Multiple P arrivals

We now set the time window between 350 and 650 s, where we observe multiple *P*-wave arrivals, namely *P, pP, PP* and *PcP* (Fig. 7). The adjoint source for the WF measurement shows pronounced oscillations for the whole signal due to the changes in amplitude. IP and ENV emphasize the part between *P* and *PP* waves. Adjoint kernels for IP, ENV and WF measurements are somewhat similar (Fig. 8). However, due to the oscillatory behaviour of its adjoint source, WF shows a complicated interaction between *P* and *PP* sensitivities. In the IP and ENV kernels, the *P*-wave sensitivity is dominant, but has nicely developed higher Fresnel zones due to *P* and *PP* interactions. In all kernels, PcP sensitivity is very small compared with the other phases.

3.1.3 Surface waves

We set the time window large enough to capture not only the fundamental mode but also overtones and multiple surface reflections, such as the PSS phase. Since the adjoint sources for the IP and ENV measurements give equal weights to all seismic phases, low-amplitude parts of the signal have high amplitudes in the adjoint source. These are the overtones at ∼1050 s and the long-period surface waves at ∼1400 s (Fig. 9). WF measurements favour high-amplitude parts of the signal, as expected. Although all misfit kernels are of course sensitive to the uppermost part of Earth's structure, they are so in different ways. When we look at the sensitivities at the surface (Fig. 10), we see that the IP kernel is narrow and has a typical interference pattern of fundamental mode and overtones (e.g. Sieminski *et al.* 2007). The ENV kernel has the same pattern but with holes inside. It is now different from the IP kernel mainly due to changes in amplitudes. The WF kernel shows a pure fundamental-mode surface-wave sensitivity.

3.1.4 Complete seismograms

For a single phase or a small portion of the seismograms, sensitivity kernels for different misfit functions are not too different in their overall appearance. The differences are in the detailed structure of the kernels. An interesting question is how these different measurements see
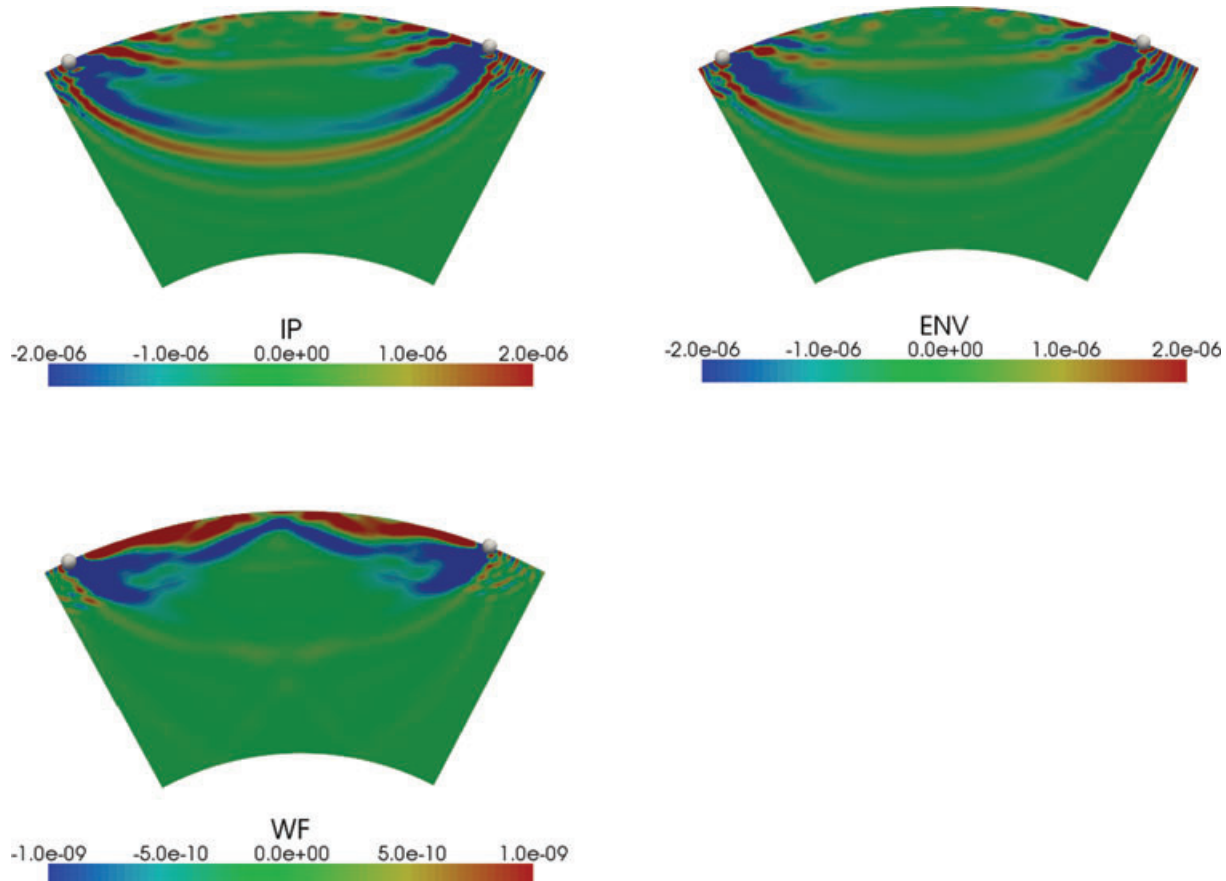

Figure 15. Source–receiver cross-sections of K_β adjoint kernels for complete seismograms computed based on the adjoint sources shown in Fig. 14.

structure when we consider entire seismograms. In Fig. 11, we present adjoint sources for WF, IP and ENV measurements for complete seismograms. The WF adjoint source favours high-amplitude parts of signals, surface waves in this case. It has no visible *P*-wave sensitivity (Fig. 12), although there are differences in the *P* waves. IP measurements provide additional information on the body-wave part of the signal due to the intrinsic weighting function in its adjoint source. The ENV kernel has the same property as the IP kernel, and shows sensitivities to both *P* and surface waves.

