Geophysical Journal International

Geophys. J. Int. (2022) **230,** 1065–1079
Advance Access publication 2022 March 09
GJI Seismology

https://doi.org/10.1093/gji/ggac098

Sensitivity kernels for receiver function misfits in a full waveform inversion workflow

J.H.E. de Jong [1], H. Paulssen,[1] T. van Leeuwen [2,3] and J. Trampert [1]

[1] *Department of Earth Sciences, Utrecht University, Princetonlaan* 8A, 3584 *CB, Utrecht, the Netherlands. E-mail: j.h.e.dejong@uu.nl*
[2] *Mathematical Institute, Utrecht University, Budapest* 6, 3548 *CD, Utrecht, the Netherlands*
[3] *Centrum Wiskunde and Informatica, Science Park* 123, 1098 *XG, Amsterdam, the Netherlands*

Accepted 2022 March 7. Received 2022 March 4; in original form 2021 August 20

SUMMARY

Receiver functions have been used for decades to study the Earth's major discontinuities by focusing on converted waves. Deconvolution, which is the mathematical backbone of the method, is assumed to remove the source time function and the far-field dependence on structure, making it a useful method to map the nearby Earth structure and its discontinuities. Ray theory, a plane incoming wavefield, and a sufficiently well-known near-receiver background velocity model are conventionally assumed to map the observations to locations in the subsurface. Many researchers are aware of the shortcoming of these assumptions and several remedies have been proposed for mitigating their consequences. Adjoint tomography with a quasi-exact forward operator is now within reach for most researchers, and we believe is the way forward in receiver function studies. A first step is to calculate adjoint sensitivity kernels for a given misfit function. Here, we derive the adjoint source for a receiver function waveform misfit. Using a spectral element forward code, we have calculated sensitivity kernels for *P*-to-*S* converted waves using several 2-D models representing an average crust with an underlying mantle. The kernels show profound differences between *P*- and *S*-wave speed sensitivity. The sensitivity to *P*-wave speed is wide-ranging and related to the scattered *P*-wavefield which interferes with that of the *P*-to-*S* converted wave. The *S*-wave speed sensitivity is more local and mostly associated to potential locations of *P*-to-*S* conversion, although more distant sensitivity is also observed. Notably, there is virtually no sensitivity to impedance. We further observe the well-known trade-off between depth of the discontinuity and wave speed, but find that considering a longer waveform that includes more surface reverberations reduces this trade-off significantly.

Key words: Waveform inversion; Body waves; Computational seismology.

1 INTRODUCTION

For more than 50 yr converted waves have been used to map Earth's major internal discontinuities (Båth & Stefánsson 1966; Vinnik 1977; Langston 1979). A phase that converts from a *P* to an *S* wave, or vice versa, when it encounters a sudden jump in wave speed or density can be used in several ways to identify and locate discontinuities. Examples are mapping of the Moho, lithosphere–asthenoshere boundary, the 410- and 660-km discontinuities, and subducting slabs (Owens & Zandt 1985; Regnier 1988; Zhang & Langston 1995; Dueker & Sheehan 1997; Bostock 1998; Rychert *et al.* 2005). The depth or topography of these sharp boundaries provides important information on compositional transitions (Owens & Zandt 1985; Zhang *et al.* 2010) or temperature variations (Dueker & Sheehan 1997; Schmandt *et al.* 2012).

Receiver functions (from here onward generally referred to as RF(s) in equations and figures) were developed to identify waves that convert close to the receiver (Phinney 1964; Vinnik 1977; Langston 1979). Originally, they were often used to model the 1-D-structure beneath individual stations (Owens & Zandt 1985; Zhang & Langston 1995; Rychert *et al.* 2005), more recently they have also been used for the 2- or 3-D mapping of discontinuities (Zhang *et al.* 2010; Schmandt *et al.* 2012; Van Stiphout *et al.* 2019). The receiver function technique relies on the decomposition of the displacement components, $u_i(\mathbf{x}_r, t)$ $(i = V, R, T$, i.e. vertical, radial, transverse), at receiver location $\mathbf{x}_r$ into

separate terms that are convolved with each other. These terms are the source time function $S(t)$, a propagator term $P_i(t)$ that describes the time dependence of wave propagation through the Earth up to the near-receiver region, a near-receiver Earth-response $E_i(\mathbf{x}_r, t)$ that represents the final propagation to the receiver, and an instrument response term $I(t)$:

$$u_i(\mathbf{x}_r, t) = S(t) * P_i(t) * E_i(\mathbf{x}_r, t) * I(t). \tag{1}$$

A receiver function is a deconvolution of one spatial component of the displacement with another to achieve cancellation of some of the aforementioned terms while retaining the near-receiver response. To this end, it is advantageous to use direct P wave and its coda for receiver function analysis because the near-receiver structure strongly contributes to the P-wave coda. Moreover, for teleseismic P waves its coda is usually well separated from other main phases. When the data are partitioned into vertical, radial and transverse (V, R and T) components, radial P-to-S receiver functions are defined as:

$$\begin{aligned} \mathrm{RF}(\mathbf{x}_r, t) &= F^{-1}\left[\frac{\widehat{u}_R(\mathbf{x}_r, \omega)}{\widehat{u}_V(\mathbf{x}_r, \omega)}\right] \\ &= F^{-1}\left[\frac{\widehat{S}(\omega)\widehat{P}_R(\omega)\widehat{E}_R(\mathbf{x}_r, \omega)\widehat{I}(\omega)}{\widehat{S}(\omega)\widehat{P}_V(\omega)\widehat{E}_V(\mathbf{x}_r, \omega)\widehat{I}(\omega)}\right] \end{aligned} \tag{2}$$

where the function $\widehat{u}_i(\omega) = F[u_i(t)]$ denotes the temporal Fourier transform F of u_i, and F^{-1} its inverse. We see that the source time function and instrument response cancel out by this procedure. An additional assumption that is commonly made to further simplify receiver function interpretation concerns the P_i-terms. The far-field propagator terms are assumed to be approximately equal, $P_R(t) \approx P_V(t)$ because in this time window, most of the arriving energy will have travelled roughly the same path. The receiver function definition then simplifies to:

$$\mathrm{RF}(\mathbf{x}_r, t) \approx F^{-1}\left[\frac{\widehat{E}_R(\mathbf{x_r}, \omega)}{\widehat{E}_V(\mathbf{x}_r, \omega)}\right] \tag{3}$$

For near-vertical P-wave incidence, the radial component of the P-wave coda predominantly encapsulates the primary P-to-S conversion, as well as later reverberations, reflected or converted near the receiver. The deconvolution within the receiver function will highlight these arrivals. Similar reasoning can be followed for S-to-P conversions that arrive as precursors to the direct S wave (Båth & Stefánsson 1966; Langston 1979; Farra & Vinnik 2000). In this case, the vertical component is deconvolved by the radial. Alternative partitioning schemes of the seismic data might be used, such as the L-Q-T scheme (Vinnik 1977) where L is in the direction parallel to the incoming P wave and Q is perpendicular to it in the radial-vertical plane. In the main body of this paper, we will only consider P-to-S receiver functions in the V-R-T coordinate system.

To interpret receiver functions, the observations are inverted to extract information on the local depth of a discontinuity. Waveform inversions of receiver functions are generally carried out for teleseismic earthquakes where the near-receiver response, $E_i(\mathbf{x}_r, t)$, is modelled for an incoming plane wave of constant slowness. Synthetic P-wave receiver functions are subsequently calculated (using eq. 3) and compared to the observed receiver functions (calculated by eq. 2). This approach has been used to model the receiver-side crust in terms of 1-D structure (Piana Agostinetti & Malinverno 2010), dipping layers (Zhang & Langston 1995) or to include the effects of anisotropy (Levin & Park 1997). To improve their signal-to-noise ratio, receiver functions are usually stacked. For a single (or widely spaced) station set-up, stacking can be performed per receiver for events with similar slowness (Kind *et al.* 1995), after application of a moveout correction for events with varying slowness values (Chen & Niu 2013), or for similar backazimuths in case of lateral variations or anisotropy (Kosarev *et al.* 1993; Farra & Vinnik 2000). When using a relatively dense array of receivers, common conversion point (CCP) stacking can be applied (Dueker & Sheehan 1997; Kind *et al.* 2002), a technique analogous to common midpoint stacking in reflection seismics that stacks rays measured at different receivers that convert at the same location. The technique relies on ray tracing, often in a 1-D reference medium, and back projection of P-to-S (or S-to-P) receiver function signals. Whereas the method produces realistic images of lateral variations of the discontinuities, it can wrongly locate conversion points due to dipping discontinuities or lateral velocity variations.

To improve upon this, various body-wave migration techniques were developed in the last two decades. Ryberg & Weber (2000) and Poppeliers & Pavlis (2003) performed acoustic (scalar) and elastic (vector) receiver function migration, but implicitly assumed that the receiver function signals are only generated by primary P-to-S conversion. This results in images that may be contaminated by converted multiples which are treated as primary conversions from artificial interfaces at greater depths. Wilson & Aster (2005) partly resolved this by including the backscattered waves from the free surface reflected wavefield. Only recently, in a study by Cheng *et al.* (2016), were all scattered and converted waves considered in 3-D. It is important to note that the aforementioned migration algorithms measure scattering strengths (Pavlis 2005) that are subsequently related to the perturbations in the material properties by point scattering (Rondenay 2009). Migration methods that are directly based on elastic wavefield scattering are more appropriate to image the material property perturbations. Bostock *et al.* (2001) and Bostock (2002) use the forward scattered P and SV wavefields from the upgoing incident wavefield as well as the backscattered response generated by its free surface reflection, and employ the generalized Radon transform or Kirchhoff migration for back projection. Liu & Levander (2013) extended their approach to 3-D, but only for scattering of the incident wavefield. A drawback of the use of the P and SV wavefields, instead of receiver functions, is that it requires estimates of the source time functions of the incident wavefield (Rondenay *et al.* 2001; Shang *et al.* 2012).

In summary, most methods use several simplifying techniques and assumptions. (i) They all assume a simple, straight, incoming plane wave of constant slowness. (ii) They use a ray-theoretical approximation to determine the wave path and the conversion point. (iii) They use a previously known background model to project the observations to the scatter location. These assumptions might strongly influence the inferences from receiver functions as (i) the incoming wave front is not necessarily a plane wave and will contains other scattered phases, near-source reverberations, and the source radiation pattern. (ii) Ray theory does not account for frequency-dependent effects where the entire Fresnel zone contributes to the waveform. (iii) Small-scale changes of the velocity model might potentially have an equally large effect on the receiver functions as the topography of the discontinuity itself. Some attempts have been made to mitigate these assumptions. For example, Lekic *et al.* (2011) have combined ray theory with Fresnel zone smoothing to account for the fact that the whole Fresnel zone contributes to the scattered wavefield. Van Stiphout *et al.* (2019) used several velocity models to draw more robust conclusions on the discontinuity topography.

