Global S-wave tomography using receiver pairs: an alternative to get rid of earthquake mislocation

Christophe Zaroli, J.J. Leveque, Bernhard S. A. Schuberth, Z. Duputel, G. Nolet

Global *S*-wave tomography using receiver pairs: An alternative to get rid of earthquake mislocation

C. Zaroli[1], J.-J. Lévêque[1], B. S. A. Schuberth[2], Z. Duputel[1], and G. Nolet[3]

[1] *Institut de Physique du Globe de Strasbourg, UMR7516, Université de Strasbourg, EOST/CNRS, France*

[2] *Ludwig-Maximilians-Universität München, Department of Earth and Environmental Sciences,*

  Theresienstrasse 41, 80333 Munich, Germany

[3] *Géoazur, Université de Nice Sophia-Antipolis, UMR6526, CNRS, France*

SUMMARY

Global seismic tomography suffers from uncertainties in earthquake parameters routinely published in seismic catalogues. In particular, errors in earthquake location and origin-time may lead to strong biases in measured body-wave delay-times and significantly pollute tomographic models. Common ways of dealing with this issue are to incorporate source parameters as additional unknowns into the linear tomographic equations, or to seek combinations of data to minimise the influence of source mislocations. We propose an alternative, physically-based method to desensitise direct *S*-wave delay-times to errors in earthquake location and origin-time. Our approach takes advantage of the fact that mislocation delay-time biases depend to first order on the earthquake-receiver azimuth, and to second order on the epicentral distance. Therefore, for every earthquake, we compute *S*-wave differential delay-times between optimised receiver pairs, such that a large part of their mislocation delay-time biases cancels out (for example origin-time fully subtracts out), while the difference of their sensitivity kernels remains sensitive to the model parameters of interest. Considering realistic, randomly distributed source mis-

location vectors, as well as various levels of data noise and different synthetic Earths, we demonstrate that mislocation-related model errors are highly reduced when inverting for such differential delay-times, compared to absolute ones. The reduction is particularly rewarding for imaging the upper-mantle and transition-zone. We conclude that using optimised receiver pairs is a suitable, low cost alternative to get rid of errors on earthquake location and origin-time for teleseismic direct *S*-wave traveltimes. Moreover, it can partly remove unilateral rupture propagation effects in cross-correlation delay-times, since they are similar to mislocation effects.

Key words: Seismic tomography – Body-waves – Earthquake source observations.

1   INTRODUCTION

Teleseismic body-waves are sensitive to the physical properties of the media through which they propagate, so that they contain a lot of information on the 3–D structure of the Earth's interior. How to extract the most relevant structural information from seismograms is still an open question. Global seismic tomography encompasses numerous ways to turn seismic wave traveltime anomalies into velocity anomalies in the mantle. Though global tomographic models could further be refined using recent theoretical developments in seismic wave propagation (e.g., Dahlen et al. 2000; Tromp et al. 2005; Fichtner et al. 2009), we believe that significative improvements could also come from an improved exploitation of the continuously expanding worldwide network of digital seismometers. Indeed, the number of permanent and temporary stations on land (or islands) has grown during the last decade, as has the number of Ocean Bottom Seismometers (OBS). The recent development of submarines MERMAIDS (Mobile Earthquake Recording in Marine Areas by Independent Divers), designed to record seismic waves under water will also further increase the amount of available seismic data (e.g., Simons et al. 2006).

Since its premises, global seismic tomography has suffered from errors in the earthquake parameters routinely published in seismic catalogues, including clock drifts, earthquake mislocations, focal mechanism errors, and, for cross-correlation delay-times, the effects of rupture propagation. In this study, we focus on errors in earthquake location and origin-time which can lead to strong biases in measurements of body-wave delay-times, and significantly pollute tomographic models; we also point out the similarity of mislocation effects with unilateral rupture directivity effects. Source locations often have errors of the order of 10 km in each direction, though larger errors or biases may exist in some

oceanic regions away from seismic networks (e.g., Kennett & Engdahl 1991; Shearer 2001). When no station is available near the earthquake, it is well known that depth accuracy of shallow events can be affected by the strong trade-off between depth and origin-time (e.g., Nolet 2008). Bolton & Masters (2001) estimate delay-time errors for *P*- and *S*-waves caused by earthquake mislocation errors to be 0.6–1.2 s and 1.6–2.5 s, respectively, assuming a typical depth uncertainty of 10 km, at an epicentral distance of 70°, and for an epicentral mislocation vector of length 10–20 km.

Let us recap the most widely used strategies to deal with uncertainties in earthquake location and origin-time in global body-wave tomography. A first approach consists in seeking combinations of data such that the influence of source errors subtracts out (e.g., Kuo et al. 1987; Woodward & Masters 1991; Woodward et al. 1993; Paulssen & Stutzmann 1996; Houser et al. 2008). One can invert for the delay-time difference of two seismic phases, e.g. *S* and *ScS*, recorded at same receiver i such that: $\delta t_i^{\mathrm{S}} - \delta t_i^{\mathrm{ScS}} = \int_{\oplus} (\mathcal{K}_i^{\mathrm{S}}(\mathbf{r}) - \mathcal{K}_i^{\mathrm{ScS}}(\mathbf{r})) \delta \ln V_S(\mathbf{r}) d^3\mathbf{r}$, where $\mathcal{K}$ denotes some traveltime sensitivity kernel. Though these observables become insensitive to errors in origin-time, there may remain some residual mislocation biases for short epicentral distances Δ_i (e.g., up to ± 1.8 s for an horizontal mislocation of 18 km), or there may be some lack of sensitivity to model parameters of interest for large Δ_i (depending on the kernel difference, $\mathcal{K}_i^{\mathrm{S}} - \mathcal{K}_i^{\mathrm{ScS}}$). Another limitation is that one cannot combine all seismic phases in this way, for a typical global shear-wave dataset. For instance, many direct *S* phases cannot be combined with other phases, such as *ScS* or *SS*, because unavailable or not measured for the same source–receiver geometry (e.g., Montelli et al. 2006; Houser et al. 2008). A second, popular approach is to incorporate corrections to the published source parameters (location and origin-time) as additional unknowns into the linear system of tomographic equations (e.g., Nolet 2008). This approach has the disadvantage that data cannot always discriminate between modifying the source parameters and changing some of the velocity anomalies (in particular in the source region). Another, more formal approach consists in mathematically desensitising the linear tomographic equations to errors in source parameters (e.g., Spencer & Gubbins 1980; Pavlis & Booker 1980; Masters et al. 1996). Basically, for each event, one considers that the travel-time residuals $\boldsymbol{d}$ can be expressed as: $\boldsymbol{d} = \mathbf{G}\boldsymbol{m} + \mathbf{A}\boldsymbol{h}$, where $\mathbf{G}$ is the matrix containing the projection of the corresponding sensitivity kernels on the model grid, $\boldsymbol{m}$ represents the unknown 3–D velocity anomalies, $\mathbf{A}$ is a matrix containing the earthquake–receiver geometry, and $\boldsymbol{h}$ is a four-components vector that describes a perturbation in location and origin-time of the earthquake. Masters et al. (1996) show that one may seek linear combinations of the data to render them insensitive to the event location, by computing a projector matrix $\mathbf{P}$ such that $\mathbf{PA}$ is zero. The major issue with this projection method is that each new time-residual ($\mathbf{P}\boldsymbol{d}$) is a linear combination of *all* the original time-residuals, so that the new sensitivity matrix ($\mathbf{PG}$) is no longer sparse, and Masters et al. (1996) report that it causes computational difficulties for massive

inverse problems. It may also be difficult to intuitively apprehend the sensitivity of such algebraically combined data to some particular model parameters.