3.2 Kernels in 3-D background models

The highest wave speed heterogeneity occurs in the uppermost part of the Earth, and therefore, for shallow earthquakes, we observe the largest differences in surface waves. To investigate kernel properties further, we focus on body-wave differences by fixing the upper-mantle and crustal parts of models for both sets of seismograms. Seismograms computed in S20RTS+Crust2.0 again play the role of *observed* seismograms whereas the S20RTS$_{um}$+Crust2.0 seismograms are now the *synthetic* seismograms, and kernels are with respect to this latter model. We used the Irian Jaya earthquake (2004 February 5, $M_w = 7.0$) and present results for the transverse component of the wavefield recorded at station ERM ($\Delta = 46°$) (path B in Fig. 2).

    Observed and *synthetic* seismograms, their IP difference and ENV ratios are shown in Fig. 13. Surface waves in both SEM seismograms are almost the same with some minor differences. *S, SS* and *ScS* phases differ most due to their sensitivities to the lowermantle. Adjoint sources obtained from IP, ENV and WF measurements, filtered between 40 and 500 s in the same way as the seismograms, are shown in Fig. 14. WF measurements naturally favour high-amplitude parts of the data. IP and ENV measurements emphasize parts where they detect the largest phase and envelope differences. The difference between WF and IP–ENV measurements is very clear from their sensitivity kernels (Fig. 15). WF is most sensitive to the *SS* phase, because the adjoint wavefield interacts with the forward seismograms and both favour the high-amplitude *SS* phase. In an inversion, most of the signal would therefore be imaged to the upper mantle, even though both models are identical here. The IP and ENV adjoint sources have highest amplitudes in the *S*-wave part, and their interaction with the forward field results in a strong lower mantle sensitivity where data and synthetic differences originate.
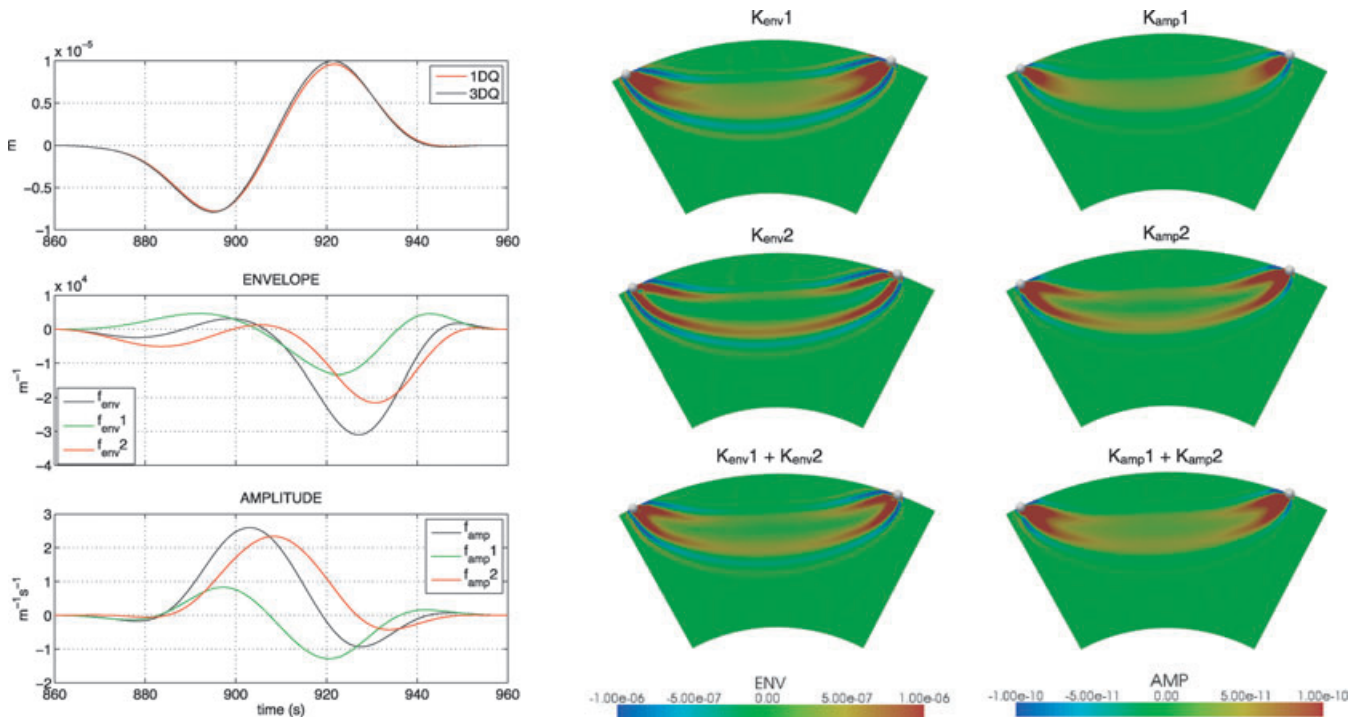