Contrary to stacking and migration methods that are based on high-frequency approximations (ray tracing), full waveform inversion (FWI) includes the full physics of the problem. FWI attempts to fit the observed data to a detailed model of the subsurface by comparing it to synthetic data and iteratively updating the model. For high-frequency ($\sim$1 s) body waves from teleseismic earthquakes used in receiver functions the problem may become computationally very expensive. Tong *et al.* (2014) and Monteiller *et al.* (2013) have therefore used a hybrid method which couples the wavefield of a 1-D global model to that of the 3-D regional structure. This technique allows for efficient modelling of teleseismic events up to a relatively high frequency (Pienkowska et al 2020; Wang *et al.* 2021a) and has been used to invert vertical and horizontal component waveforms (Monteiller *et al.* 2015; Beller *et al.* 2018; Wang *et al.* 2021b). Hansen & Schmandt (2017) have chosen to use ray theory and point scattering for the forward problem in order to calculate sensitivity kernels of receiver function misfits. With scattering only from the direct incoming wave and the free surface reflected *P* and *S* wavefields, the underlying physics is similar to Bostock *et al.* (2001) but is now applied to receiver functions.

We propose to analyse receiver functions by applying the adjoint method described by Tarantola (1984a), Plessix (2006), Tromp *et al.* (2005) and Fichtner *et al.* (2006). We will calculate sensitivity kernels for receiver functions directly and investigate what model parameters contribute to their waveforms. Contrary to ray-theoretical method of Hansen & Schmandt (2017) and the hybrid technique of Tong *et al.* (2014) and Monteiller *et al.* (2015), we use the full forward and adjoint wavefield calculations to obtain sensitivity kernels of receiver function misfits. We are thus able to assess to what extend the receiver function truly approximates the near-receiver Earth response. Even though our method is computationally more expensive, it uses all the information stored in the wavefield, and therefore allows us a more rigorous analysis of receiver function sensitivity.

In Section 2, we derive the mathematical tools for calculating receiver functions' sensitivity kernels. In Section 3, several examples, illustrating relevant cases, are provided. This is followed by a discussion and final conclusions.

2 SENSITIVITY KERNELS FOR RECEIVER FUNCTIONS

2.1 The adjoint method

During any seismic inversion, one needs to minimize a misfit between the synthetic data $\mathbf{u}(\mathbf{x}_r, \mathbf{m}, t)$ in a given model $\mathbf{m}$ at position $\mathbf{x}_r$ and at time t and the corresponding observed data $\mathbf{d}$, written as $\chi\,(\mathbf{u}(\mathbf{x}_r, \mathbf{m}, t), \mathbf{d}(\mathbf{x}_r, t))$. The misfit quantifies the difference between the observed and synthetic data and its perturbation ($\delta\chi$) captures the influence of model perturbations ($\Delta\mathbf{m}$) on the synthetic data.

$$\delta\chi = \chi(\mathbf{m} + \Delta\mathbf{m}) - \chi(\mathbf{m}) \tag{4}$$

The Fréchet derivative, that is the derivative of the objective function with respect to the model parameters, usually expressed in the form of sensitivity kernels (see below) are needed to update the model. In principle, these derivatives could be calculated by finite differences, but considering that a model usually consists of a large amount of parameters and that it would require two full forward calculations to determine the derivative for each one of them, this is generally avoided. An alternative is the adjoint method, which allows for a drastic reduction in the number of calculations (Tarantola 1984a). The adjoint field is defined such that its interaction with the forward wavefield gives the derivatives of the objective function with respect to the model parameters (Fichtner *et al.* 2006; Liu & Tromp 2006). The adjoint field can be calculated as the backward propagation of a field generated by the adjoint source (Tromp *et al.* 2005). Only two calculations are required in total: one forward and one adjoint calculation where the adjoint sources are propagated into the medium at the receiver locations. The waveform adjoint source is defined as the derivative of the objective function, χ, with respect to the synthetic waveform, $\mathbf{u}$ (Tromp *et al.* 2005; Fichtner *et al.* 2006; Bozdağ *et al.* 2011):

$$f_i^\dagger(\mathbf{x}, \mathbf{x}_r, t) = \partial_{u_i}\chi(\mathbf{u}\,(\mathbf{x}_r, T-t), \mathbf{d}(\mathbf{x}_r, T-t))\,\delta(\mathbf{x}-\mathbf{x}_r) \tag{5}$$

or, with a change of variables,

$$f_i^\dagger(\mathbf{x}, \mathbf{x}_r, T-t) = \partial_{u_i}\chi(\mathbf{u}\,(\mathbf{x}_r, t), \mathbf{d}(\mathbf{x}_r, t))\,\delta(\mathbf{x}-\mathbf{x}_r) \tag{6}$$

where $f_i^\dagger$ denotes the ith component of the adjoint source, with † indicating that time is reversed in adjoint calculations (i.e the time variable is $T-t$). Throughout this work we assume that all forward calculations are for the time window $[0, T]$, and all adjoint calculations for $[T, 0]$.

The adjoint wavefield is the solution of the wave equation, denoted by operator $\mathcal{L}$, and excited by the adjoint source:

$$\mathcal{L}\mathbf{u}^{\dagger}(\mathbf{x}, T-t) = \mathbf{f}^{\dagger}(\mathbf{x}, T-t) \tag{7}$$

whereas the forward field is defined as:

$$\mathcal{L}\mathbf{u}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t) \tag{8}$$

for some seismic source $\mathbf{f}$. The sensitivity kernels for density ρ, the bulk modulus κ and the shear modulus μ are then defined as a time integral of the forward field and the backpropagated adjoint field (Tromp *et al.* 2005):

$$\begin{aligned}
K_{\rho}(\mathbf{x}) &= -\int_0^T \rho(\mathbf{x})\mathbf{u}^{\dagger}(\mathbf{x}, T-t)\partial_t^2\mathbf{u}(\mathbf{x}, t)dt \\
K_{\kappa}(\mathbf{x}) &= -\int_0^T \kappa(\mathbf{x})\left(\nabla\cdot\mathbf{u}^{\dagger}(\mathbf{x}, T-t)\right)(\nabla\cdot\mathbf{u}(\mathbf{x}, t))\,dt \\
K_{\mu}(\mathbf{x}) &= -\int_0^T 2\mu(\mathbf{x})\mathbf{D}^{\dagger}(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t)dt
\end{aligned} \tag{9}$$

where $\mathbf{D}$ represents the traceless strain deviator. In this study, we will use a linear combination of these kernels that correspond to the more widely used wave-speed parameters (Tromp *et al.* 2005):

$$\begin{aligned}
K_{\rho'} &= K_{\rho} + K_{\kappa} + K_{\mu} \\
K_{\alpha} &= 2\left(\frac{\kappa + \frac{4}{3}\mu}{\kappa}K_{\kappa}\right) \\
K_{\beta} &= 2\left(K_{\mu} - \frac{4\mu}{3\kappa}K_{\kappa}\right)
\end{aligned} \tag{10}$$

where α denotes the *P*-wave speed, β corresponds to the *S*-wave speed and ρ' represents the rescaled density. It is interesting to note that the ρ'-kernel also corresponds to the impedance kernel, the product of density and velocity (Zhu *et al.* 2009). These kernels together contribute to the misfit perturbation such that:

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

Intuitively, the sensitivity kernels or Fréchet derivatives thus tell us what locations in the modelled region potentially contribute to the misfit between the synthetic and observed data. They are used to update the synthetic model from one iteration to the next (Fichtner *et al.* 2006).

2.2 Adjoint source for receiver functions

The adjoint source depends on the choice of misfit (eq. 6). We thus need to define the objective function, χ, which we take as the least-squares difference between synthetic and observed receiver functions.

$$\begin{aligned}
\chi &= \frac{1}{2}\left\|\mathbf{RF}^{syn}(\mathbf{x}_r, \mathbf{m}) - \mathbf{RF}^{obs}(\mathbf{x}_r)\right\|^2 \\
&= \frac{1}{2}\left[\mathbf{RF}^{syn}(\mathbf{x}_r, \mathbf{m}) - \mathbf{RF}^{obs}(\mathbf{x}_r)\right]^T\left[\mathbf{RF}^{syn}(\mathbf{x}_r, \mathbf{m}) - \mathbf{RF}^{obs}(\mathbf{x}_r)\right]
\end{aligned} \tag{12}$$

where $\mathbf{RF}$ is the discretized receiver function represented in the time domain as a vector. The vector is denoted by the use of boldface. Contrary to the analytical derivation in the previous section, we will approach most of the mathematical analysis in this section using vectors.

A receiver function is a deconvolution which is defined as a simple division in the frequency domain (eq. 2). It is therefore more intuitive to also express the objective function (eq. 12) in the frequency domain. The misfit χ represents the total energy of the difference between synthetic and observed receiver functions. Recalling Parseval's theorem (Wider 1961), χ can then simply be expressed in the frequency domain as

$$\chi = \frac{1}{2}\left\|\widehat{\mathbf{RF}}^{syn}(\mathbf{x}_r, \mathbf{m}) - \widehat{\mathbf{RF}}^{obs}(\mathbf{x}_r)\right\|^2 \tag{13}$$

where $\widehat{\mathbf{RF}}$ stands for the discretized receiver function vector in the frequency domain. We can insert this into the Fourier transformed eq. (6):

$$\begin{aligned}\widehat{\mathbf{f}}_i^*(\mathbf{x}_r, \mathbf{x}) &= \partial_{\widehat{\mathbf{u}}_i} \chi\ \delta(\mathbf{x}-\mathbf{x}_r) \\ &= \partial_{\widehat{\mathbf{u}}_i}\left(\frac{1}{2}\left\|\widehat{\mathbf{RF}}^{syn}(\mathbf{x}_r, \mathbf{m}) - \widehat{\mathbf{RF}}^{obs}(\mathbf{x}_r)\right\|^2\right)\delta(\mathbf{x}-\mathbf{x}_r) \\ &= \left[\frac{\partial\widehat{\mathbf{RF}}^{syn}(\mathbf{x}_r, \mathbf{m})}{\partial\widehat{\mathbf{u}}_i(\mathbf{x}_r, \mathbf{m})}\right]^* \odot \left[\widehat{\mathbf{RF}}^{syn}(\mathbf{x}_r, \mathbf{m}) - \widehat{\mathbf{RF}}^{obs}(\mathbf{x}_r)\right]\delta(\mathbf{x}-\mathbf{x}_r) \\ &= \left[\frac{\partial\widehat{\mathbf{RF}}^{syn}(\mathbf{x}_r, \mathbf{m})}{\partial\widehat{\mathbf{u}}_i(\mathbf{x}_r, \mathbf{m})}\right]^* \odot \Delta\widehat{\mathbf{RF}}(\mathbf{x}_r, \mathbf{m},)\delta(\mathbf{x}-\mathbf{x}_r)\end{aligned} \tag{14}$$