In this study, we present a physically-based, low cost alternative to desensitise teleseismic long-period direct *S*-wave delay-times to errors in earthquake location and origin-time. Our approach consists in computing, for each event, differential *S* delay-times between receiver pairs (i, j), and solving for: $\delta t_i^{\mathrm{S}} - \delta t_j^{\mathrm{S}} = \int_{\oplus} (\mathcal{K}_i^{\mathrm{S}}(\mathbf{r}) - \mathcal{K}_j^{\mathrm{S}}(\mathbf{r})) \delta \ln V_S(\mathbf{r}) d^3\mathbf{r}$. We will discuss how to efficiently select optimised receiver pairs (i, j), so that the differential delay-times $\delta t_i^{\mathrm{S}} - \delta t_j^{\mathrm{S}}$ become nearly insensitive to source mislocations, while the differences of sensitivity kernels $\mathcal{K}_i^{\mathrm{S}} - \mathcal{K}_j^{\mathrm{S}}$ remain sensitive to model parameters of interest. Our motivations are to take advantage of: 1) Our knowledge that mislocation delay-time biases depend to first order on the earthquake-receiver azimuth, and to second order on the epicentral distance; 2) The continuously expanding worldwide broadband seismic networks, which we believe can make it possible to routinely use such optimised receiver pairs in global body-wave tomography. The benefits of our approach will be illustrated with several tomographic tests using realistic synthetic data sets, biased by randomly distributed source mislocation vectors. We will use two different synthetic Earth models (Gaussian Random Field and Geodynamic models) of shear-velocity anomalies distributed in the whole-mantle to produce statistically relevant structural time-residuals. Our synthetic data will also include various levels of noise. Our goal will therefore be to show that mislocation-related model errors are highly reduced when inverting for such differential delay-times, $\delta t_i^{\mathrm{S}} - \delta t_j^{\mathrm{S}}$, compared to absolute ones, δt_i^{S}.

2   SYNTHETIC DATA WITH REALISTIC STATISTICS

2.1   Receivers and earthquakes

Our approach aims at taking advantage of the rapidly expanding worldwide seismic networks. Houser et al. (2008) report that, in the last decade, it has become common to record an earthquake on more than 200 worldwide broadband seismometers. In order to set up our synthetic experiment, therefore, we consider a dense, realistic spatial distribution of 738 receivers and 144 earthquakes, as shown in Fig. 1. Earthquake locations corresponding to events with a body-wave magnitude $5.5 \leq m_{\mathrm{b}} \leq 6.5$ are extracted from the Harvard GCMT catalogue. Receiver locations correspond to real locations of permanent and temporary broadband seismometers, including OBS.

2.2   Teleseismic *S* traveltimes

Unlike e.g. Lou et al. (2013), we aim at working with the framework of finite-frequency tomography (e.g., Dahlen et al. 2000; Zaroli et al. 2010; Mercerat & Nolet 2013; Zaroli et al. 2013; Mercerat

et al. 2014). Thus, the ith datum δt_i corresponds to the time-lag maximising the cross-correlation of an observed *S*-waveform, $u_i^{\text{obs}}(t)$, with its corresponding ray-theoretical synthetic waveform, $u_i^{\text{syn}}(t)$, and over the time-window $[t_1,\ t_2]$:

$$\delta t_i(T) = \{\tau \in \mathbb{R},\ \int_{t_1}^{t_2} u_i^{\text{syn}}(t) u_i^{\text{obs}}(t - \tau) dt = \max\}, \tag{1}$$

The observed and synthetic waveforms are filtered around a central period T prior to cross-correlation measurements (e.g., Zaroli et al. 2010), so that the time-delay is frequency-dependent. Because teleseismic *S*-waves often have their maximum of energy around 20 s period, we shall use $T = 20\,\text{s}$ throughout this study. Following Dahlen et al. (2000), the linear problem, to be solved for $m(\mathbf{r})$, is:

$$\delta t_i(T) = \int_{\oplus} \mathcal{K}_i(\mathbf{r}; T) m(\mathbf{r}) \mathrm{d}^3\mathbf{r}\,. \tag{2}$$

The volume $\oplus = \mathcal{V}_i(T)$ is limited to the region where the amplitude of the finite-frequency sensitivity (Fréchet) kernel $\mathcal{K}_i(\mathbf{r}; T)$ is significant (e.g., Zaroli et al. 2013), and $m(\mathbf{r})$ represents unknown 3–D shear-velocity perturbations, $\delta \ln V_S(\mathbf{r})$, with respect to the 1–D reference velocity model IASP91 (Kennett & Engdahl 1991), at each point $\mathbf{r}$ in the medium. The linearity of Eq. 2 is guaranteed over a wide range of anomaly amplitudes (Mercerat & Nolet 2013), which will be central to our method for relating differential delay-times to differences of finite-frequency kernels (cf., Eq. 10). Formally, one should not write an equality sign in Eq.2, since the measured time-lags are affected by effects other than the 3–D structure (cf., Eq. 3), but it is a common way to do so (e.g., Nolet 2008). The dependence on the period T will be dropped hereafter for ease of notation.

We aim at computing realistic *S*-wave time-residuals δt_i, where i denotes the receiver index for each earthquake. Considering our source–receiver geometry, the total number of absolute delay-times δt_i we have, for direct *S* phases, is $N_{\text{abs}} = 54652$. After correction for physical dispersion due to intrinsic anelastic processes, δt_i can be expressed as:

$$\delta t_i = \delta t_i^{\text{3D}} + \delta t_i^{\text{N}} + \delta t_i^{\text{X}}, \tag{3}$$

where δt_i^{3D}, δt_i^{N} and δt_i^{X} are residual-times caused by 3–D shear-velocity anomalies, measurement noise, and earthquake mislocation (including clock drift), respectively.

To compute the structural delay-times δt_i^{3D}, one needs to design a 3–D synthetic Earth model, that we refer to as m^{true}. This true-model should contain 3–D shear-velocity anomalies throughout the entire mantle, and should allow us to yield structural data with realistic statistics. We generate a true-model using a Gaussian Random Field (GRF) with an exponential correlation function (correlation length $\sim$400 km), as shown in Fig. 2. Note that the seismic heterogeneities in our true-model are characterised by both short and long wavelengths, with some sharp discontinuities. Figure 2(b, left) shows the histogram of shear-velocity anomalies ($\delta \ln V_S$), which follows a normal distribution with

mean 0 % and standard-deviation 1 %. Structural delay-times are therefore computed as:

$$\delta t_i^{3D} = \int_{\oplus} \mathcal{K}_i(\mathbf{r}) m^{true}(\mathbf{r}) \mathrm{d}^3 \mathbf{r}. \qquad (4)$$

As shown on Fig. 2(b, right), the histogram of δt_i^{3D} also follows a normal distribution, $\mathcal{N}(\mu^{3D}, \sigma^{3D})$, with mean $\mu^{3D} = 0\,\mathrm{s}$ and standard-deviation $\sigma^{3D} = 3.1\,\mathrm{s}$. Our distribution of synthetic structural delay-times is therefore in agreement with Bolton & Masters (2001), who found a standard-deviation of 3.2 s for real data. In Sect. 4.3.1, another true-model input (more "Earth-like") will be considered to compute a second set of structural time-residuals for further testing our method.

To compute the noise-related residual-times δt_i^N, we randomly draw N_{abs} samples from a normal distribution, such that:

$$\delta t_i^N \sim \mathcal{N}(\mu^N, \sigma^N), \qquad (5)$$

with mean $\mu^N = 0\,\mathrm{s}$ and standard-deviation $\sigma^N = 0.7\,\mathrm{s}$, according to the realistic estimates derived by Zaroli et al. (2010) and Bolton & Masters (2001) for $\sim$20 s dominant period teleseismic *S* waves. In Sect. 4.3.2, a "pessimistic" case consisting in twice noisier data (i.e., $\sigma^N = 1.4\,\mathrm{s}$) will also be considered for better quantifying the expected benefits from using receiver pairs.