Figure 16. Construction of anelastic envelope (ENV) and amplitude (AMP) kernels for first arriving *S* waves. Left: adjoint sources for ENV (middle plot) and AMP (bottom plot) measurements are shown for dispersion and amplitude terms in their corresponding expressions (eq. 52), separately (green and red lines, respectively). The total anelastic adjoint sources are simply the sum of these two terms (blue lines). *S*-wave seismograms are shown in the top plot. The elastic structure (isotropic PREM) is fixed, but 1-D and 3-D *Q* models are used to compute SEM seismograms, which are referred to as *synthetic* and *observed* seismograms, respectively. The seismograms are computed for path B shown in Fig. 2. Middle: source–receiver cross-sections of anelastic K_β adjoint kernels for ENV measurements for dispersion (upper plot) and amplitude (middle plot) terms. Their sum gives the total anelastic adjoint kernel for ENV measurements (bottom plot). Right: source–receiver cross-sections of anelastic K_β adjoint kernels for AMP measurements for dispersion (upper plot) and amplitude (middle plot) terms. Their sum gives the total anelastic adjoint kernel for AMP measurements (bottom plot).

3.3 *Q* kernels

We have so far presented elastic kernels based on various misfit functions. Amplitudes, however are also significantly affected by anelastic properties of Earth's interior. In Fig. 16, *Q* kernels based on ENV and classical AMP misfits are presented. To understand the behaviour of misfits for the anelastic case, we also show adjoint sources and kernels for dispersion and amplitude terms (first and second terms, respectively, in the anelastic adjoint source in eq. 52). *Q* kernels are constructed with the adjoint wavefield generated by the adjoint source given in eq. (52). The reference frequency ω_0 in the dispersion term is chosen to be 212 s, which is the central period related to the resolution of our simulations determined based on the maximum and minimum periods of the absorption band. We considered windowed *S* waves on transverse components computed in 1-D PREM using 1-D PREM *Q* and 3-D *Q* (Dalton *et al.* 2008) models representing *synthetic* and *observed* data, respectively. Since we fixed the wave speed model in both synthetic and observed seismograms, the entire amplitude anomaly may be attributed to anelastic structure. Both ENV and AMP kernels for the dispersion term show similarities with elastic AMP kernels. Note that this dispersion term depends on the selected reference frequency. The second term of the adjoint sources is responsible for the amplitude anomaly. The AMP kernel for this amplitude term is like a banana-doughnut TT kernel, as expected, since the Hilbert transform in the anelastic displacement wavefield makes it similar to elastic velocity seismograms used in TT adjoint sources. The ENV kernel is similar, however, time-dependent normalizations make the kernel broader. The total anelastic kernels are simply the sum of the dispersion and amplitude terms. As shown for the elastic case, the ENV misfit can easily be extended to complete seismograms. This is also the case for anelastic kernels.
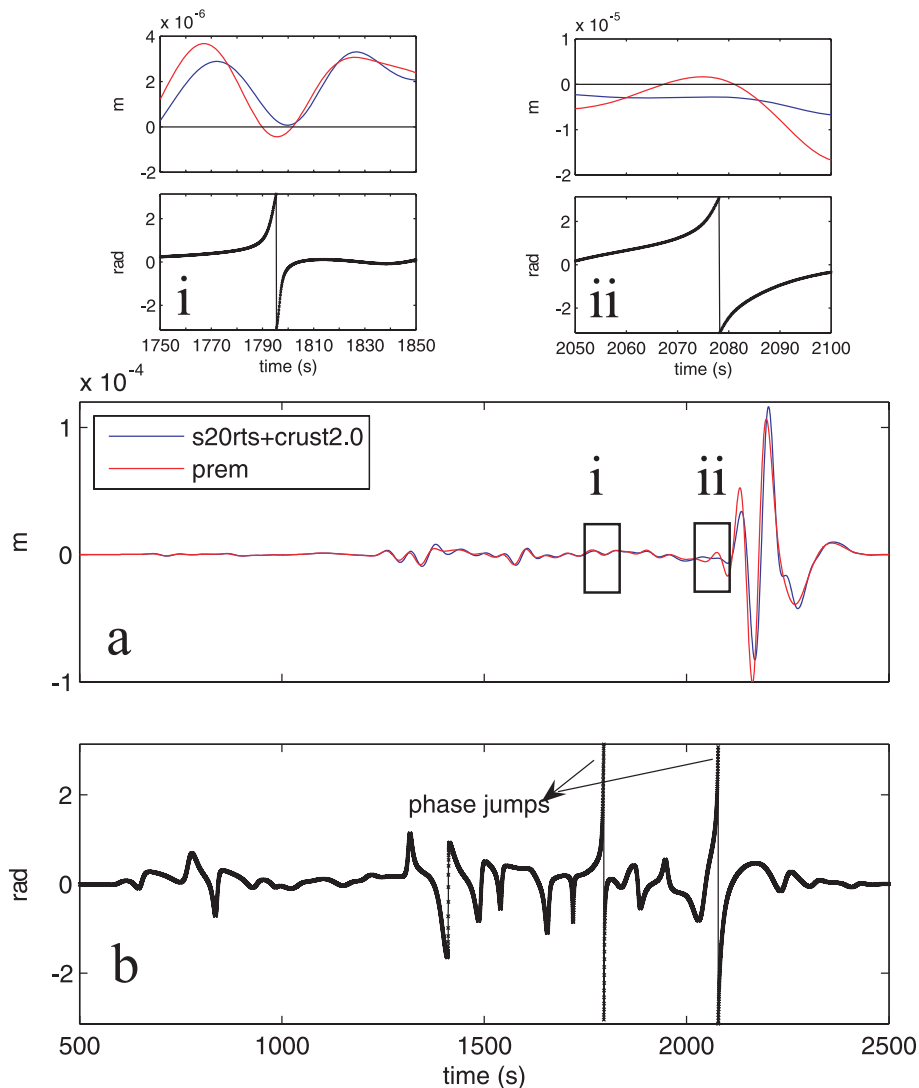

Figure 17. An example of phase jumps in instantaneous phase measurements. (a) SEM seismograms for path C in Fig. 2, bandpass filtered between 40 and 500 s. (b) Instantaneous phase difference between seismograms. In (i) and (ii), parts of the seismograms with phase jumps are enlarged.