Here $\widehat{\mathbf{f}}_i^*$ denotes the discretized adjoint source vector in the frequency domain, with the complex conjugate being equivalent to the time reversal in eq. (6), meaning that $\widehat{f}_i^*(\omega) \equiv F\left[f_i^\dagger(T-t)\right]$. The $\partial_{\widehat{\mathbf{u}}_i}$ represents the derivative with respect to the $\widehat{\mathbf{u}}_i$-vector, that is the discretized ith component of the wavefield in the frequency domain, and $\odot$ represents pointwise multiplication. For the radial and vertical components of the adjoint source, this means:

$$\begin{aligned}\widehat{\mathbf{f}}_R^*(\mathbf{x}) &= \left[\frac{1}{\widehat{\mathbf{u}}_V(\mathbf{x}_r, \mathbf{m})}\right]^* \odot \Delta\widehat{\mathbf{RF}}(\mathbf{x}_r, \mathbf{m})\delta(\mathbf{x}-\mathbf{x}_r) \\ \widehat{\mathbf{f}}_V^*(\mathbf{x}) &= \left[-\frac{\widehat{\mathbf{u}}_R(\mathbf{x}_r, \mathbf{m})}{\widehat{\mathbf{u}}_V^2(\mathbf{x}_r, \mathbf{m})}\right]^* \odot \Delta\widehat{\mathbf{RF}}(\mathbf{x}_r, \mathbf{m})\,\delta(\mathbf{x}-\mathbf{x}_r)\end{aligned} \tag{15}$$

Using the inverse Fourier transform, we can define the adjoint source in the time domain and use an appropriate time window ($W(t)$) to isolate the desired arrival's waveform.

$$\begin{aligned}f_R^\dagger(\mathbf{x}, T-t) &= W(t)F^{-1}\left[\frac{\Delta\widehat{RF}(\mathbf{x}_r, \mathbf{m}, \omega)}{\widehat{u}_V^*(\mathbf{x}_r, \mathbf{m}, \omega)}\right]\delta(\mathbf{x}-\mathbf{x}_r) \\ f_V^\dagger(\mathbf{x}, T-t) &= W(t)F^{-1}\left[-\frac{\Delta\widehat{RF}(\mathbf{x}_r, \mathbf{m}, \omega)\widehat{u}_R^*(\mathbf{x}_r, \mathbf{m}, \omega)}{\widehat{u}_V^{2*}(\mathbf{x}_r, \mathbf{m}, \omega)}\right]\delta(\mathbf{x}-\mathbf{x}_r)\end{aligned} \tag{16}$$

To be more explicit about the specific time and frequency dependence, the terms in eq. (16) are again expressed as scalar functions rather than the vector notation used previously.

2.3 Practical implementation

The spectral division of the receiver function is unstable for near-zero amplitudes of $\widehat{\mathbf{u}}_V(\mathbf{x}_r, \mathbf{m}, \omega)$. Consequently, there exist several methods to perform the deconvolution (Oldenburg 1981; Sipkin & Lerner-Lam 1992; Gurrola *et al.* 1995; Park & Levin 2000). One of the most widely used methods is that of a water-level deconvolution accompanied with a Gaussian low-pass filter (Langston 1979). In this method, which we will apply as well, the frequency-domain division is stabilized by multiplying both nominator and denominator by the complex conjugate of $\widehat{\mathbf{u}}_V$, and adding a small damping factor, ϵ to the denominator. The Gaussian low-pass filter with a width ω_a, $\widehat{G}(\omega) = e^{-\frac{\omega^2}{4\omega_a{}^2}}$, is implemented to smooth the receiver function and to minimize the effect of side lobes, where we use 0.5 Hz for ω_a. The receiver function is then defined as:

$$\widehat{\mathrm{RF}}(\mathbf{x}_r, \omega) = \frac{\widehat{u}_R(\mathbf{x}_r, \omega)\,\widehat{u}_V^*(\mathbf{x}_r, \omega)}{\widehat{u}_V(\mathbf{x}_r, \omega)\,\widehat{u}_V^*(\mathbf{x}_r, \omega) + \epsilon}\,\widehat{G}(\omega) \tag{17}$$

The adjoint source of the stabilized receiver function (eq. 17) is approached in the same manner as the original receiver function deconvolution. The relevant components of the adjoint source (eq. 16) are then given by:

$$\begin{aligned}f_R^\dagger(\mathbf{x}, T-t) &= W(t)\,F^{-1}\left[\frac{\Delta\widehat{\mathrm{RF}}(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_V(\mathbf{x}_r, \mathbf{m}, \omega)}{\widehat{u}_V(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_V^*(\mathbf{x}_r, \mathbf{m}, \omega) + \epsilon}\right]\delta(\mathbf{x}-\mathbf{x}_r) \\ f_V^\dagger(\mathbf{x}, T-t) &= W(t)\,F^{-1}\left[-\frac{\Delta\widehat{\mathrm{RF}}(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_R^*(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_V^2(\mathbf{x}_r, \mathbf{m}, \omega)}{\widehat{u}_V^2(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_V^{2*}(\mathbf{x}_r, \mathbf{m}, \omega) + \epsilon}\right]\delta(\mathbf{x}-\mathbf{x}_r)\end{aligned} \tag{18}$$

For our examples in the next section, we used a value for the damping factor ϵ equal to 1 per cent of the maximum value of the denominator ($[\widehat{u}_V(\mathbf{x}_r, \omega)\,\widehat{u}_V^*(\mathbf{x}_r, \omega)]$). For the adjoint sources, ϵ is set to 1 per cent of the maximum value of $[\widehat{u}_V(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_V^*(\mathbf{x}_r, \mathbf{m}, \omega)]$ and $[\widehat{u}_V^2(\mathbf{x}_r, \mathbf{m}, \omega)\,\widehat{u}_V^{2*}(\mathbf{x}_r, \mathbf{m}, \omega)]$ for the radial and vertical components, respectively.

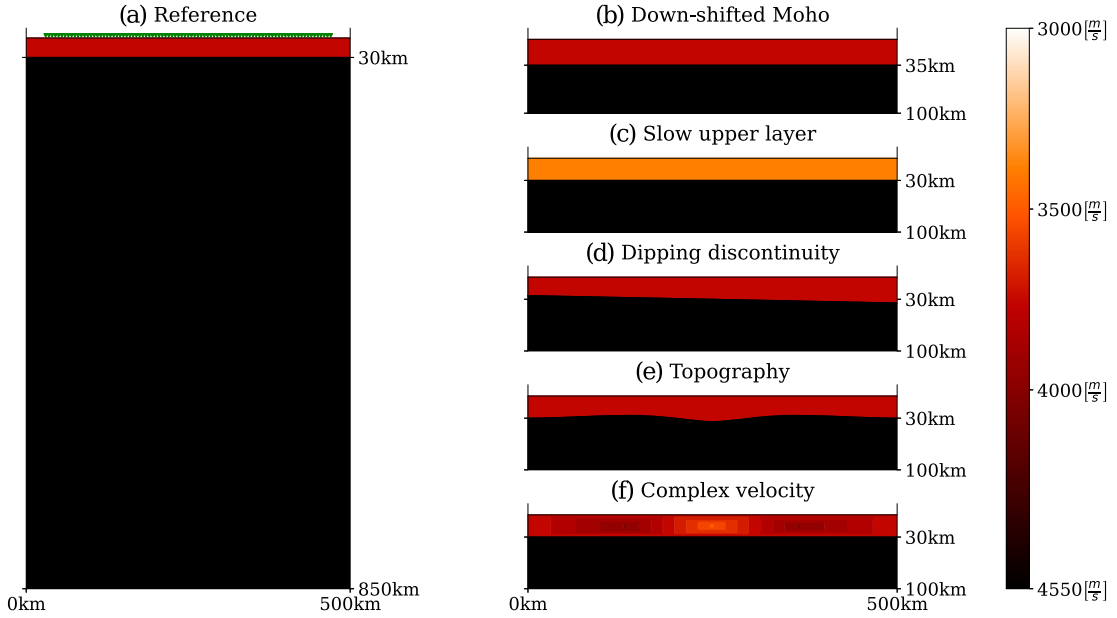

Figure 1. (a) The reference model with a horizontal Moho at 30 km depth and the 89-station receiver array in green. (b) Upper 100 km of a model with the Moho-discontinuity shifted 5 km downwards, to 35 km. (c) 100 km of model with a lower velocity in the upper layer. (d) Upper 100 km of the model with a discontinuity that dips toward the right. (e) Upper 100 km of a model with topography on the Moho discontinuity. (f) Upper 100 km of a model with a more complex, varying velocity in the crustal layer. The colours represent *Vs*-velocity. All the parameters and basic features are given in Table 1.

Table 1. Model parameters and description. 'ref.' implies the same value as the reference model. The parameters for the bottom layer are the same for all models: $\rho_{\text{bot}} = 3300$ [kg m^{-3}], $\alpha_{\text{bot}} = 7970$ [m s^{-1}] and $\beta_{\text{bot}} = 4550$ [m s^{-1}].

Model	ρ_{top} [kg m^{-3}]	α_{top} [m s^{-1}]	β_{top} [m s^{-1}]	Discontinuity
Reference	2800	6520	3760	Horizontal at 30 km
Downshift	ref.	ref.	ref.	Horizontal at 35 km
Slow upper layer	ref.	6420	3390	ref.
Dipping	ref.	ref.	ref.	Increasing, 25 to 35 km
Topography	ref.	ref.	ref.	Varying, 26.4 to 35 km
Complex velocity	ref.	6060–6980	3500–4020	ref.

3 RECEIVER FUNCTION SENSITIVITY KERNELS FOR 2-D MODELS

3.1 Synthetic models

To assess receiver functions' sensitivity, we have calculated adjoint sources and sensitivity kernels using the spectral element code SPECFEM2D (Tromp *et al.* 2008). Our simple models represent the Moho discontinuity that separates the crust from the mantle below with parameters that are loosely based on average values for the crust and upper mantle. The models are 500 km wide and 850 km deep. An array of 89 receivers, spaced 5 km apart, spans the surface starting at 30 km from the left (Fig. 1a). Sources with a source time function of the first derivative of a Ricker wavelet and with a dominant frequency of 0.84 Hz are put at a depth of 800 km or more to emulate wavefields arriving from below. In the reference model, for which the kernels are calculated, the Moho discontinuity is located at a depth of 30 km. The other models, which will provide the 'observed data', are subtle variations of this reference model. We initially focus on two of these perturbed models, Figs 1(b) and (c), that constitute a uniform, 1-D perturbation: in one of the these models the Moho is shifted downwards in its entirety to a depth of 35 km. In the other model the crustal velocity is uniformly lowered with respect to the reference model in such a way the first converted *P*-to-*S* wave (*Ps*) arrives at approximately the same time as in case of the shifted Moho. Additionally, we consider three extra synthetic models: one with a dipping discontinuity, one with some topography on the Moho discontinuity, and one with a more complex velocity in the crustal layer. Figs 1(d)–(f) show the upper 100 km of these models (the lower layer is again equal to the reference model). The parameter values and most important features of all the models are given in Table 1.