To compute a global set of mislocation time-residuals, δt_i^X, we first need to randomly generate mislocation vectors between true (i.e., exact) and original (i.e., false) source locations: $\delta\mathbf{x}_s = \mathbf{x}_s^{true} - \mathbf{x}_s^0$. As seen in Sect. 1, Bolton & Masters (2001) estimate that *S* delays due to event mislocation should follow a normal distribution with mean 0 s and standard deviation $\simeq$1.6–2.5 s, corresponding to epicentral mislocation vectors of length $\delta\ell \simeq$10–20 km, respectively. Therefore, we need to generate 144 triplets of mislocation parameters $(\delta x, \delta y, \delta z)$, that are compatible with those realistic statistics. One may *a priori* assume that δx and δy follow normal distributions with null means and same standard deviations, that is: $\sigma_x = \sigma_y$. For the horizontal mislocation $\delta\ell$ to be on average $\simeq$10 km, we have to set: $\sigma_x = \sigma_y \simeq 8$ km. It is well known that, for teleseismic body-wave delay-times, errors in horizontal mislocation $(\delta x, \delta y)$ tend to dominate the error budget in the mislocation time-residuals, because the errors in origin time (δt_0) and depth (δz) tend to mutually cancel out on average (e.g., Bolton & Masters 2001; Nolet 2008). As detailed in Sect. 3, we aim at computing differential *S* delays for receiver pairs, which intrinsically are insensitive to the origin-time. Thus, for sake of simplicity we discard the errors in origin-time from the earthquake parameters, but compensate for this by making sure that the horizontal mislocation $(\delta x, \delta y)$ remains on average predominant. To do this, we assume that the vertical mislocation (δz) follows a normal distribution with null mean and standard deviation: $\sigma_z \simeq \sigma_x/2$. This choice, which is related to the amount of compensation between the errors in origin-time and source depth, is not crucial since using optimised receiver pairs will also largely remove the errors in source-depth (cf., Sect. 3). At this stage, we randomly draw 144 triplets $(\delta x, \delta y, \delta z)$ from the

three normal distributions $\mathcal{N}(0, \sigma_x)$, $\mathcal{N}(0, \sigma_y)$ and $\mathcal{N}(0, \sigma_z)$. Mislocation time-residuals are computed at all receivers i, for each event, as:

$$\delta t_i^{\mathrm{X}} = t_i^{\mathrm{1D}}(\mathbf{x}_s^{\mathrm{true}}) - t_i^{\mathrm{1D}}(\mathbf{x}_s^0), \tag{6}$$

where $t_i^{\mathrm{1D}}(\mathbf{x}_s^{\mathrm{true}})$ and $t_i^{\mathrm{1D}}(\mathbf{x}_s^0)$ denote the *S* phase ray-theoretical traveltimes, in the 1–D Earth model IASP91, from true and original source locations, respectively, to receiver i's location. Traveltimes t_i^{1D} are computed with the TauP Toolkit (Crotwell et al. 1999). This allows us to generate an "optimistic" set (i.e., $\delta\ell \simeq 10\,\mathrm{km}$) of 54652 mislocation time-residuals, that follows a normal distribution with null mean and standard-deviation $\sigma^{\mathrm{X}} \simeq 1.4\,\mathrm{s}$ (close to the 1.6 s estimated by Bolton & Masters (2001)). This "optimistic" set of 144 mislocation vectors $\delta\mathbf{x}_s(\delta x, \delta y, \delta z)$ is then multiplied by two in order to generate a second "pessimistic" set (i.e., $\delta\ell \simeq 20\,\mathrm{km}$) of mislocation time-residuals, that follows a normal distribution with null mean and standard-deviation $\sigma^{\mathrm{X}} \simeq 2.8\,\mathrm{s}$ (close to the 2.5 s estimated by Bolton & Masters (2001)).

2.3   Influence of source mislocation on traveltimes

In practice, mislocation residual-times, δt_i^{X}, are commonly encountered in earthquake seismology (e.g., Kikuchi & Kanamori 1982; Zhan et al. 2014). For a teleseismic *S*-wave recorded at receiver i, one can write:

$$\delta t_i^{\mathrm{X}} \approx -\frac{||\delta\mathbf{x}_s||_2}{c} \times \cos \Phi_i, \tag{7}$$

where $||\delta\mathbf{x}_s||_2 = ||\mathbf{x}_s^{\mathrm{true}} - \mathbf{x}_s^0||_2$ denotes the Euclidean norm of the mislocation vector between true and original source locations, $c = V_S(\mathbf{x}_s^0)$ denotes the shear-wave phase velocity in IASP91 within the source region, and Φ_i is the angle between $\delta\mathbf{x}_s$ and the ray path at $\mathbf{x}_s^0$. A bit of geometry leads to:

$$\cos \Phi_i = \cos \mathrm{i}_{0,i} \cos \eta + \sin \mathrm{i}_{0,i} \sin \eta \cos \widetilde{\varphi}_i, \tag{8}$$

where $\mathrm{i}_{0,i}$ is the ray take-off angle at $\mathbf{x}_s^0$, η is the angle of $\delta\mathbf{x}_s$ with respect to the vertical axis, and $\widetilde{\varphi}_i = \widetilde{\varphi} - \varphi_i$ is the azimuth difference between $\delta\mathbf{x}_s$ (azimuth $\widetilde{\varphi}$) and the departing ray at $\mathbf{x}_s^0$ (earthquake-receiver azimuth, φ_i). Thus, using Eqs. 7 and 8, one can predict the effect of source mislocation on direct *S* phase delays. Figure 3 shows an example of mislocation residual-times, δt_i^{X}, for direct *S*-waves recorded at receivers i with epicentral distance $\Delta_i \in [28°, 99°]$, for an earthquake located in the Indian Ocean with mislocation parameters $(\delta x, \delta y, \delta z) = (18.3, -8.3, 2.5)\,\mathrm{km}$. The residual δt_i^{X} varies from -3.2 s to +2.1 s, which is quite a large range of variation compared to structural residual-times (cf., Sect. 2.2). Figure 3 shows that δt_i^{X} is dominated at first order by a sinusoidal-like dependence on the earthquake-receiver azimuth φ_i, though it also depends at second order on the epicentral distance Δ_i (i.e., on the ray take-off angle $\mathrm{i}_{0,i}$). Indeed, the mislocation budget is usually

dominated by its horizontal component (cf., Sect. 2.2), and in the case of a purely horizontal mislocation ($\eta = \pi/2$), the Eq. 8 leads to $\delta t_i^{\mathrm{X}} \propto \sin \mathrm{i}_{0,i} \cos \widetilde{\varphi}_i$. The effect of an error in depth location only depends on the epicentral distance, since for a purely vertical mislocation (η equal to 0 or π) the Eq. 8 leads to $\delta t_i^{\mathrm{X}} \propto \cos \mathrm{i}_{0,i}$. Moreover, note that unilateral rupture directivity effects on cross-correlation delay-times can lead to a similar sinusoidal-like pattern (cf., Appendix A). In the following, we show that source mislocation (and unilateral rupture propagation) effects can partly be removed by using well-chosen receiver pairs.