4 DISCUSSION

All misfit functions determine their specific way in which particular measurements sense Earth's structure. Sensitivities for an isolated seismic phase, although not the same, show similar characteristics for all common misfit functions. For longer waveforms, different misfit functions exhibit quite different properties. WF favours the highest amplitude parts of seismograms only when a wave train is considered. The complementary IP and ENV misfits extract information from low-amplitude phases as well due to intrinsic weighting in adjoint sources while remaining numerically well behaved. Cross-correlation–based measurements, by construction, should only be applied to isolated seismic phases. All our simulations were based on noise-free data. Noise is easily dealt with for isolated phases since only phases above the noise level are selected. For WF, noise is also not a major issue since it focuses on highest amplitudes only. IP and ENV analyse phase and amplitude differences at each individual time step, independent of the overall strength of the signal at that particular time. Noise can thus affect small amplitude parts of the seismogram. We are going to discuss these issues in detail later.

4.1 Comparison of misfit functions

TT and AMP measurements give a robust estimation of phase and amplitude differences and characterize all information contained in waveforms if made at different frequencies. Their kernels are numerically well behaved and well documented (e.g. Luo & Schuster 1991; Marquering *et al.* 1999; Dahlen *et al.* 2000; Tromp *et al.* 2005; Tape *et al.* 2007). The drawback is that they need waveforms to be similar in shape and require isolating seismic phases from seismograms prior to measurements, and thus elaborate processing is required to pick every available phase in a seismogram. Recently, an automatic phase picking algorithm has been developed (e.g. Maggi *et al.* 2009) specifically tailored to adjoint tomography and has successfully been applied in a regional study (Tape *et al.* 2009).

  WF measurements, if applied to a whole wave train, favour higher amplitude parts of seismograms. Losing contributions of low-amplitude phases in sensitivity kernels can clearly be a disadvantage. We illustrated such adverse behaviour in Figs 14 and 15. To overcome this problem, it could be envisaged to define a misfit function based on relative rather than absolute waveform differences. This would lead to an expression similar to eq. (28), where envelopes are replaced by corresponding signals. Although eq. (28) is numerically well behaved, its waveform equivalent can be highly unstable due the many zero crossings of signals in the denominator. Similar to TT measurements, WF may also be used on isolated phases, where each phase is weighted by its energy to avoid small amplitude phases. However, this requires the use of a phase picking algorithm as well. The major disadvantage of WF comes from mixing phase and amplitude information in a single observable and is known to be highly non-linear with respect to Earth's structure (e.g. Gauthier *et al.* 1986; Luo & Schuster 1991).

  Working with the Hilbert transform is a natural way to separate phase and amplitude information. This can also be done in the time–frequency domain, as in the method proposed by Fichtner *et al.* (2008). The advantages of IP–ENV measurements in the time domain are less data processing and easier implementation. Separation of phase and amplitude is important to take full advantage of less non-linear phase information. Working on phase first is probably a sensible starting point for full waveform inversion. As demonstrated by Tape *et al.* (2010), matching phase alone improves amplitude misfit considerably. The similarity of our IP and ENV kernels confirms this. Adjusting the
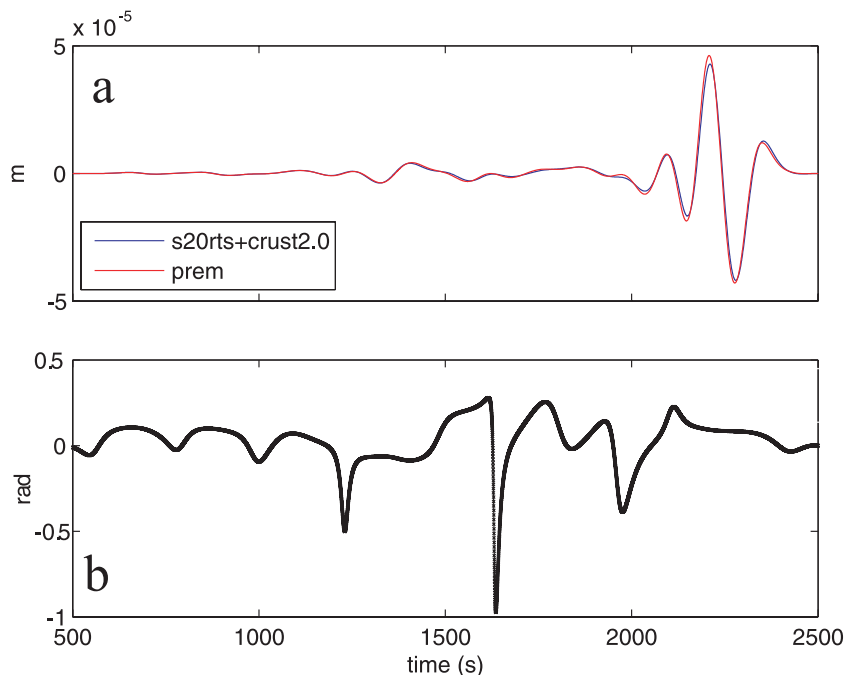

Figure 18. Same as for Fig. 17, but filtered between 100 and 500 s.

phase will also adjust elastic contributions in ENV misfits. As demonstrated in Section 3.3, ENV misfit can easily be extended to incorporate anelasticity. The remaining ENV misfit can then be adjusted by changing intrinsic attenuation. IP and ENV measurements thus contain complementary and seperable information on Earth structure, and remain numerically well behaved if applied to full waveforms.