3.2 *Ps* sensitivity kernels for a single source–receiver pair

We first consider the sensitivity kernels of a single source–receiver pair. In Fig. 2, we show the synthetic traces, receiver functions and adjoint sources, using data from the models of Figs 1(a)–(c). Several arrivals can be observed in the traces and the receiver functions. The first arrival,

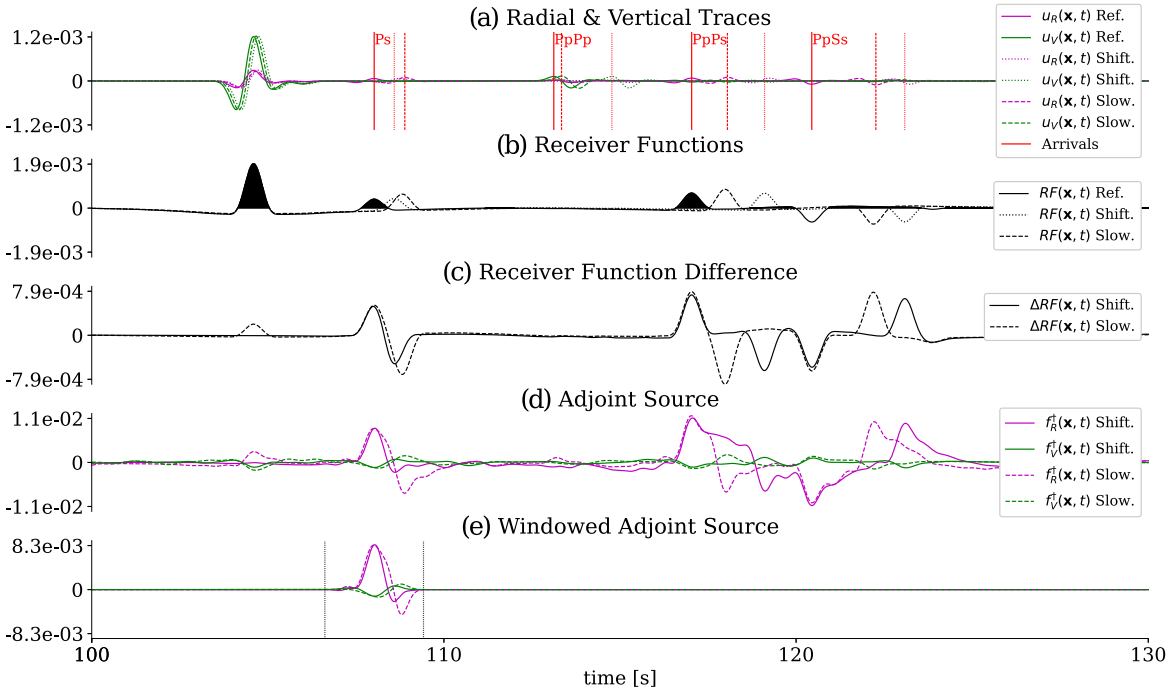

Figure 2. From top to bottom: (a) Traces for the reference model ('Ref', solid line; Fig. 1a), 'Downshifted Moho' ('Shift', dotted line; Fig. 1b), and 'Slow upper layer' ('Slow', dashed line; Fig. 1c). Vertical red lines indicate the arrival times of the *Ps*-, *PpPp*-, *PpPs*- and *PpSs*-phases (solid, dashed and dotted for the different models). (b) Receiver functions. (c) Receiver function difference between the reference model for the 'Downshift' model ('Shift', solid line) and the 'Slow upper layer' ('Slow', dashed line). (d) The adjoint sources for the 'Downshift' and the 'Slow upper layer' models. (e) Adjoint sources with a window applied to isolate the *Ps*-phase. Green lines represent the vertical components and magenta lines the radial components. The source is located at $x = 480$ km and $z = 800$ km, the receiver is located at $x = 275$ km.

the direct *P* wave, is the most pronounced. Next, the *Ps*-phase arrives. Several seconds later, the *PpPp*-phase that reflects at the free surface and subsequently back up at the Moho discontinuity, can be observed in the vertical trace (u_V), but in the receiver function this arrival is undetected. Two more arrivals are observed in the receiver function, the *PpPs*-, and the *PpSs*-phases. For these arrivals the difference between the two models of Figs 1(b) and (c) becomes more pronounced, as the arrival times diverge. The two perturbed models ('Downshift' and 'Slow upper layer') were chosen in such a way that their *Ps*-phases both arrive roughly a second later than the *Ps*-arrival in the reference model. The time window for the adjoint source is 2.8 s wide and chosen to include the *Ps*-arrivals of all the models.

The sensitivity kernels of Fig. 3 correspond to the adjoint sources of the *Ps* time window of Fig. 2 and show several interesting features. The α-kernels demonstrate the sensitivity to the *P*-wave speed. In them, we observe two striking features: (i) a strong sensitivity to an outer 'halo' and (ii) an 'inner sensitivity' around the *Ps* ray path. Because the kernels are built from the interaction of the forward and adjoint wavefield, we look at the forward and adjoint propagation in more detail. Fig. 4 shows snapshots of the amplitude of the wavefields at $t =$ 90 and 105 s. They illustrate that the outer 'halo' arises from the interaction of the forward *P* wavefield (*P*) with the *P* wavefield of the adjoint field (P_{adj}). In the β-kernels, some sensitivity to this outer sensitivity can be observed as well, albeit weaker than that of the α-kernels. This sensitivity to the *S*-wave speed can easily be understood from the fact that the β-kernel is a linear combination of both the κ- and μ-kernels (eq. 10), both of which influence *P*-wave propagation. Although this sensitivity is diminished in the β-kernel, it does not completely cancel out. In the top of the upper layer *P*-wave sensitivity is observed (Fig. 3b). This relates to scattering from the free-surface reflected *P* wave.

The 'inner sensitivity' around the *Ps* ray path can also be explained with Fig. 4. In the region below the discontinuity, it is produced by the interaction of the forward *P* wave (*P*) with the Sp_{adj} wave of the adjoint field, that is the adjoint *S* wave that converted to a *P* wave at the discontinuity. Thus, the forward *P* wave and the adjoint Sp_{adj} wave together produce the sensitivity to the *Ps* waveform. Alternatively, the *Ps* sensitivity in the upper layer is obtained by interaction of the *P*-to-*S* converted wave of the forward field *Ps* with the adjoint *S* wave S_{adj}, and can be predominantly observed in the β-kernel. In both the α- and β-kernels, we observe an asymmetry with significantly weaker sensitivity on the left side. This arises from the interaction of the forward and adjoint wavefield which are weaker on the left side due to a smaller scattering coefficient.

We observe that the sensitivity to impedance (ρ') is low. This suggests that receiver functions are not well suited to image impedance contrasts. Lastly, we note that the kernels of the Downshift model (Figs 3a and b) and the Slow upper layer model (Fig. 3c) are very similar. This indicates that it will be difficult to distinguish between velocity perturbations in the upper layer and perturbations in the depth of the discontinuity using receiver function Ps waveforms.

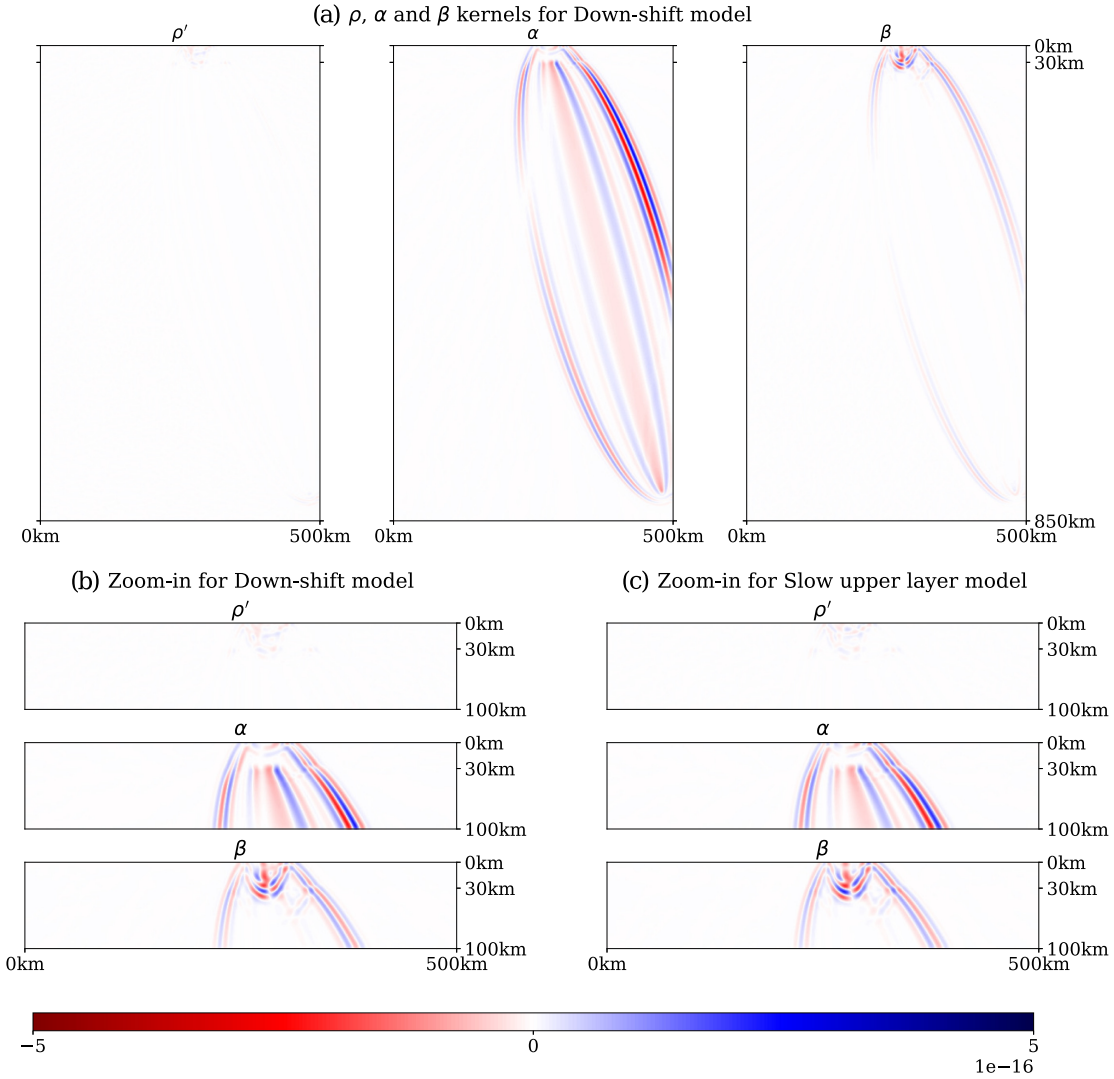

Figure 3. Sensitivity kernels for the same source–receiver pair as the data in Fig. 2 (source: $x = 480$ km, $z = 800$ km and receiver: $x = 275$ km). (a) and (b) Sensitivity kernels for the 'Downshifted Moho' model of Fig. 1(b) for impedance (ρ'), *P*-wave speed (α) and *S*-wave speed (β). (b) Zoom-in of the upper 100 km of (a). (c) Upper region of the sensitivity kernels for the 'Slow upper layer' model (Fig. 1c).

3.3 *Ps* sensitivity for multiple source–receiver pairs

Interpretations of receiver functions are often made using stacks of multiple source–receiver pairs, binned by station or CCP (Dueker & Sheehan 1997; Rondenay 2009). To emulate this, we modelled 10 sources for which we recorded the resulting seismograms at all 89 receivers of the surface array. For each source, we calculated the adjoint sources at all receivers, and used them to generate the full adjoint wavefield and calculate the sensitivity kernel. In adjoint calculations, the contribution to the misfit is back projected from the receivers into the entire subregion of the model space. So rather than back projecting the data to a CCP, we obtain the full subsurface sensitivity by simply summing the kernels of all source–receiver pairs. The summed kernels of the *Ps* time windows for the two models of Figs 1(b) and (c) are shown in Fig. 5.