3   A RATIONALE FOR OPTIMISED RECEIVER PAIRS

Our idea is to compute, for each event, differential delay-times of *S* phases simultaneously recorded at some pairs of receivers (i, j), that is: $\delta t_{ij} = \{\delta t_i - \delta t_j\}$. Using Eq. 3, such differential data can be formally expressed as:

$$\delta t_{ij} = \{\delta t_i^{\mathrm{3D}} - \delta t_j^{\mathrm{3D}}\} + \{\delta t_i^{\mathrm{N}} - \delta t_j^{\mathrm{N}}\} + \{\delta t_i^{\mathrm{X}} - \delta t_j^{\mathrm{X}}\}. \tag{9}$$

The linear tomographic problem (Eq. 2) becomes:

$$\delta t_{ij} = \int_{\oplus} (\mathcal{K}_i(\mathbf{r}) - \mathcal{K}_j(\mathbf{r})) m(\mathbf{r}) d^3\mathbf{r}. \tag{10}$$

We want that the new mislocation term, $\delta t_{ij}^{\mathrm{X}} = \{\delta t_i^{\mathrm{X}} - \delta t_j^{\mathrm{X}}\}$, subtracts out for well-chosen (optimised) receiver pairs (i, j). Using Eqs. 7 and 8, one can also predict the effect of source mislocation on the differential residual-time $\delta t_{ij}^{\mathrm{X}}$, for a couple of receivers (i, j):

$$\delta t_{ij}^{\mathrm{X}} \approx -\frac{||\delta \mathbf{x}_s||_2}{c} \times (\cos \Phi_i - \cos \Phi_j). \tag{11}$$

To minimise these mislocation residual-times, we propose to focus on receiver pairs with similar earthquake-receiver azimuths and epicentral distances, yet with different data sampling of the Earth. Indeed, to cancel out the mislocation term $\delta t_{ij}^{\mathrm{X}}$, one needs $\cos \Phi_i \to \cos \Phi_j$, meaning that the two receivers (i, j) should have similar earthquake-receiver azimuths and epicentral distances (i.e., ray take-off angles). However, doing so would lead to a pair of receivers so close to each other that the differential kernel would go to zero for a large number of model parameters. Since two receivers should neither be too close nor too far, we found after some trials that, for *S* wave time-residuals measured at $T = 20$ s period, a relevant compromise is to select receiver pairs (i, j) such that:

$$\{|\varphi_i - \varphi_j| \to 0° \ \& \ |\Delta_i - \Delta_j| \to 35°\}. \tag{12}$$

Let $\xi_{ij} = |\cos \Phi_i - \cos \Phi_j|$ denotes the term in $\delta t_{ij}^{\mathrm{X}}$ that should be reduced for relevant receiver pairs; note that ξ_{ij} ranges from 0 to 2 for random pairs (i, j). According to our criterion (Eq. 12),

an "ideal" receiver pair (i, j) with, for instance, $\{\varphi_i = \varphi_j, \Delta_i = 50°, \Delta_j = \Delta_i + 35°\}$, leads to: $\xi_{ij} = 0.046$ for a purely vertical mislocation, and $0 \leq \xi_{ij} \leq 0.121$ for a purely horizontal one. As expected, such a receiver pairing should allow us to strongly reduce the mislocation-related errors in differential data. Figure 4 shows an example of the difference of two finite-frequency kernels, $\mathcal{K}_i - \mathcal{K}_j$, for the same ideal receiver pair ($\varphi_{ij} = |\varphi_i - \varphi_j| = 0°$, and $\Delta_{ij} = |\Delta_i - \Delta_j| = 35°$). The differential kernel remains sensitive to velocity anomalies in a large part of the mantle, though it may sometimes be weaker at shallow depth (depending on the anomaly size and its location). For example, we compare the relative sensitivity of absolute versus differential data to some shear-velocity anomaly located in the source vicinity. The situation is sketched in Fig. 4, where a square-shaped anomaly of 200 km edge is depicted within the transition-zone. It shows that such an anomaly may become less visible when taking delay-time differences, though this should not prevent us from using receiver pairs in global tomography (cf., Sect. 4.2).

In this study, for each receiver i we select its best-partner receiver j such that:

$$j = \{k \in [1; N_{\text{sta}}^\star], \quad \mathcal{W}_{ik}(\varphi_{ik}, \Delta_{ik}) = \max\}, \tag{13}$$

where $N_{\text{sta}}^\star$ is the number of stations for which the target seismic phase has been measured, and where the weight functional $\mathcal{W}_{ij}$ should lead us to select the most adequate receiver pairs (i, j) according to our criterion (Eq. 12). We found that, for our data, the weights can be set as follows to meet our needs:

$$\begin{cases} \mathcal{W}_{ij} = \mathcal{W}_{ij}^\varphi \times \mathcal{W}_{ij}^\Delta \\ \mathcal{W}_{ij}^\varphi = e^{-(\varphi_{ij}\sqrt{\ln 2}/C_1)^2} \\ \mathcal{W}_{ij}^\Delta = \dfrac{e^{-((35-\Delta_{ij})\sqrt{\ln 2}/C_2)^2} - e^{-(35\sqrt{\ln 2}/C_2)^2}}{1 - e^{-(35\sqrt{\ln 2}/C_2)^2}}, \end{cases} \tag{14}$$

where: $C_1 = 25°$; $C_2 = 35°$, if $\Delta_{ij} \leq 35°$, and $C_2 = 35/3°$, otherwise. As illustrated in Fig. 5(a), the weights $\mathcal{W}_{ij}^\varphi$ and $\mathcal{W}_{ij}^\Delta$ are built to reach their maxima for $\varphi_{ij} = 0°$ and $\Delta_{ij} = 35°$, respectively. Note that the $\sqrt{\ln 2}$ factor causes the weight $\mathcal{W}_{ij}^\varphi$, for instance, to decrease by a factor of 2 as φ_{ij} increases by C_1.

Finding an optimum and more general criterion for pairing receivers is beyond the scope of this study, whose main goal is to quantify the possible advantages of using receiver pairs in global *S* wave tomography. The crucial point is to meet the physical basis of our criterion (Eqs. 12–14), namely the need to build pairs such that their differential kernels tend to be insensitive to the errors in origin-time and source location, while remaining sensitive to the mantle structure of interest. Thus, we only need an empirical expression for the weight functional, $\mathcal{W}_{ij}$, to automate the selection of receiver pairs based upon our *a priori* physical expectations of what relevant pairs should be.

Considering our earthquake–receiver geometry (Fig. 1), we report that 96 % (resp., 99.5%) of all our best-partner receiver pairs (i, j) are characterised with weights $\mathcal{W}_{ij}^\varphi$ and $\mathcal{W}_{ij}^\Delta$ both superior to 0.8

(resp., 0.6). We did not discard the very small number of pairs with too low weights to facilitate further comparisons of tomographic models based on either absolute or differential data. For example, Fig. 5d shows all the receiver pairs corresponding to the same earthquake as in Fig. 3. Here, we only select the best-partner pairs, for each event, to keep the number of differential data ($N_{\mathrm{diff}} = 45875$) comparable to the number of absolute ones ($N_{\mathrm{abs}} = 54652$). However, if significantly increasing the number of differential data does not represent a prohibitive computational issue, it could be interesting to consider *several* partners i for each receiver j, provided that their weights $\mathcal{W}_{ij}$ are greater than some threshold (cf., Sect. 4.3.3). Doing so would increase the number of differential data, and hopefully could lead to a better recovery of the coherent structural information through the inversion process.