4.2 Phase jumps in instantaneous phase measurements

In IP measurements, phase difference is measured between $\pm\pi$ rad. Thus, cycle skips occur whenever there is more than one cycle difference between observed and synthetic data. We present such an example in Fig. 17 (the seismograms are for path C in Fig. 3). We observe phase jumps between 1800 s and right before the surface waves around 2050 s. Since we are working within the limits of Born theory, our observed and synthetic data should differ by $\pm\pi/2$ rad at most. It is easy to detect such cycle skips in IP differences and eliminate these parts of seismograms by windowing, similar to Fichtner *et al.* (2008). To avoid cycle skip problems in phase speed measurements, Ekström *et al.* (1997) used long-period waveforms first, gradually increasing the frequency content of data in subsequent iterations in the inversion. We observe similar behaviour for our IP measurements (Fig. 18) and recommend working from long to short periods as iterations proceed.

4.3 Effect of noise

All numerical experiments so far have been based on noise-free synthetic seismograms. An important issue, however, in real cases is the noise content of data. Effects of noise should be small in the case of TT, AMP and WF measurements, since they favour high amplitude parts of data by only selecting high signal-to-noise seismic phases. IP and ENV measurements are amplitude normalized and every seismic phase has equal weight. This could potentially increase noise contributions in kernels compared to amplitude-dependent misfit functions. One obvious
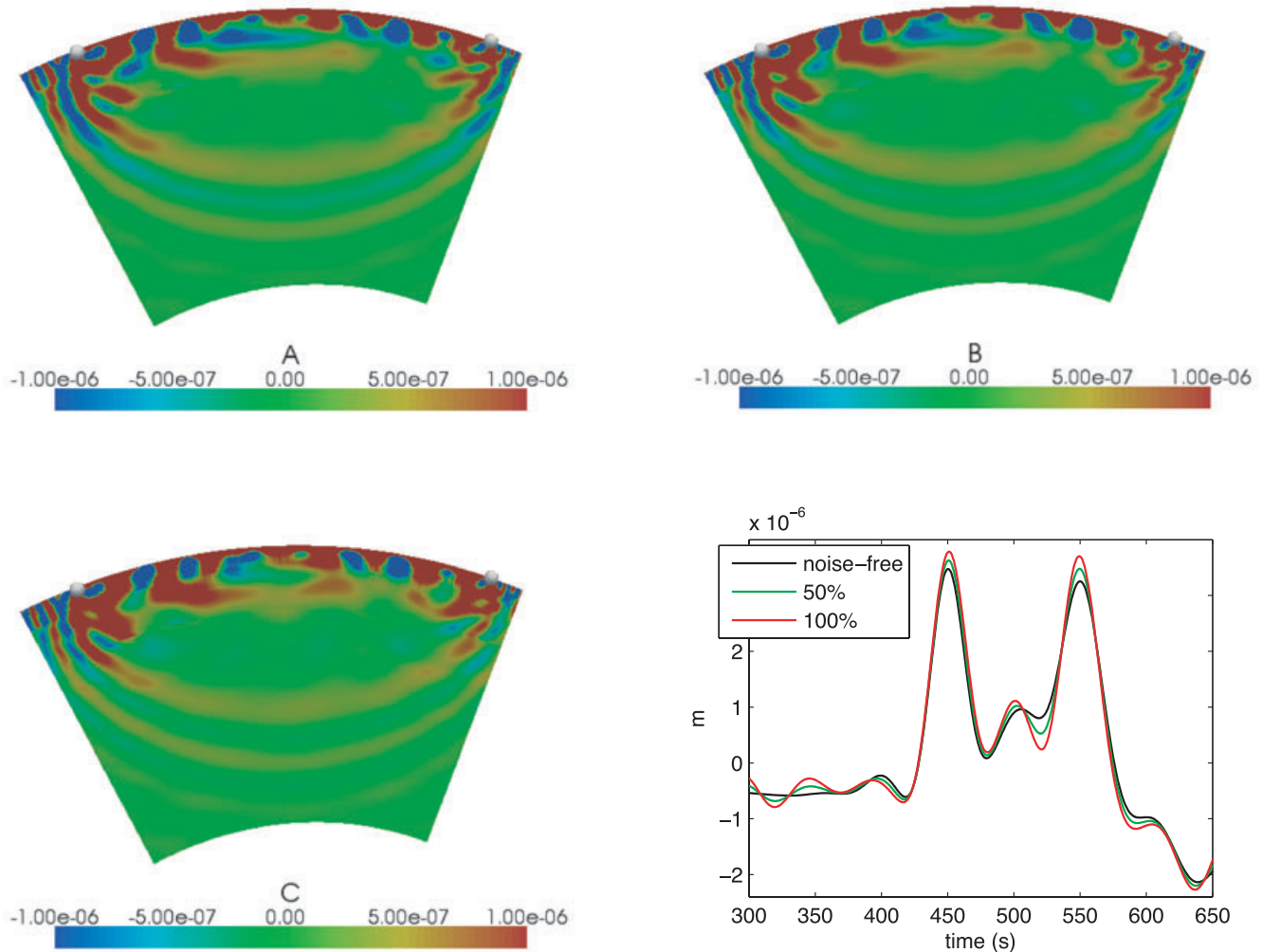

Figure 19. Effects of noise on sensitivity kernels computed for full seismograms presented in Fig. 3 using IP measurements. Uniformly distributed random noise is added to the observed data and bandpass filtered between 40 and 500 s (right-bottom plot). Noise is scaled to 50 per cent (B) and 100 per cent (C) of the maximum of the *P*-wave amplitude. In (A), the noise-free kernel is shown.

strategy is to isolate parts of the signal with a high signal-to-noise ratio (e.g. Maggi *et al.* 2009), but we would like to investigate the behaviour of kernels for full waveforms.