Several interesting features of the sensitivity are present in these kernels. The β-kernel of the 'Downshift' model shows a strong sensitivity to the area around the discontinuity, particularly directly beneath it (Figs 5a and b). The blue area corresponds to the region of the model that belongs to the lower layer in the reference model and to the upper layer in the 'Downshift' model. The strong sensitivity to this region represents the downshift of the discontinuity and indicates that this region shifts from the fast lower layer to the slower upper layer. The β-kernel of the 'Slow upper layer' model shown in Fig. 5(c) is nearly identical to the kernel in Fig. 5(b). So instead of a decrease in velocity in the upper layer, the kernel could easily be misinterpreted as a downward shift of the discontinuity.

The α-kernels for the 'Downshift' and 'Slow upper layer' models are also highly similar (Figs 5b and c), and, analogous to the single source–receiver kernels (Fig. 3), there is a strong sensitivity to the model beneath the discontinuity. We also observe sensitivity related to backscattering of the free surface reflected *P* wav field within the upper layer (the blue zone in the middle of the crustal layer). Additionally, we note a strong dependence of the sensitivity on the incidence angle, that is, the angle of the wave propagation with the discontinuity

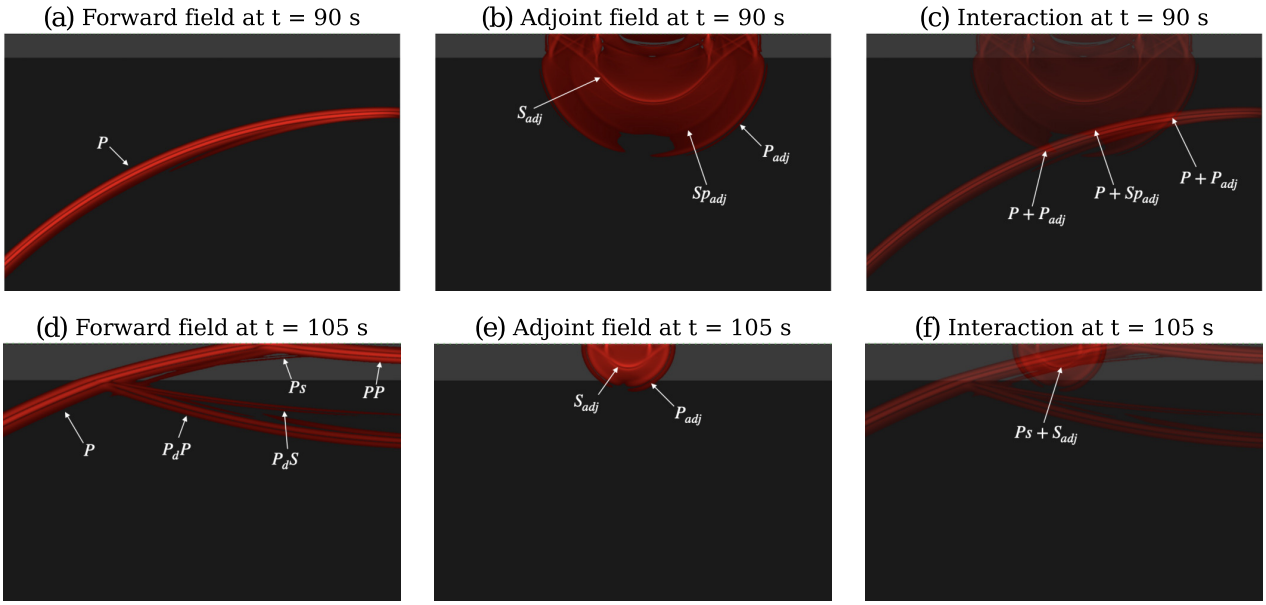

Figure 4. The normalized displacement of the forward and adjoint wavefields and their interaction at two snapshots: $t = 90$ and 105 s (windows for the later snapshots are slightly zoomed-in). The interaction panels highlight regions that have a high amplitude in the sensitivity kernels. (a)–(c) The outer sensitivity is produced by interaction of the forward P wavefield (P) with the adjoint P wavefield ($P_{\rm adj}$) giving $P + P_{\rm adj}$ sensitivity. Both the interaction of the forward P wave with the adjoint S-to-P converted wave ($P + Sp_{\rm adj}$) in (c) and the interaction of the forward P-to-S converted wave with the adjoint S wave ($Ps + S_{\rm adj}$) in (f) produce the Ps-scattering sensitivity.

normal. The kernels of source–receiver pairs with a source at the far left side and a receiver at the far right side (or vice versa) have larger amplitudes than more vertically aligned source–receivers pairs (Fig. 5a). These high incidence angle source–receiver pairs contribute more significantly to the overall sensitivity. This arises from the dependency of the P-to-S conversion coefficient in the incidence angle, resulting in a higher amplitude of the Ps-phase on the radial component for higher incidence angles, and thus a higher Ps-amplitude of the receiver function and a higher amplitude of the adjoint field. The ρ'-kernels again display a relatively weak sensitivity to impedance. Their sensitivity is predominantly confined to the upper region of the models, and is associated to reflections from the free surface and the Moho discontinuity with arrival times similar to that of the Ps phase.

Additionally, we calculated sensitivity kernels for the other models of Table 1. Interestingly, the β-kernels of in Figs 6(a) and (b) retrieve the dip and topography of the discontinuity of the models in Figs 1(d) and (e) very well. However, like before, both sets of kernels also show strong sensitivity to P wave speed below the discontinuity and S wave speed above the discontinuity. Moreover, the similarity of the kernels for the 'Topography' model (Fig. 6b) and the 'Complex velocity' model (Fig. 6c) suggests that it may be difficult to distinguish between upper layer velocity perturbations and perturbations in the depth of the discontinuity.

3.4 Multiple arrivals

Figs 5 and 6 illustrate that the ability to retrieve the input models with sensitivity kernels is improved by using more data, that is by using more receiver function Ps waveforms from multiple sources and receivers simultaneously. Another way of extending data is to include the multiples of the upper layer (Zhu & Kanamori 2000). The arrival times of two P-to-S converted multiples are marked by vertical red lines in the receiver functions of Fig. 2. These multiples are the $PpPs$-phase, the P-wave free surface reflection of the direct P wave that reflects back up as an S wave at the discontinuity, and the $PpSs$-phase, the S-wave free surface reflection of the direct P wave that reflect back up at the discontinuity. Fig. 7 shows the adjoint sources for a time window of roughly 20 s wide that include not just the Ps-phase, but also the $PpPs$- and $PpSs$-phase. The longer adjoint source increases the amount of information encapsulated within the sensitivity kernels, but it also increases their complexity, as the more complex adjoint field interacts with the forward wavefield.

When comparing the full coda kernels for multiple sources and receivers in Figs 8(a) and (b), we observe that they differ more than those in Fig. 5. By including the later arrivals, the kernels sample the up- and downward P or S legs of the $PpPs$- and $PpSs$ phases. We observe a stronger sensitivity to the upper layer in the α-kernels of Figs 8(a) and (b). However, the α-kernel in Fig. 8(b) shows a different sensitivity to the upper layer P-wave velocity than the kernel of Fig. 8(a); contrary to the α-kernel of the downshifted model, the α-kernel of the 'Slow upper layer' model indicates a decrease in P-wave speed. In the upper layer of the β-kernels in Fig. 8(a), we observe that the sensitivity to S-wave speed is strong in the region between 30 and 35 km depth, suggesting a significantly decrease in velocity there, that is, a downshift of the boundary. In the kernel of Fig. 8(b) ('Slow upper layer') the width of this strong sensitivity region is smaller, indicating a smaller downward shift of the layer boundary. Notably, we see the results of interactions between the direct P wave of the forward field and

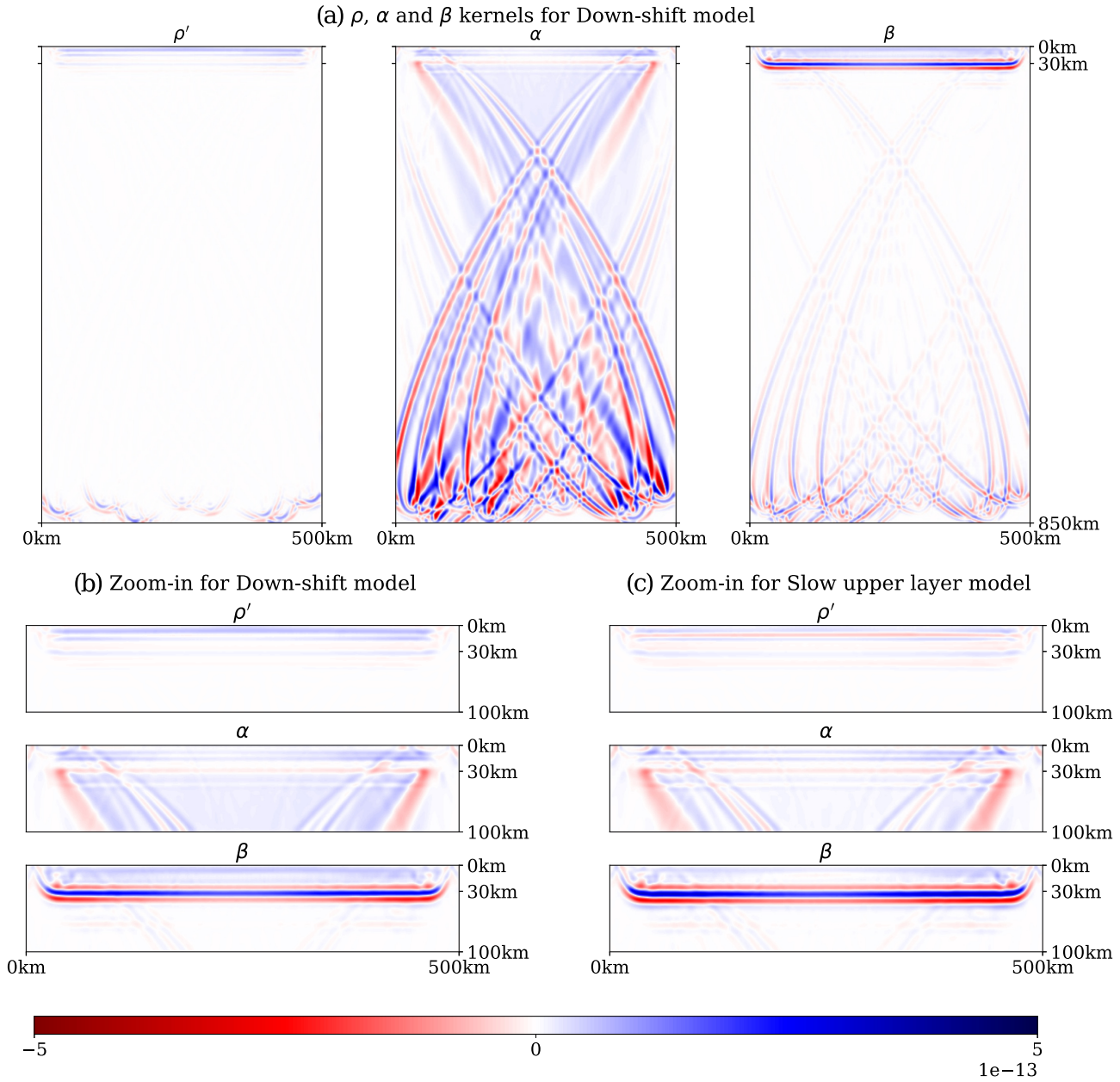

Figure 5. Sensitivity kernels for 10 sources and 89 receivers. The 10 sources are located at: (1) $x = 30$ km, $z = 820$ km, (2) $x = 70$ km, $z = 845$ km, (3) $x = 110$ km, $z = 810$ km, (4) $x = 165$ km, $z = 835$ km, (5) $x = 205$ km, $z = 840$ km, (6) $x = 250$ km, $z = 805$ km, (7) $x = 300$ km, $z = 815$ km, (8) $x = 395$ km, $z = 825$ km, (9) $x = 425$ km, $z = 830$ km, (10) $x = 480$ km, $z = 800$ km. (a) and (b) for the Downshift model of Fig. 1(b) and (c) for the Slow upper layer model of Fig. 1(c).

the various phases of the adjoint field in the lower layer at around 100 km depth, leading to 'echos' of the discontinuity in the β-kernels. The multiple arrival kernels for the 'Topography' and 'Complex velocity' model are shown in Figs 8(c) and (d), and will be considered in detail in the discussion.