4   GLOBAL TOMOGRAPHY USING RECEIVER PAIRS

4.1   The inverse problem

The linear direct problem for either the absolute or the differential delay-times can be written in the usual formulation:

$$d = \mathbf{G}m, \tag{15}$$

where d (of size N) and m (of size M) denote vectors of (absolute or differential) data and model parameters, respectively. The $\mathbf{G}$ matrix represents the projection of sensitivity kernels (or kernel differences) onto the model grid. Following Zaroli et al. (2013), we use a data-driven, irregular, model parameterisation (spherical triangular prisms and spherical layers), and analytical ray-based finite-frequency traveltime sensitivity kernels. We assume that the *prior* covariance matrices of the data, $\mathbf{C}_{\mathrm{d}}$, and of the model parameters, $\mathbf{C}_{\mathrm{m}}$, follow Gaussian probability functions, such that the optimum estimate of model parameters, m, can be obtained by minimising (e.g., Tarantola 1987):

$$f(m) = (\mathbf{G}m - d)^t \mathbf{C}_{\mathrm{d}}^{-1}(\mathbf{G}m - d) + m^t \mathbf{C}_{\mathrm{m}}^{-1}m. \tag{16}$$

Doing so leads to solving for m a system of normal equations:

$$\begin{pmatrix} \mathbf{G} \\ \mathbf{C}_{\mathrm{d}}^{\frac{1}{2}}\mathbf{C}_{\mathrm{m}}^{-\frac{1}{2}} \end{pmatrix} m = \begin{pmatrix} d \\ 0 \end{pmatrix}. \tag{17}$$

We use simple covariance matrices for the data, $\mathbf{C}_{\mathrm{d}} = \sigma_{\mathrm{d}}^2 \mathbf{I}_N$ (uniform data errors), and for the model, $\mathbf{C}_{\mathrm{m}} = \sigma_{\mathrm{m}}^2 \mathbf{I}_M$, where $\mathbf{I}_N$ and $\mathbf{I}_M$ are identity matrices of size $N \times N$ and $M \times M$, respectively. The value of $\lambda = \sigma_{\mathrm{d}}/\sigma_{\mathrm{m}}$ allows us to regularise the problem by damping the model norm. In our experience, a simple regularisation parameter (damping, λ) is sufficient to obtain smoothed model solutions since finite-frequency kernels integrate over a large volume (several Fresnel zones). For each damping value, λ, we use LSQR (e.g., Paige & Saunders 1982; Grunberg 2006) to find the

corresponding model solution, $\boldsymbol{m}(\lambda)$. The LSQR algorithm is an iterative row action method that converges to solution:

$$\boldsymbol{m}(\lambda) \;=\; \{\boldsymbol{m} \in \mathbb{R}^M, \; ||\boldsymbol{d} - \mathbf{G}\boldsymbol{m}||_2^2 + \lambda^2 ||\boldsymbol{m}||_2^2 = \min\}. \tag{18}$$

There is plenty of regularisation strategies to find an appropriate damping value for the model solution $\boldsymbol{m}(\lambda)$, though they are rarely fully objective when applied to real data whose uncertainties are often just best guesses (e.g., Hansen & O'leary 1993; Montelli et al. 2004; Boschi et al. 2006; Nolet 2008; Zaroli et al. 2013). Since in our case we know what is the true-model, m^{true}, the most natural definition for an optimal damping value, λ^{opt}, is:

$$\lambda^{\mathrm{opt}}\{\boldsymbol{m}(\lambda)\} \;=\; \{\lambda \in \mathbb{R}_+, \; ||\boldsymbol{m}(\lambda) - \boldsymbol{m}^{\mathrm{true}}||_2 = \min\}, \tag{19}$$

where $\boldsymbol{m}^{\mathrm{true}}$ denotes the vector of true-model parameters (projection of m^{true} onto the model grid). The model solution $\boldsymbol{m}(\lambda^{\mathrm{opt}})$ is the closest, from the ℓ^2-norm point of view, to the true-model. The damping derived from Eq. 19 is similar to what a subjective choice could have yield, for instance based upon an *L-curve* analysis. Using our knowledge of the true-model to infer the damping value is thus not a crucial point. However, using this truly optimal damping will allow us to objectively perform a further, fruitful error analysis in Sect. 4.2.

4.2   Results and proof-of-concept

In the context of using direct *S*-waves in global tomography, our goal is to show that it is beneficial to invert for differential rather than absolute delay-times, provided that earthquake mislocations are of the order of 10–20 km. Let us define notations for particular tomographic models and their associated optimal damping values (cf., Table 1). Subscript k will alternatively refer to three mislocation cases:

$$\begin{cases} \mathrm{k} = 0 & \Leftrightarrow \quad \text{``no misloc.''} \\ \mathrm{k} = 1 & \Leftrightarrow \quad \text{``}\sigma^{\mathrm{X}} = 1.4\,\mathrm{s}\text{''} \\ \mathrm{k} = 2 & \Leftrightarrow \quad \text{``}\sigma^{\mathrm{X}} = 2.8\,\mathrm{s}\text{''}. \end{cases} \tag{20}$$

Let us consider the *absolute-models*, $\boldsymbol{A}_{\mathrm{k}}(\lambda)$, and the *differential-models*, $\boldsymbol{D}_{\mathrm{k}}(\lambda)$, resulting from inversions (λ denotes some damping value) of absolute and differential *S* delay-times affected or not by source mislocations, respectively:

$$\boldsymbol{A}_{\mathrm{k}}(\lambda) \;\Leftarrow\; \begin{cases} \delta t_i = \delta t_i^{\mathrm{3D}} + \delta t_i^{\mathrm{N}} & \textit{if} \quad \mathrm{k} = 0 \\ \delta t_i = \delta t_i^{\mathrm{3D}} + \delta t_i^{\mathrm{N}} + \delta t_i^{\mathrm{X}} & \textit{if} \quad \mathrm{k} = \{1; 2\}, \end{cases} \tag{21}$$

and

$$\boldsymbol{D}_{\mathrm{k}}(\lambda) \;\Leftarrow\; \begin{cases} \delta t_{ij} = \delta t_{ij}^{\mathrm{3D}} + \delta t_{ij}^{\mathrm{N}} & if \quad \mathrm{k} = 0 \\ \delta t_{ij} = \delta t_{ij}^{\mathrm{3D}} + \delta t_{ij}^{\mathrm{N}} + \delta t_{ij}^{\mathrm{X}} & if \quad \mathrm{k} = \{1; 2\}. \end{cases} \tag{22}$$

One can define their associated optimal damping values, for $\mathrm{k} = \{0; 1; 2\}$, as follows:

$$\begin{cases} \lambda^{A_{\mathrm{k}}} \;=\; \lambda^{\mathrm{opt}}\{\boldsymbol{A}_{\mathrm{k}}(\lambda)\} \\ \lambda^{D_{\mathrm{k}}} \;=\; \lambda^{\mathrm{opt}}\{\boldsymbol{D}_{\mathrm{k}}(\lambda)\}. \end{cases} \tag{23}$$

Our goal therefore consists in showing that, for $\mathrm{k} = \{1; 2\}$, the differential-models $\boldsymbol{D}_{\mathrm{k}}(\lambda^{D_{\mathrm{k}}})$ are less differing from the true-model $\boldsymbol{m}^{\mathrm{true}}$, when compared to the absolute-models $\boldsymbol{A}_{\mathrm{k}}(\lambda^{A_{\mathrm{k}}})$, meaning that errors on earthquake locations are generating (significantly) less errors into the model solutions if one inverts for such differential (δt_{ij}) rather than absolute (δt_i) residual-times.

4.2.1   *Reference absolute-model*

First, note that the best tomographic solution within our reach, if S delay-times were free of mislocation biases (i.e., $\mathrm{k} = 0$), would require to use absolute rather than differential data, in order to fully exploit the kernels. The corresponding reference absolute-model, $\boldsymbol{A}_0(\lambda^{A_0})$, hence represents the best achievable tomographic model given our data geometry. The term $\{\boldsymbol{m}^{\mathrm{true}} - \boldsymbol{A}_0(\lambda^{A_0})\}$ therefore represents the *basic* errors for any absolute-model, and is related to two factors: 1) Limited data coverage; 2) Applied regularisation to deal with measurement noise. Figure 6 displays at several depths through the mantle a model comparison between $\boldsymbol{A}_0(\lambda^{A_0})$ and $\boldsymbol{m}^{\mathrm{true}}$. Since only teleseismic S phases are used, $\boldsymbol{A}_0(\lambda^{A_0})$ mostly differs with $\boldsymbol{m}^{\mathrm{true}}$ at shallow depth (upper-mantle, transition-zone), and preferentially beneath oceanic regions (lower data coverage). At greater depth (mid lower-mantle and deeper), model differences largely decrease.