We tested effects of noise on IP measurements with the same seismograms used in Sections 3.1 and 3.2. We first generated uniformly distributed random numbers representing noise and scaled them to varying levels of *P*- and *S*-wave amplitudes. Then we added the scaled random noise to S20RTS+Crust2.0 seismograms, which played the role of *real* seismograms in both cases.

In the first example, we added random noise, scaled by 50 and 100 per cent of the maximum *P*-wave amplitude, to seismograms calculated in S20RTS+Crust2.0 (Fig. 3). The kernels computed with this noise are shown in Fig. 19. When the noise level increases, we start observing minor differences in kernels compared to the noise-free case, however overall sensitivities in kernels remain the same.

In the second example, random noise was added to seismograms calculated in S20RTS +Crust2.0 (Fig. 13). Remember that in this case the reference seismogram corresponds to S20TRS$_{um}$ + Crust2.0, and thus the upper-mantle part in both simulations is identical. We added random noise, scaled by 10 and 50 per cent of the maximum *S*-wave amplitude (Fig. 20). We observe that the higher the noise level, the stronger the amplitude of *S* kernels. High levels of noise also slightly changes the upper-mantle sensitivity. Please note that, for this example, the observed phase shift for the noise-free case is quite small, and thus perturbations due to noise become comparable to perturbations due to signal. However, overall sensitivity for *S* waves remains the same (Fig. 20C), and the kernel still recognizes that the data dominantly acquired a difference in the lower mantle. We do not observe any cases where significant spurious structures are added to the kernels due to noise in seismograms but a more detailed investigation needs to confirm this. We expect ENV measurements have similar behaviour to IP measurements, since ENV ratios have a similar normalizing term in the adjoint source. Defining an ENV misfit based on differences between envelopes of synthetic and observed data would be less sensitive to noise, however, in this way ENV kernels would be similar to WF kernels favouring higher amplitude parts of seismograms.
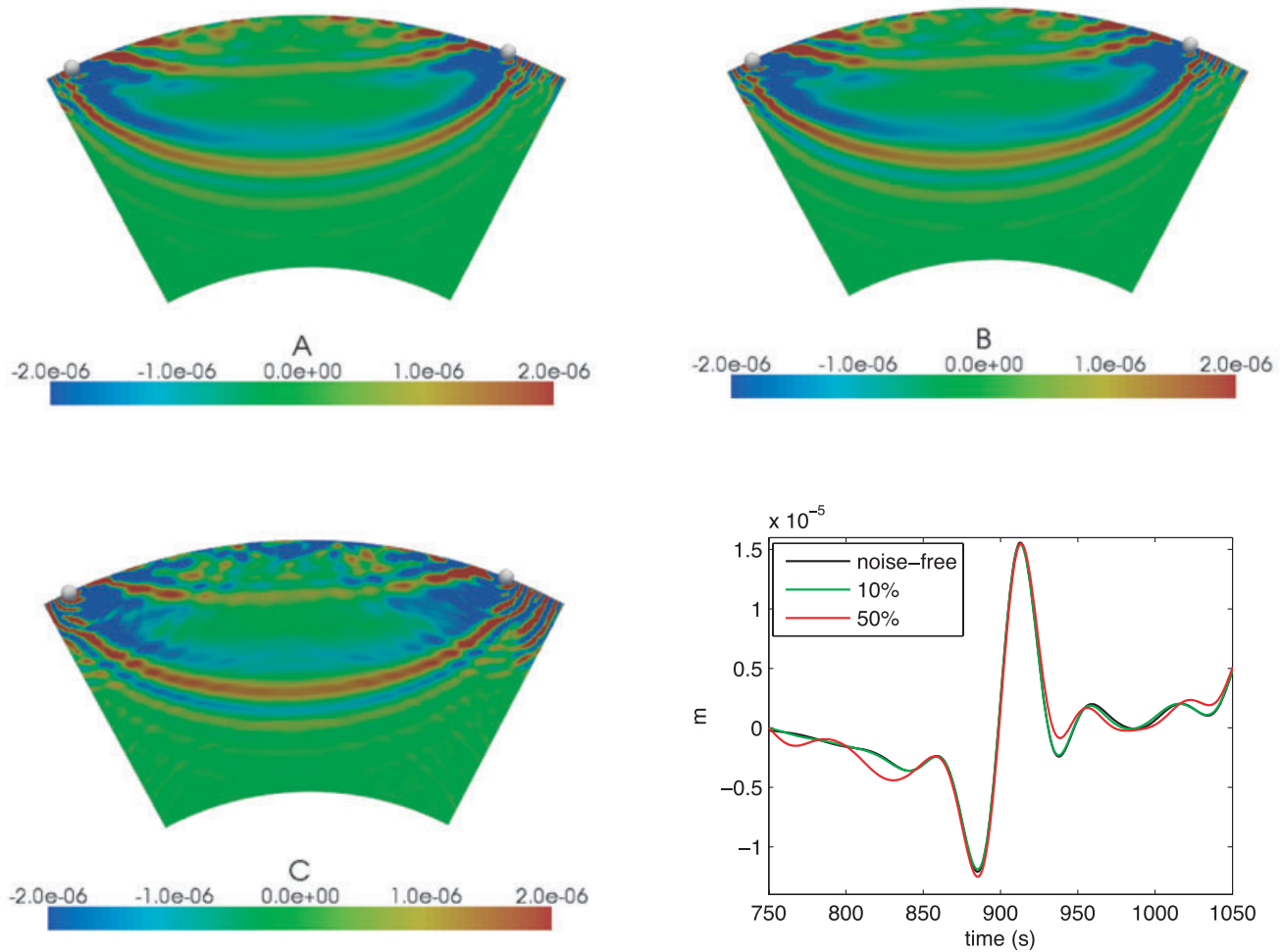