4 DISCUSSION

In the Introduction of this paper, we described various methods that have been employed to extract information from waveforms using receiver functions. An important simplification that is applied in most studies is that of ray theory. However, our sensitivity kernels provide evidence that the use of ray theory to establish conversion point locations and from that infer discontinuity depth and topography can be misleading. It has previously been demonstrated that the infinite frequency approximation of ray theory is inaccurate when it comes to assessing the contribution of various model regions and that actually the whole Fresnel zone sensitivity needs to be considered, for example, the 'banana-doughnuts' introduced by, for example, Woodward (1992), Marquering *et al.* (1998, 1999), Dahlen *et al.* (2000) and

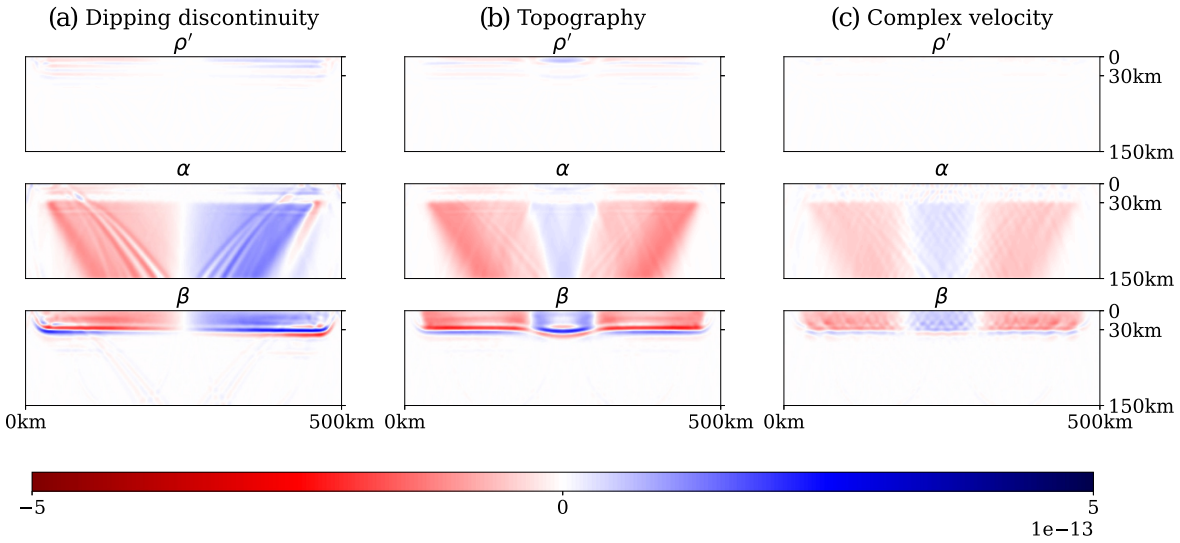

Figure 6. Sensitivity kernels for 10 sources and 89 receivers for the three final models (Figs 1d–f). Only the upper 150 km of the models is shown.

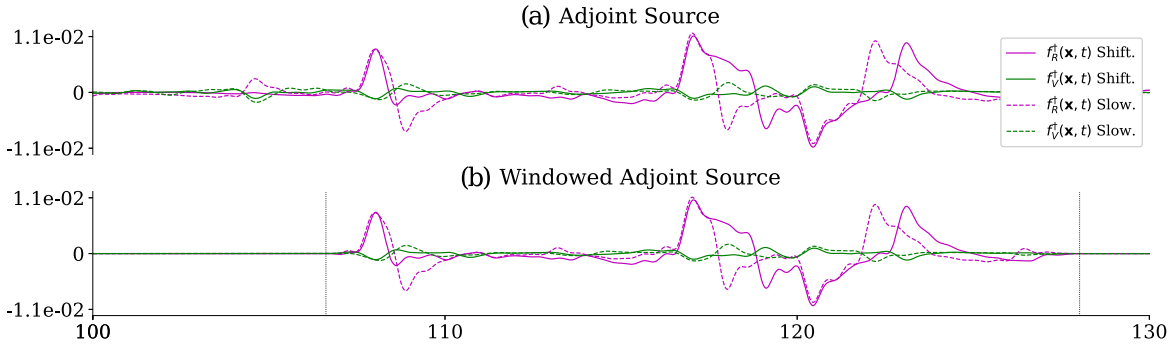

Figure 7. Same adjoint sources as in Fig. 2, now time-windowed to include the later arrivals of *PpPs* and *PpSs*.

Tromp *et al.* (2005). Our kernels demonstrate that the same holds for receiver functions. We show that the strongest sensitivity is not limited to the ray path or to the conversion point, but to a much wider region. This is illustrated in Fig. 9, where we compare the ray path of a single source-receiver pair to its α- and β-kernels. The sensitivity observed in these kernels shows the width of the region that contributes to this receiver function, and indicates that it is severely underestimated by the ray theoretical approximation.

Attempts have been made to account for this underestimation of sensitivity by 'spreading' the weight of receiver functions according to a (wavelength dependent) distance to the ray at the discontinuity (Lekic *et al.* 2011). The weighting factor is used to attribute the individual source–receiver pairs to multiple CC bins, allowing them to influence the inferred depth of a region on the discontinuity rather than a point. Fig. 9(e) compares this weighting factor to the sensitivity at the discontinuity depth. It is obvious that the weighting factor of CCP-stacking cannot directly be related to waveform sensitivity as these are different parameters. The weighting factor is an estimate of the first Fresnel zone at the discontinuity whereas our kernels indicate sensitivity to wave speed perturbations. The main point of comparison that we intent to stress is the width and shape of region that influences the observed receiver function. In Fig. 9(e), we observe the highest β-sensitivity at the discontinuity outside first Fresnel zone estimate. Additionally, we find significant α- and β-sensitivity even further from the conversion point that are related to the 'outer halo' observed in Fig. 3, indicating that the width of the influencing region is severely underestimated.

This 'outer halo' sensitivity in these single source–receiver kernels arises from the interaction of the primary P wave of the adjoint field with the primary P wave of the forward field ($P + P_{\text{adj}}$; Fig. 4). We therefore interpret this 'halo' as representing receiver functions' sensitivity to PP scatterers arriving in the same time window as the Ps phase. It implies that a velocity anomaly at the location of this 'halo' might affect the receiver function waveform within the Ps time window. When we consider kernels of the multiple sources and receivers (Figs 5 and 6), the relative contribution of the PP-scatterer sensitivity decreases. The contribution of the 'inner sensitivity' then dominates as it interferes constructively. This sensitivity is particularly well illustrated by the kernels of Fig. 6. It arises, as inferred from Fig. 4, from the interaction of the forward P wave with adjoint S-to-P conversion ($P + S_{\text{adj}}p$) and the interaction of the forward P-to-S conversion with adjoint S wave ($Ps + S_{\text{adj}}$). These two types of interaction produce different sensitivities above and below the discontinuity: a stronger α-sensitivity below the discontinuity and a dominant β-sensitivity above the discontinuity. This dominant β-sensitivity in the upper layer explains why receiver functions are inverted for crustal shear velocity (Owens *et al.* 1984).

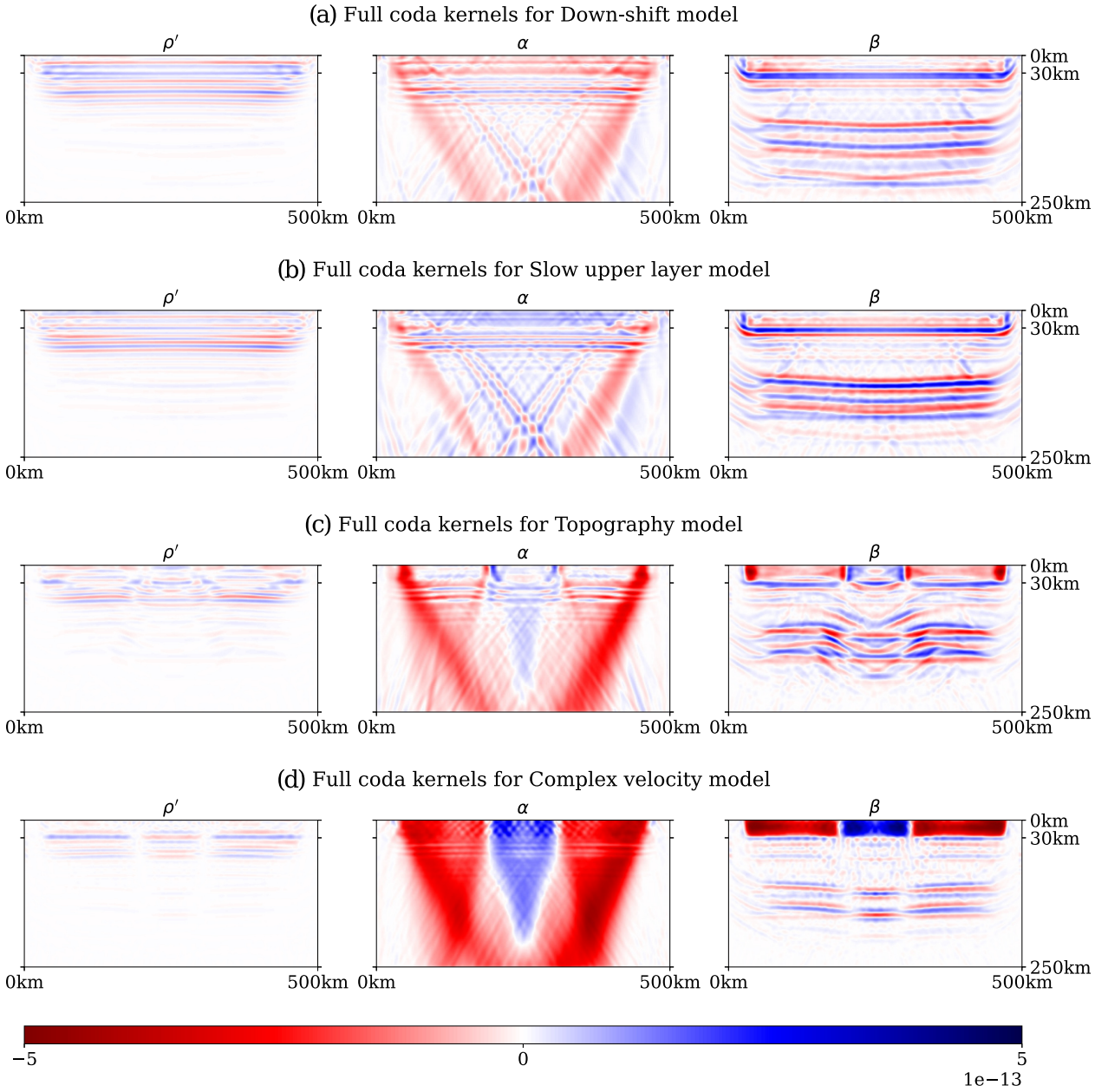

Figure 8. Upper 250 km of the multiple source–receiver kernels generated by the adjoint source time window that includes the *Ps*, *PpPs* and *PpSs* phases (Fig. 7) for four of the models shown in Fig. 1.