4.2.2   *Extra errors for absolute- and differential-models*

We are then interested in comparing the *extra* errors for absolute- and differential-models, that is: $\{\boldsymbol{A}_{\mathrm{k}}(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_0})\}$ and $\{\boldsymbol{D}_{\mathrm{k}}(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_0})\}$, respectively, where $\mathrm{k} = \{1; 2\}$. The word *extra* refers to the supplementary model errors (in addition to the basic errors) involved by using mislocation-biased data and/or differential kernels. Figures 7(a, d) and 8(a, d) show tomographic pictures of the extra errors for absolute- and differential-models, and their histograms are shown in Fig. 9(a, c). We report that, while the true-model varies in amplitude (up to) $\uparrow \pm 3.5\,\%$, the extra errors for absolute-models can locally be very significant within the upper-mantle ($\uparrow \pm 3\,\%$, if $\mathrm{k} = 1$; $\uparrow \pm 4\,\%$, if $\mathrm{k} = 2$) and transition-zone ($\uparrow \pm 2.5\,\%$, if $\mathrm{k} = 1$; $\uparrow \pm 3\,\%$ if $\mathrm{k} = 2$), though they are weaker at the top of mid lower-mantle ($\uparrow \pm 1\,\%$, if $\mathrm{k} = 1$; $\uparrow \pm 2\,\%$ if $\mathrm{k} = 2$). Meanwhile, we report that the extra

errors for differential-models are significantly lower within the upper-mantle ($\uparrow \pm 2\,\%$, if k $= \{1; 2\}$), transition-zone ($\uparrow \pm 2\,\%$, if k $= \{1; 2\}$), and at the top of mid lower-mantle ($\uparrow \pm 1\,\%$, if k $= \{1; 2\}$). Note that the extra errors for differential-models do not vary much for both mislocation cases, which is an evidence for the model insensitivity to source mislocations when using such differential data. We conclude that it is very rewarding to invert for differential, rather than absolute, *S* delay-times, provided that receivers could appropriately be paired (according to Eq. 12) and mislocation residual-times would statistically be characterised by $\sigma^{\mathrm{X}} = 1.4$–$2.8\,\mathrm{s}$.

4.2.3 *Contributors to extra errors*

Last but not least, we are interested in (quantitatively) identifying the contributors to the extra errors for absolute- and differential-models. This can easily be achieved if we realise that the extra errors can be decomposed, for absolute- and differential-models, such that:

$$
\begin{cases}
\boldsymbol{A}_{\mathrm{k}}(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_0}) = \{\boldsymbol{A}_{\mathrm{k}}(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_{\mathrm{k}}})\} \\
\qquad\qquad\qquad + \{\boldsymbol{A}_0(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_0})\},
\end{cases}
\tag{24}
$$

and

$$
\begin{cases}
\boldsymbol{D}_{\mathrm{k}}(\lambda^{D_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_0}) = \{\boldsymbol{D}_{\mathrm{k}}(\lambda^{D_{\mathrm{k}}}) - \boldsymbol{D}_0(\lambda^{D_{\mathrm{k}}})\} \\
\qquad\qquad\qquad + \{\boldsymbol{D}_0(\lambda^{D_{\mathrm{k}}}) - \boldsymbol{D}_0(\lambda^{D_0})\} \\
\qquad\qquad\qquad + \{\boldsymbol{D}_0(\lambda^{D_0}) - \boldsymbol{A}_0(\lambda^{A_0})\},
\end{cases}
\tag{25}
$$

respectively, and where k $= \{1; 2\}$. The situation is illustrated in Figs. 7 and 8.

On the one hand, Eq. 24 tells us that the extra errors for absolute-models result from two terms: 1) The first term $\{\boldsymbol{A}_{\mathrm{k}}(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_{\mathrm{k}}})\}$ is due to adding mislocation biases in the data, while keeping the same regularisation parameter; 2) The second term $\{\boldsymbol{A}_0(\lambda^{A_{\mathrm{k}}}) - \boldsymbol{A}_0(\lambda^{A_0})\}$ is due to increasing the regularisation parameter, to deal with mislocation-biased data. The extra errors for absolute-models are dominated by the first term (mislocation) as shown in Figs. 7(b) and 8(b). A non-negligible contribution may also come from the second term (damping) in case of strong mislocation biases, in particular at the top of mid lower-mantle, cf. Fig. 8(c). Indeed, to deal with large mislocation biases (k $= 2$) in absolute data, one has to significantly increase the regularisation parameter ($\lambda^{A_2} \gg \lambda^{A_0}$).

On the other hand, Eq. 25 tells us that the extra errors for differential-models result from three terms: 1) The first term $\{\boldsymbol{D}_{\mathrm{k}}(\lambda^{D_{\mathrm{k}}}) - \boldsymbol{D}_0(\lambda^{D_{\mathrm{k}}})\}$ is due to adding mislocation biases in the data, while keeping the same regularisation parameter; 2) The second term $\{\boldsymbol{D}_0(\lambda^{D_{\mathrm{k}}}) - \boldsymbol{D}_0(\lambda^{D_0})\}$ is due to increasing the regularisation parameter, to deal with mislocation-biased data; 3) The third term $\{\boldsymbol{D}_0(\lambda^{D_0}) - \boldsymbol{A}_0(\lambda^{A_0})\}$ is due to the use of kernel differences, which can itself be decomposed into the sum of a (predominant) first part $\{\boldsymbol{D}_0(\lambda^{D_0}) - \boldsymbol{A}_0(\lambda^{D_0})\}$ involving a lack of sensitivity to some

model parameters and a second part $\{\boldsymbol{A}_0(\lambda^{D_0}) - \boldsymbol{A}_0(\lambda^{A_0})\}$ involving the need for more stringent damping to deal with enlarged noise-related errors in differential data (cf., Sect. 4.3.2). The extra errors for differential-models are essentially not influenced by the first term (mislocation) as shown in Figs. 7(e) and 8(e). This result confirms our appropriate receiver pairing for getting rid of a large part of mislocation effects in differential delay-times. There is therefore no need to impose some extra damping for differential-models to deal with mislocation-related errors ($\lambda^{D_k} \approx \lambda^{D_0}$), and the second term (damping) of the extra errors for differential-models is very weak, cf. Figs. 7(f) and 8(f). Finally, the dominant term of the extra errors for differential-models is the third one (kernel difference), as shown in Figs. 7(g) and 8(g). Even though using kernel differences necessarily involves a lack of sensitivity to some model parameters, our results (Sect. 4.2.2) clearly show that the overall balance is largely in favour of differential data via the use of optimised receiver pairs.

4.3   Discussion

In the following, we aim at discussing several points that the reader could wonder about before applying our approach to some *real* global body-wave dataset. We also mention some perspectives towards fully getting rid of earthquake mislocation using multiple receiver pairs.