Figure 20. Effect of noise analysis on sensitivity kernels computed for full seismograms presented in Fig. 13 using IP measurements. Uniformly distributed random noise is added to the observed data and bandpass filtered between 40 and 500 s (right-bottom plot). Noise is scaled to 10 per cent (B) and 50 per cent (C) of the maximum of the *S*-wave amplitude. In (A), the noise-free kernel is shown.

5 CONCLUSIONS

We propose new misfit functions based on instantaneous phase differences and envelope ratios for full waveform tomography. We determine associated adjoint sources and examine their properties by computing finite-frequency adjoint kernels based on a spectral-element method. We compare the newly introduced misfit functions with more commonly used waveform, cross-correlation traveltime and classical amplitude misfit functions.

IP and ENV measurements are a natural way of separating phase and AMP information in data. They provide complementary information for full waveform inversion. In case of isolated seismic phases, kernel properties are not much different from those of commonly used measurements. Their potential lies in their application to full seismograms. Because of the normalizing weighting functions in their adjoint sources, all seismic phases are treated with equal importance. IP and ENV measurements remain numerically well behaved, and thus allow us to treat the complete seismogram without having to pick certain seismic phases.

IP measurements are prone to cycle skips, but this can easily be prevented by starting the inversion at long periods and gradually moving to higher frequencies in subsequent iterations. This is common practice in waveform-based phase speed measurements. In the presence of noise, kernels remain well behaved, and we have not identified cases where significant spurious structure is introduced into kernels. The cautious user can always isolate high signal-to-noise ratios in seismograms, which is easily accomplished by inspecting the envelope of the signal. Therefore, the IP–ENV combination is a promising approach in full waveform inversion.

ACKNOWLEDGMENTS

REFERENCES

Akçelik, V. *et al.*, 2003. High resolution forward and inverse earthquake modeling on terascale computers, in *Proc. ACM/IEEE Supercomputing SC'2003 Conference,* published on CD-ROM and at www.scconference.org/sc2003.

Aki, K. & Richards, P., 1980. *Quantitative Seismology, Theory and Methods,* W.H. Freeman, San Francisco, CA.

Barnes, A.E., 2007. A tutorial on complex seismic trace analysis, *Geophysics,* **72**(6), 33–43.

Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.,* **F897,** 81.

Bhattacharyya, J., Masters, G. & Shearer, P., 1996. Global lateral variations of shear wave attenuation in the upper mantle, *J. geophys. Res.,* **101,** 22 273–22 289.

Boschi, L. & Dziewonski, A., 2000. Whole Earth tomography from delay times of P, PcP and PKP phases: lateral heterogeneities in the outer core and radial anisotropy in the mantle?, *J. geophys. Res.,* **105**(B6), 13 675–13 696.

Boschi, L., Ampuero, J., Peter, D., Mai, P., Soldati, G. & Giardini, D., 2007. Petascale computing and resolution in global seismic tomography, *Phys. Earth planet. Inter.,* **163,** 245–250.

Capdeville, Y., Chaljub, E., Vilotte, J. & Montagner, J., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global Earth models, *Geophys. J. Int.,* **152,** 34–67.

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.

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

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

Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic non-linear waveform inversion: application to real data, *Geophysics,* **55,** 527–538.

Dahlen, F. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton University Press, Princeton, New Jersey.

Dahlen, F.A. & Baig, A., 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,** 157–174.

Dalton, C.A. & Ekström, G., 2006. Global models of surface wave attenuation, *J. geophys. Res.,* **111,** B05317, doi:10.1029/2005JB003997.

Dalton, C.A., Ekström, G. & Dziewonski, A.M., 2008. The global attenuation structure of the upper mantle, *J. geophys. Res.,* **113,** B09303, doi:10.1029/2007JB005429.

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

Ekström, G., Tromp, J. & Larson, E., 1997. Measurements and global models of surface wave propagation, *J. geophys. Res.,* **102,** 8137–8157.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2008. Theoretical background for continental and global scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.,* **175,** 665–685.

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

Fink, M., 1997. Time reversed acoustics, *Phys. Today,* **50,** 34–40.

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

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

Gu, Y.J., Dziewonski, A.M., Su, W.J. & Ekström, G., 2001. Models of mantle shear velocity and discontinuities in the pattern of lateral heterogeneities, *J. geophys. Res.,* **106,** 11 169–11 199.

Gung, Y. & Romanowicz, B., 2004. Q tomography of the upper mantle using three-component long-period waveforms, *Geophys. J. Int.,* **157,** 813–830.

Hudson, J., 1977. Scattered waves in the coda of P, *J. Geophys.,* **43,** 359–374.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.,* **139,** 806–822.

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

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

Komatitsch, D. & Vilotte, J., 1998. The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Li, X. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.,* **101**, 22 245–22 272.

Liu, H.P., Anderson, D.L. & Kanamori, H., 1976. Velocity dispersion due to anelasticity: implications for seismology and mantle composition, *Geophy. J. R. astr. Soc,* **47**, 41–58.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.,* **96**, 2383–2397.