However, the α-sensitivity in the bottom layer is of the same magnitude as the β-sensitivity in the the upper layer. This is an important observation, because the basic idea behind receiver functions is that, apart from elimination of the source time function, the far-field contribution of the wavefield can be neglected. Yet, our results show that receiver functions are sensitive to the far-field *P*-wave structure. This means that the assumption $P_R(t) \approx P_V(t)$, which leads to the cancelling out of these terms in the deconvolution of eq. (3), may not be accurate. So while receiver functions will remove the source term and instrument response, one needs to be aware that the full wave path contributes to receiver functions, and that the distant velocity structure needs to be taken in consideration as well.

A well-known problem in receiver function inversions is the trade-off between discontinuity depth and velocity (Ammon *et al.* 1990). By comparing the sensitivity kernels of various models that were carefully chosen, we tried to investigate to what extend FWI might help to resolve this problem. Our first two models contained 1-D perturbations to the reference model: in the first we shifted the discontinuity downward, and in the second we uniformly changed the velocity in upper layer such that both types of perturbations had a very similar effect on the *Ps* waveform. Naturally, the resulting kernels for individual receiver functions (Fig. 3) and for the multiple receiver functions (Fig. 5) are very similar. This implies that it remains difficult to directly distinguish between 1-D depth and 1-D velocity changes in inversions. However, the sensitivity kernels of Fig. 6 show more pronounced differences. In these models 2-D perturbations were applied: either a gradual change in discontinuity topography or gradually changing velocities in the upper layer. Although the α-kernels of Figs 6(b) and (c) are

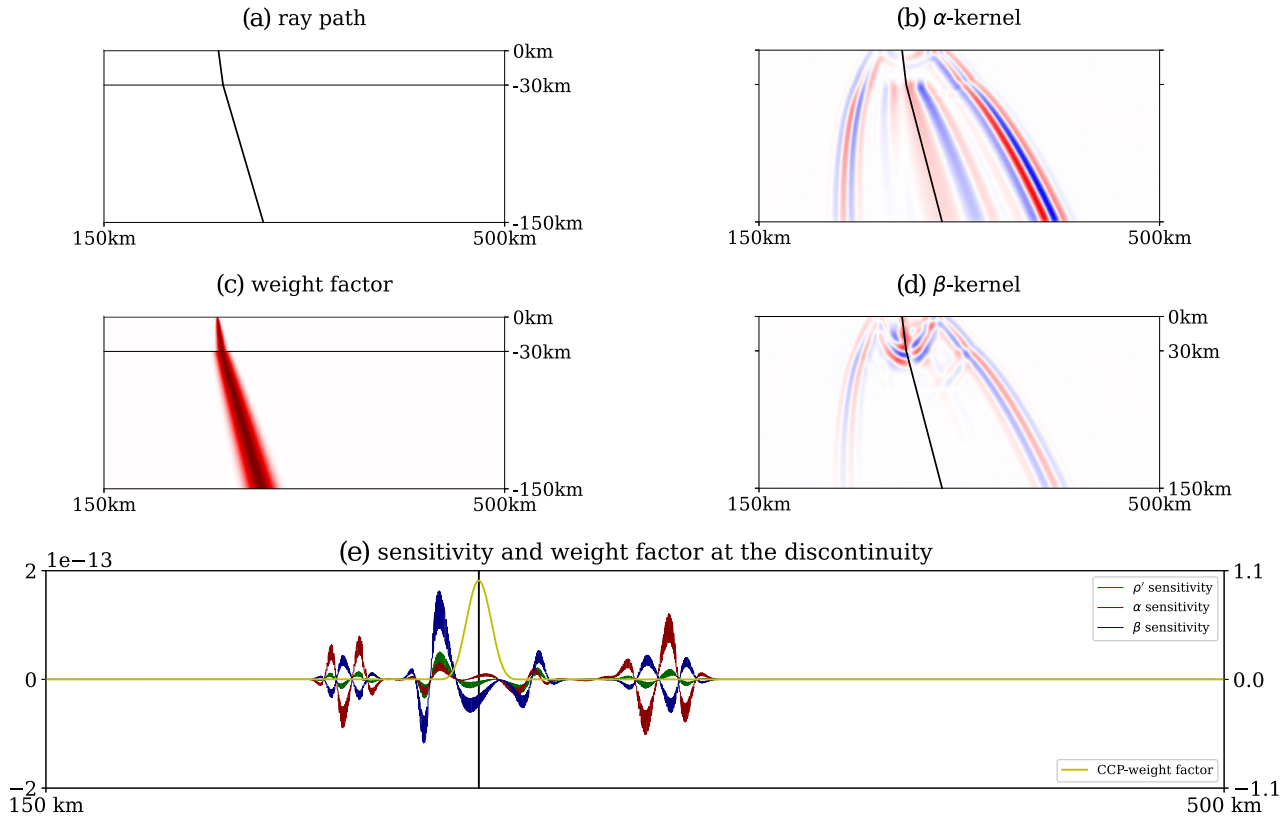

Figure 9. Comparison of (a) ray path with the (c) Fresnel-zone smoothed weight factor from Lekic *et al.* (2011), (b) α- and (d) β-sensitivity kernels. (e) The sensitivity of ρ', α and β at the discontinuity depth (30 km) compared with the CCP-weight factor by Lekic *et al.* (2011) for waves with a frequency of 0.84 Hz. The scale on the left-hand axis corresponds to the sensitivity values and on the right-hand axis to the weighting factor. The black, vertical line represents the ray theoretical location of the conversion point.

still very similar at first sight, the discontinuity topography is clearly retrieved in the β-kernel of Fig. 6(b), whereas there is no significant topography in the β-kernel of Fig. 6(c). Likewise, in Fig. 6(a), we retrieve the dip of the discontinuity. Therefore, we speculate that it might be possible to distinguish between 2-D perturbations in topography and velocity with an iterative FWI using multiple events and a wide array of receivers.

Because of the remaining similarity between the aforementioned kernels of Figs 6(b) and (c), we further investigated whether a longer receiver function time window, including multiple reflected and converted phases from the upper layer, would result in significant differentiation between the kernels related to depth perturbations and those related to perturbed velocities. Such multiples have already been incorporated in receiver functions studies, both in single station receiver function applications (Zhu & Kanamori 2000) as well as in migration schemes (Wilson & Aster 2005). The 1-D perturbed kernels (Fig. 8) show some increase in depth-velocity differentiation in the multiple source–receiver kernels. However, the differences between the kernels remain rather small. Additionally, we obtain relatively strong β-sensitivity at depths of around 100 km caused by the interaction of a more complex adjoint field with the forward P wave leading to spurious 'echos' of the discontinuity. The kernels of the 2-D models (Figs 8c and d), on the other hand, are significantly more differentiated. The kernels of the 'Complex velocity model' show an overall increased sensitivity to the velocity parameters, particularly in the upper layer compared to the kernels of the 'Topography model'. Despite the increased complexity of the kernels, they demonstrate that applying a larger receiver function time window better differentiates between discontinuity depth and velocity variations. Although the parameters and geometry of our 2-D models were chosen with the purpose to illustrate full-waveform receiver function sensitivity kernels and these choices directly affect the specific details of our kernels, we do not expect our main observations to substantially change in more realistic settings, which we will pursue in 3-D in future work.

In the last two decades migration has proven to be an important method to map discontinuities from receiver function observations. The adjoint method that we use is closely related to the various migration techniques described in the Introduction. In particular, there are similarities with the method of elastic wavefield scattering by Bostock *et al.* (2001). Both employ back projection of scattered phases to image features within the reference model. In fact, migration has been described as the first step of FWI (Tarantola 1984b). An important advantage of migration with respect to earlier methods is that, similar to FWI, no assumptions are made on the topography and orientation of the discontinuity. However, all migration techniques rely heavily on the accuracy of the velocity model. Without a reasonable velocity model, the high-amplitude arrivals will not be back projected to their correct locations and thus will fail to focus.

Compared to migration, the strength of FWI and the adjoint method is that it is able to iteratively update the velocity model. It has therefore a smaller dependence on the reference velocity model, which in the case of FWI is the starting model (Tarantola 1984a). Through the full-waveform sensitivity kernels, such as we have calculated and shown in this study, the user has insight into what part of the velocity model might contribute to the observed misfit between modelled and observed data, and how the model might be updated. Contrary to sensitivity kernels calculated with ray theory (Hansen & Schmandt 2017), our full waveform sensitivity kernels include all finite-frequency effects. However, full waveform calculations come at a high computational cost and may not be feasible for every type of study. That cost might be reduced by applying the aforementioned hybrid modelling (Monteiller *et al.* 2013; Tong *et al.* 2014). However, to incorporate all the possible contributions of the full global wavefield, some of which we have demonstrated to have a significant effect on a receiver function's sensitivity, one has to account for all possible secondary phases (Wang *et al.* 2021a).

5 CONCLUDING REMARKS

We conclude that some of the main assumptions made in receiver functions studies may be insufficient or may result in inaccurate conclusions, because they ignore important contributions from other parts of the wavefield. We have shown that the assumption of ray theory is inadequate when it comes to mapping receiver function arrivals. Our study shows that using single locations at discontinuities or even an estimate of the first Fresnel zone to account for finite-frequency effects to map receiver function arrivals severely underestimates the contribution of the broader *Ps* wavefield at the discontinuity.