4.3.1   Earth-like true-model

First, one may wonder whether using a more Earth-like true-model, instead of a Gaussian Random Field (GRF) model, could modify our conclusions? For instance, it is well known that 3–D velocity anomalies inside the (real) mantle are not uniformly distributed (e.g., higher amplitudes in the upper-mantle). Thus, we have done supplementary calculations with another, more Earth-like, true-model denoted $m_{\text{Geody}}^{\text{true}}$. We used the shear-wave velocity structure corresponding to the high-resolution geo-dynamic model S09-M2-Q by Schuberth et al. (2009, 2012), here referred to as *Geody*. This geody-namic model relies on three assumptions: 1) Large-scale flow structure related to plate motion his-tory; 2) Radial three-layer (lithosphere, upper and lower mantle) viscosity profile in agreement with postglacial rebound and geoid observations; 3) Isochemical whole mantle flow with pyrolite com-position. Shear wave velocities are computed by converting the temperature field of the mantle to elastic parameters and density using thermodynamic models of mantle mineralogy. Figure 10 shows this second true-model, with respect to IASP91, after projection onto our own model parameterisation. It compares well to global tomographic models in terms of spectral characteristics and magnitude of velocity anomalies, though it cannot correctly predict the exact location and pattern of structure in the deep mantle. In addition, it shows slightly different, yet realistic, *S*-wave delay-time statistics (cf., Fig. 10(b)). Note that $m_{\text{Geody}}^{\text{true}}$ is characterised by large provinces in the mantle with very weak amplitudes

($\delta \ln V_S(\mathbf{r}) \approx 0$), as shown on the tomographic cross-sections in Fig. 10(c, d). Such a feature precludes to analyse the extra errors for differential-models related to the use of kernel differences. This is the reason why we first used the GRF true-model, with amplitudes uniformly distributed in the whole mantle. We now compute a new set of structural delay-times as:

$$\delta t_{i,\,\text{Geody}}^{\text{3D}} = \int_{\oplus} \mathcal{K}_i(\mathbf{r}) m_{\text{Geody}}^{\text{true}}(\mathbf{r}) \mathrm{d}^3\mathbf{r} \,, \tag{26}$$

and then a new set of absolute time-residuals:

$$\delta t_{i,\,\text{Geody}} = \delta t_{i,\,\text{Geody}}^{\text{3D}} + \delta t_i^{\text{N}} + \delta t_i^{\text{X}}. \tag{27}$$

Finally, using exactly the same receiver pairs (source–receiver geometry is identical), we compute the extra errors for the new absolute- and differential-models, as shown in Fig. 9(b, d). The extra errors are, again, significantly larger for absolute- than for differential-models, for the two considered mislocation cases. We conclude that our results do not depend much on the input true-model, provided that it (grossly) reproduces the statistics of real data.

4.3.2 Noisy data

The noise-related errors are enlarged when taking delay-time differences (σ^{N} is multiplied by a factor of $\sqrt{2}$), so that the inversion of differential data may require more stringent regularisation parameter to suppress their induced model errors (cf., Sect. 4.2.3). Therefore, one may wonder whether the total extra errors could be larger in differential- rather than -absolute models for different levels of noise in the data. We have already shown that there was no reason to be worried about this hypothesis in the (realistic) case of $\sigma^{\text{N}} = 0.7\,\text{s}$. Here, we aim at verifying that our results still hold in a more pessimistic case where $\sigma^{\text{N}} = 1.4\,\text{s}$. We have then multiplied by two the original noise-related delay-times:

$$\delta t_{i,\,\text{Pessi}}^{\text{N}} = \delta t_i^{\text{N}} \times 2 \,. \tag{28}$$

Then, using this new set of $\delta t_{i,\,\text{Pessi}}^{\text{N}}$ we have recomputed all the extra errors for the absolute- and differential-models, with the same mislocation regimes ($\sigma^{\text{X}} = 1.4$ and $2.8\,\text{s}$) and true-model inputs (GRF and Geody). The new results are shown in Fig. 9(e–h), and clearly demonstrate the interest of using receiver pairs with much noisier data. In particular, it still works fairly well in the worst-case scenario, i.e. weak mislocation-related errors ($\sigma^{\text{X}} = 1.4\,\text{s}$) and large noise-related errors ($\sigma^{\text{N}} = 1.4\,\text{s}$), as shown in Fig. 9(e,f). We conclude that the use of optimised receiver pairs in global *S*-wave tomography should be beneficial for real data applications, no matter whether the earthquake location catalogue is fairly accurate ($\sigma^{\text{X}} = 1.4\,\text{s}$) or less ($\sigma^{\text{X}} = 2.8\,\text{s}$), and the level of noise in the dataset is relatively weak ($\sigma^{\text{N}} = 0.7\,\text{s}$) or large ($\sigma^{\text{N}} = 1.4\,\text{s}$).

4.3.3   Sparse receiver coverage

We have previously shown that, with our synthetic coverage, using all the pairs of best-partner receivers (Eqs. 12–14) was highly beneficial to decrease the mislocation-related errors in the model space. When dealing with present-time data, source–receiver geometry may be sparser, mainly because not all stations are recording simultaneously. It is thus likely that some of those best-partner pairs could not be relevant enough for getting rid of source mislocation. This should not be a brick wall that prevents us from using all other appropriate pairs, and the best option would therefore be to invert for a mix of absolute (δt_i) and differential (δt_{ij}) delay-times. The related question that then arises is how to set up a criterion indicating when to switch to absolute data. This could be done by setting a threshold value on weights $\mathcal{W}_{ij}^{\varphi}$ and $\mathcal{W}_{ij}^{\Delta}$ ensuring that both the azimuth and distance criteria of Eq. 12 are "fairly" met. What could be these minimal weights? Formally answering this question is beyond the scope of this study. However, we can get a first idea from our synthetic case. In this successful case, 99.5% of the selected pair weights, $\mathcal{W}_{ij}^{\varphi}$ and $\mathcal{W}_{ij}^{\Delta}$, are greater than 0.6, and 96% of them are greater than 0.8. Note that $\mathcal{W}_{ij}^{\varphi}$ and $\mathcal{W}_{ij}^{\Delta}$ being greater than 0.8 corresponds to: $\varphi_{ij} \leq 14°$ and $22° \leq \Delta_{ij} \leq 42°$, respectively. As mentioned in Sect. 3, using a threshold criterion, instead of a best partner criterion, may significantly increase the number of selected differential data. Though it could lead to a better recovery of the coherent structural information, it would also lead to heavier computational issues. The appropriate minimal values for $\mathcal{W}_{ij}^{\varphi}$ and $\mathcal{W}_{ij}^{\Delta}$ should thus result from some compromise between the amount of differential data and the degree of compatibility with Eq. 12.

4.3.4   Other seismic phases

Figures 7(a) and 8(a) are a reminder that earthquake mislocations may represent severe limitations to the final resolution of global *S*-wave tomographic models. This is particularly true for the upper-mantle and transition zone, mainly in regions where earthquakes occur. Removing such bias may be crucial, for instance when jointly inverting for body- and surface-waves, so that body-waves do not bring biased informations in regions of common data sensitivity (upper-mantle and transition-zone). Although we chose to focus on direct *S*-waves in this study, one could also desensitise direct *P*-wave delay-times using similar receiver pairs. Bolton & Masters (2001) report that mislocation effects are even more troublesome for *P*-wave tomography, so that the benefits from using receiver pairs could even potentially be greater. Future work will also consist in tuning our criteria for efficiently combining receivers having recorded other kind of seismic phases (e.g., *ScS*, *SS*).

4.3.5   *Towards fully getting rid of mislocation using multiple receiver pairs*

We aim at showing that, for each event, it is theoretically possible to fully get rid of source mislocation effects, yet sampling the regions of interest, by combining S phases recorded at four well-chosen receivers. Consider two receiver pairs (i, j) and (k, l) such that:

$$\begin{cases} \Delta_i \approx \Delta_j \approx \Delta_k \approx \Delta_l \\ |\varphi_i - \varphi_k| \approx 180° \\ |\varphi_j - \varphi_l| \approx 180° \\ i \neq j \neq k \neq l \end{cases} \tag{29}$$

Using Eqs. 8 and 11, one can demonstrate that the data combination $\{\delta t_{ij} + \delta t_{kl}\}$ leads to fully get rid of errors in source location and origin-time, that is:

$$\delta t_{ij}^{\mathrm{X}} + \delta t_{kl}^{\mathrm{X}} \approx 0 \,. \tag{30}$$

To maximise the sensitivity of such multiple kernel differences, $\{\mathcal{K}_i - \mathcal{K}_j + \mathcal{K}_k - \mathcal{K}_l\}$, to model parameters, one should favour receiver pairs (i, j) and (k, l) such that:

$$|\varphi_i - \varphi_j| \approx |\varphi_k - \varphi_l| \to 90° \,. \tag{31}$$

A dense enough receiver coverage would be required to find four receivers (i, j, k, l) verifying Eqs. 29 and 31. In order to exploit the most relevant information in mislocation-biased data, the best approach would therefore consist in inverting for a mix of absolute-delays (δt_i), simple differential-delays (δt_{ij}), and multiple differential-delays ($\delta t_{ij} + \delta t_{kl}$), depending upon the actual source–receiver geometry to be dealt with. We postpone a more formal study on the feasibility of using such multiple receiver pairs in global body-wave tomography (e.g., weights definition, sensitivity of multiple kernel differences).