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

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime tomography, *Geophysics,* **56**, 645–653.

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

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.

Masters, G., Johnson, S., Laske, G. & Bolton, H., 1996. A shear velocity model of the mantle, *Philos. Trans. R. Soc. Lond. A,* **354**, 1385–1411.

Mégnin, C. & Romanowicz, B., 2000. The 3D shear velocity structure of the mantle from the inversion of body, surface and higher mode waveforms, *Geophys. J. Int.,* **143**, 709–728.

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

Nolet, G., 1987. Waveform tomography, in *Seismic Tomography, with Applications in Global Seismology and Exploration Geophysics,* pp. 301–322, ed. Nolet, G., Reidel, Dordrecht.

Perz, M., Sacchi, M. & O'Byrne, A., 2004. Instantaneous phase and the detection of lateral wavelet stability, *The Leading Edge,* **23**, 639–643.

Pratt, R., 1999. Seismic waveform inversion in the frequency domain. Part I: Theory and verification in a physical scale model, *Geophysics,* **64**, 888–901.

Reid, F., Woodhouse, J. & van Heijst, H., 2001. Upper mantle attenuation and velocity structure from measurements of differential s phases, *Geophys. J. Int.,* **145**, 615–630.

Ritsema, J., van Heijst, H. & Woodhouse, J., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science,* **286**, 1925–1928.

Ritsema, J., Rivera, L., Komatitsch, D., Tromp, J. & van Heijst, H., 2002. The effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios, *Geophys. Res. Lett.,* **29**, doi:10.1029/2001GL013831.

Romanowicz, B., 1995. A global tomographic model of shear attenuation in the upper mantle, *J. geophys. Res.,* **100**, 12 375–12 394.

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

Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.,* **1**, 458–462, doi:10.1038/ngeo231.

Su, W.J., Woodward, R.L. & Dziewonski, A.M., 1994. Degree 12 model of shear velocity heterogeneity in the mantle, *J. geophys. Res.,* **99**, 4945–4980.

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.

Taner, M.T., Koehler, F. & Sheriff, R.E., 1979. Complex seismic trace analysis, *Geophysics,* **44**(6), 1041–1063.

Tanimoto, T., 1995. Formalism for traveltime inversion with finite frequency effects, *Geophys. J. Int.,* **121**, 103–110.

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

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

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

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

Tarantola, A., 1987. *Inverse Problem Theory,* Elsevier, Amsterdam, the Netherlands.

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

Trampert, J. & Woodhouse, J., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.,* **122**, 675–690.

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

Tromp, J., Komatitsch, D. & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Commun. Comput. Phys.,* **3**, 1–32.

Wu, R. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics,* **50**, 582–595.

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

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

Zhou, H.W., 1996. A high resolution P wave model of the top 1200 km of the mantle, *J. geophys. Res.,* **101**, 27 791–27 810.

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

APPENDIX

In this appendix we show how the second term in gradients of IP and ENV misfits, presented in Sections 2.3 and 2.4, may be written in terms of perturbations in the displacement wavefield δs. Let us consider the second term in the gradients of IP and ENV misfits in eqs (25) and (32), namely

$$g_r^\phi \equiv - \int_0^T \left[\phi_r^{\text{obs}}(t) - \phi_r(t) \right] \frac{s_r(t)}{E_r(t)^2} \delta \mathcal{H}\{s_r(t)\} \, dt, \tag{A1}$$

$$g_r^E \equiv \int_0^T \ln \left[\frac{E_r^{\text{obs}}(t)}{E_r(t)} \right] \frac{\mathcal{H}\{s_r(t)\}}{E_r(t)^2} \delta \mathcal{H}\{s_r(t)\} \, dt. \tag{A2}$$

Let us define the Hilbert transform of $s_r(t)$ as

$$\mathcal{H}\{s_r(t)\} = h(t) * s_r(t), \tag{A3}$$

where

$$h(t) \equiv -\frac{1}{\pi t}. \tag{A4}$$

Note that $h(-t) = -h(t)$. We then have

$$
\begin{aligned}
\int f_i(t)\delta\mathcal{H}\{s_i(t)\}\mathrm{d}t &= \int f_i(t) \int h(t-t')\delta s_i(t')\,\mathrm{d}t'\mathrm{d}t \\
&= \int \int h(t-t')f_i(t)\mathrm{d}t\,\delta s_i(t')\,\mathrm{d}t' \\
&= \int \int h(-(t'-t))\,f_i(t)\,\mathrm{d}t\,\delta s_i(t')\,\mathrm{d}t' \\
&= -\int \int h(t'-t)\,f_i(t)\,\mathrm{d}t\,\delta s_i(t')\mathrm{d}t' \\
&= -\int \mathcal{H}\{f_i(t)\}\,\delta s_i(t)\,\mathrm{d}t.
\end{aligned}
\tag{A5}
$$

Then eq. (A1) and (A2) become respectively

$$g_r^\phi \equiv \int_0^T \mathcal{H}\left\{\left[\phi_r^{\mathrm{obs}}(t) - \phi_r(t)\right]\frac{s_r(t)}{E_r(t)^2}\right\}\delta s_r(t)\,, \tag{A6}$$

$$g_r^E \equiv -\int_0^T \mathcal{H}\left\{\ln\left[\frac{E_r^{\mathrm{obs}}(t)}{E_r(t)}\right]\frac{\mathcal{H}\{s_r(t)\}}{E_r(t)^2}\right\}\delta s_r(t)\,. \tag{A7}$$