Additionally, we have demonstrated that it is incorrect to assume that the far-field propagation term can be ignored in receiver function studies. Our kernels illustrate that receiver functions have a significant sensitivity to the far-field *P*-wave velocity structure. They also show the importance of the scattered *P* wavefield which interferes with that of the converted arrivals. This contribution is reduced when a dense array of receivers is used with multiple events, because it leads to the destructive interference of the *P*-wave scattering sensitivity. We further demonstrate that the well-known trade-off between velocity and discontinuity depth in receiver function studies is reduced when a receiver function time window is used that includes multiple free-surface reverberations. In none of the kernels we observe significant sensitivity to impedance, and we conclude that impedance has no significant effect on the receiver function's waveform.

FWI can be seen as the next step after migration. Although it comes at a relatively high computational cost, FWI is able to properly assess what information is contained in a receiver function waveform, and how it might map into the subsurface.

ACKNOWLEDGMENTS

This work has been financed by the research programme DeepNL of the Dutch Research Council (NWO) under project number DeepNL.2018.033. We used the open source software SPECFEM2D (available at https://geodynamics.org/cig/software/specfem2d/), for which we would like to thank its developers. Additionally, we would like to thank Ping Tong and an anonymous reviewer for their constructive comments and suggestions.

DATA AVAILABILITY

The synthetic data generated for this study are available upon request.

REFERENCES

Ammon, C.J., Randall, G.E. & Zandt, G., 1990. On the nonuniqueness of receiver functions, *J. geophys. Res. Solid Earth,* **95,** 15303–15318.

Båth, M. & Stefánsson, R., 1966. S-P conversions at the base of the crust, *Ann. Geophys.,* **19,** 119–130.

Beller, S., Monteiller, V., Operto, S., Nolet, G., Paul, A. & Zhao, L., 2018. Lithospheric architecture of the South-Western Alps revealed by multiparameter teleseismic full-waveform inversion, *Geophys. J. Int.,* **212,** 1369–1388.

Bostock, M.G., 1998. Mantle stratigraphy and evolution of the Slave province, *J. geophys. Res. Solid Earth,* **103,** 21183–21200.

Bostock, M.G., 2002. Kirchhoff-approximate inversion of teleseismic wavefield, *Geophys. J. Int.,* **149,** 787–795.

Bostock, M.G., Rondenay, S. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 1. Theory for oblique incidence, *J. geophys. Res.,* **106**(12), 30771–30782.

Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185,** 845–870.

Chen, Y. & Niu, F., 2013. Ray-parameter based stacking and enhanced preconditioning for stable inversion of receiver function data, *Geophys. J. Int.,* **194,** 1682–1700.

Cheng, C., Bodin, T. & Allen, R.M., 2016. Three-dimensional pre-stack depth migration of receiver functions with the fast marching method: a Kirchhoff approach, *Geophys. J. Int.,* **205,** 819–829.

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

Dueker, K.G. & Sheehan, A.F., 1997. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hotspot track, *J. geophys. Res. Solid Earth,* **102,** 8313–8327.

Farra, V. & Vinnik, L., 2000. Upper mantle stratification by P and S receiver functions, *Geophys. J. Int.,* **141,** 699–712.

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

Gurrola, H., Baker, G.E. & Minster, J.B., 1995. Simultaneous time-domain deconvolution with application to the computation of receiver functions, *Geophys. J. Int.,* **120,** 537–543.

Hansen, S.M. & Schmandt, B., 2017. P and S wave receiver function imaging of subduction with scattering kernels, *Geochem. Geophys. Geosys.,* **18**(12), 4487–45027.

Kind, R., Kosarev, G.L. & Petersen, N.V., 1995. Receiver functions at the stations of the German regional seismic network, *Geophys. J. Int.,* **121,** 191–202.

Kind, R. *et al.*, 2002. Seismic images of crust and upper mantle beneath Tibet: evidence for Eurasian plate subduction, *Science,* **298**(5596), 1219–1221.

Kosarev, G.L., Petersen, N.V. & Vinnik, L.P., 1993. Receiver functions for the Tien Shan Analog Broadband Network: contrast in the evolution of structures across the Talasso-Fergana Fault, *J. geophys. Res. Solid Earth,* **98,** 4437–4448.

Langston, C.A., 1979. Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. geophys. Res. Solid Earth,* **120,** 537–543.

Lekic, V., French, S.W. & Fischer, K.M., 2011. Lithospheric thinning beneath rifting regions of Southern California, *Science,* **334,** 783–787.

Levin, V. & Park, J., 1997. P-SH conversions in a flat-layered medium with anisotropy of arbitrary orientation, *Geophys. J. Int.,* **131,** 253–266.

Liu, K. & Levander, A., 2013. Three-dimensional Kirchhoff-approximate generalized Radon transform imaging using teleseismic P-to-S scattered waves, *Geophys. J. Int.,* **192,** 1196–1216.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.,* **96**(6), 2383–2397.

Marquering, H., Dahlen, F.A. & Nolet, G., 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.

Monteiller, V., Chevrot, S., Komatitsch, D. & Fuiji, N., 2013. A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.,* **192,** 230–247.

Monteiller, V., Chevrot, S., Komatitsch, D. & Wang, Y., 2015. Three-dimensional full waveform inversion of short-period teleseismic wavefield based upon the SEM-DSM hybrid method, *Geophys. J. Int.,* **202,** 811–827.

Oldenburg, D.W., 1981. A comprehensive solution to the linear deconvolution problem, *Geophys. J. Int.,* **65**(2), 331–357.

Owens, T.J. & Zandt, G., 1985. The response of the continental crust-Mantle boundary observed on broadband teleseismic receiver functions, *Geophys. Res. Lett.,* **12,** 705–708.

Owens, T.J., Zandt, G. & Taylor, S.R., 1984. Seismic evidence for an ancient rift beneath the Cumberland Plateau, Tennessee: A detailed analysis of broadband teleseismic P waveforms, *J. geophys. Res. Solid Earth,* **89,** 7783–7795.

Park, J. & Levin, V., 2000. Receiver functions from multiple-tape spectral correlation estimates, *Bull. seism. Soc. Am.,* **90**(6), 1507–1520.

Pavlis, G.L., 2005. Direct imaging of the coda of teleseismic P waves, *Geophys. Monogr. Ser.,* **157,** 171–185.

Phinney, R.A., 1964. Structure of the Earth's crust from from spectral behavior of long-period body waves, *J. geophys. Res.,* **69,** 2997–3017.

Piana Agostinetti, N. & Malinverno, A., 2010. Receiver functions inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181,** 858–872.

Pienkowska, M., Monteiller, V. & Nissen-Meyer, T., 2020. High-frequency global wavefield for local 3-D structures by wavefield injection and extrapolation, *Geophys. J. Int.,* **225**(3), 1782–1798.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Poppeliers, C. & Pavlis, G.L., 2003. Three-dimensional, pre-stack plane wave migration of teleseismic P-to-S converted phases: 1. Theory, *J. geophys. Res. Solid Earth,* **108**(B2). doi:10.1029/2001JB000216.

Regnier, M., 1988. Lateral variation of upper mantle structure beneath New Caledonia determined from P-wave receiver function: evidence for a fossil subduction zone, *Geophys. J. Int.,* **95,** 561–577.

Rondenay, S., 2009. Upper mantle imaging with array recordings of converted and scattered teleseismic waves, *Surv. Geophys.,* **30,** 377–405.

Rondenay, S., Bostock, M.G. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves III. Application to the Cascadia 1993 data set, *J. geophys. Res. Solid Earth,* **106**(B12), 30795–30807.

Ryberg, T. & Weber, M., 2000. Receiver functions arrays: a reflection seismic approach, *Geophys. J. Int.,* **141,** 1–11.

Rychert, C.A., Fisher, K.M. & Rondenay, S., 2005. A sharp lithosphere-astenosphere boundary imaged beneath eastern North America, *Nature,* **436,** 542–545.

Schmandt, B., Dueker, K., Humphreys, E. & Hansen, S., 2012. Hot mantle up-welling across the 660 beneath Yellowstone, *Earth planet. Sci. Lett.,* **331–332,** 224–236.

Shang, X., de Hoop, M. & van der Hilst, R., 2012. Beyond receiver functions: passive source reverse time migration and inverse scattering of converted waves, *Geophys. Res. Lett.,* **39**(L15308). doi: 10.1029/2012GL052289.

Sipkin, S.A. & Lerner-Lam, A.L., 1992. Pulse-shape distortion introduced by broadband deconvolution, *Bull. seism. Soc. Am.,* **81**(1), 238–258.

Tarantola, A., 1984a. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tarantola, A., 1984b. Linearized inversion of seismic reflection data, *Geophys. Prospect.,* **32,** 998–1015.

Tong, P., Komatitsch, D., Tseng, T.L., Hunh, S.H., Chen, C.W., Basini, P. & Liu, Q., 2014. A 3-D spectral-element and frequency-wave number hybrid method for high resolution seismic array imaging, *Geophys. Res. Lett.,* **41**(20), 7025–7034.

Tromp, J., Komatitsch, D. & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Commun. Comput. Phys.,* **2,** 1–32.

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

Van Stiphout, A.M., Cottaar, S. & Deuss, A.F., 2019. Receiver functions mapping of mantle transition zone discontinuities beneath Alaska using scaled 3-D velocity corrections, *Geophys. J. Int.,* **219,** 1432–1446.

Vinnik, L., 1977. Detection of waves converted from P to SV in the mantle, *Phys. Earth planet. Inter.,* **15,** 39–45.

Wang, K., Wang, Y., Song, X., Tong, P., Liu, Q. & Yang, Y., 2021. Full-waveform inversion of high-frequency teleseismic body waves based on multiple plane-wave incidence: Methods and practical applications, *Bull. seism. Soc. Am.,* **XX,** 1–15.

Wang, K., Yang, Y., Jiang, C., Tong, P., Liu, T. & Liu, Q., 2021. Adjoint tomography of ambient noise data and teleseismic P waves: methodology and applications to central California, *J. geophys. Res.,* **126,** e2021JB02164. doi:10.1029/2021JB021648.

Wang, Y. *et al.*, 2016. The deep roots of the western Pyrenees revealed by full waveform inversion of teleseismic P waves, *Geology,* **44**(6), 475–478.

Wider, D.V., 1961. *Advanced Calculus,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey , second edn.

Wilson, D. & Aster, D., 2005. Seismic imaging of the crust and upper mantle using regularized joint receiver functions, frequency-wave number filtering, and multi-mode Kirchhoff migration, *J. geophys. Res. Solid Earth,* **110**(B05305). doi: 10.1029/2004JB003430.

Woodward, M.J., 1992. Wave equation tomography, *Geophysics,* **57,** 15–26.

Zhang, J. & Langston, C.A., 1995. Constraints on oceanic lithosphere structure from deep-focus regional receiver functions inversions, *J. geophys. Res. Solid Earth,* **100**(B11), 22187–22196.

Zhang, Z., Yuan, X., Chen, Y., Tian, X., Kind, R., Li, X. & Teng, J., 2010. Seismic signature of the collision between the east Tibetan escape flow and the Sichuan Basin, *Earth planet. Sci. Lett.,* **292,** 254–264.

Zhu, H., Luo, Y., Nissen-Meyer, T., Morency, C. & Tromp, J., 2009. Elastic imaging and time-lapse migration based on adjoint methods, *Geophysics,* **74**(6), WCA167–WCA177.

Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res. Solid Earth,* **105,** 2969–2980.