5   CONCLUSION

From its first applications, global seismic tomography has suffered from uncertainties in earthquake parameters, including clock time-drift, earthquake mislocation, and, for cross-correlation delay-times, the effects of rupture propagation. In this study, we have focussed on errors in earthquake location and origin-time. Teleseismic S-wave residual-times are commonly affected by mislocation biases with standard-deviation $\sigma^{\mathrm{X}} \sim 2\,\mathrm{s}$, while their 3–D structural part corresponds to $\sigma^{\mathrm{3D}} \sim 3\,\mathrm{s}$. Thus, earthquake uncertainties can represent severe limitations to the improvement of global tomographic models. In this study, we have presented an alternative, physically-based method to desensitise teleseismic long-period direct S-wave delay-times to errors in earthquake location and origin-time. Our approach takes advantage of the fact that mislocation delay-time biases depend to first order on the earthquake-

receiver azimuth, and to second order on the epicentral distance. For each event, differential delay-times are computed between well-chosen receiver pairs. We have shown the feasibility of selecting receiver pairs such that differential data become nearly insensitive to source mislocation parameters (e.g., origin-time fully subtracts out), while the (finite-frequency) kernel differences remain sensitive to model parameters of interest. We found that a good compromise is to favour receiver pairs with similar earthquake-receiver azimuths, and whose epicentral distances differ by $\sim 35\,°$. Considering realistic, randomly distributed source mislocation vectors, as well as various levels of data noise and different synthetic Earths, we have shown that mislocation-related model errors could highly be reduced when inverting for such differential delay-times, compared to absolute ones. In particular, we have shown how much it could be rewarding in the upper-mantle and transition-zone. We conclude that using optimised receiver pairs is a suitable, low cost alternative to get rid of errors on earthquake location and origin-time for teleseismic direct *S*-wave traveltimes. Moreover, it can partly get rid of unilateral rupture propagation effects in cross-correlation delay-times, since they are similar to mislocation effects. The same benefits should hold for teleseimic direct *P*-waves. Finally, using receiver pairs could help us to better exploit the weak finite-frequency effects (e.g., wavefront-healing) recently observed on teleseismic *S*-wave cross-correlation delay-times (e.g., Zaroli et al. 2010), and lead to improved imaging of small-scale 3–D velocity anomalies in future global tomographic models.

ACKNOWLEDGMENTS

REFERENCES

Aki, K. & Richards, P., 2002. *Quantitative Seismology, 2nd ed.*, Univ. Science Books, Sausalito, CA.

Bolton, H. & Masters, G., 2001. Traveltimes of P and S from global digital seismic networks: implication for the relative variation of P and S velocity in the mantle, *J. Geophys. Res*, **106**, 13,527–13,540.

Boschi, L., Becker, T. W., Soldati, G., & Dziewonski, A. M., 2006. On the relevance of born theory in global seismic tomography, *Geophys. Res. Lett.*, **33**(L06302, doi:10.1029/2005GL025063).

Crotwell, H., Owens, T., & Ritsema, J., 1999. The Taup toolkit: flewible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.*, **70**(2), 154–160.

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

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

Grunberg, M., 2006. *Conception d'une méthode de maillage 3D parallèle pour la construction d'un modèle de Terre réaliste par la tomographie sismique*, Ph.D. thesis, Strasbourg University, France.

Hansen, C. & O'leary, D., 1993. The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.*, **14**(6), 1487–1503.

Houser, C., Masters, G., Shearer, P. M., & Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.*, **174**, 195–212.

Kennett, B. & Engdahl, E., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429–465.

Kikuchi, M. & Kanamori, H., 1982. Inversion of complex body waves, *Bull. seism. Soc. Am.*, **72**(2), 491–506.

Kuo, B.-Y., Forsyth, D., & Wysession, M., 1987. Lateral heterogeneity and azimuthal anisotropy in the north atlantic determined from SS-S differential travel times, *J. Geophys. Res*, **92**, 6421–6436.

Lou, X., van der Lee, S., & Lloyd, S., 2013. Aimbat: A python/matplotlib tool for measuring teleseismic arrival times, *Seismol. Res. Lett.*, **84**(85-93).

Masters, G., Johnson, S., Laske, G., Bolton, H., & Davies, J. H., 1996. A shear-velocity model of the mantle, *Phil. Trans. Roy. Soc. A.*, **354**, 1385–1410.

Mercerat, E., D. & Nolet, G., 2013. Comparison of ray- and adjoint-based sensitivity kernels for body-wave seismic tomography, *Geophys. Res. Lett.*, **39**(L12301, 6PP.), doi:10.1029/2012GL052002.

Mercerat, E., D., Nolet, G., & Zaroli, C., 2014. Cross-borehole tomography with correlation delay times, *GEOPHYSICS*, **79**(1), (R1–R12, doi: 10.1190/geo2013-0059.1).

Montelli, R., Nolet, G., Masters, G., Dahlen, F. A., & Hung, S.-H., 2004. Global P and PP traveltime tomography: Rays versus waves, *Geophys. J. Int.*, **158**, 636–654.

Montelli, R., Nolet, G., Dahlen, F. A., & Masters, G., 2006. A catalogue of deep mantle plumes. new results from finite-frequency tomography, *Geochem. Geophys. Geosyst.*, **7**(11), DOI: 10.1029/2006GC001248.

Nolet, G., 2008. *A breviary of seismic tomography*, Cambridge University Press, Cambridge, UK.

Paige, C. C. & Saunders, M., 1982. Lsqr: An algorithm for sparse, linear equations and sparse least squares, *A.C.M. Trans. Math. Softw.*, **8**, 43–71.

Paulssen, H. & Stutzmann, E., 1996. On PP-P differential travel time measurements, *Geophys. Res. Lett.*, **23**, 1833–1836.

Pavlis, G. & Booker, J., 1980. The mixed discrete-continuous inverse problem: Application to the simultaneous determination of earthquake hypocenters and velocity structure, *J. Geophys. Res*, **85**, 4801–4810.

Schuberth, B. S. A., Bunge, H.-P., Steinle-Neumann, G., Moder, C., & Oeser, J., 2009. Thermal versus elastic heterogeneity in high-resolution mantle circulation models with pyrolite composition: high plume excess temperatures in the lowermost mantle, *Geochem. Geophys. Geosyst.*, **10**(1), Q01W01.

Schuberth, B. S. A., Zaroli, C., & Nolet, G., 2012. Synthetic seismograms for a synthetic earth: long-period P- and S-wave traveltime variations can be explained by temperature alone, *Geophys. J. Int.*, **188**(3), 1393–1412.

Shearer, P. M., 2001. Improving global seismic event locations using source-receiver reciprocity, *Bull. seism. Soc. Am.*, **91**, 594–603.

Simons, F., Nolet, G., Babcock, J., Davis, R., & Orcutt, J., 2006. A future for drifting seismic networks, *Eos Transactions American Geophysical Union*, **31**, 305–307.

Spencer, C. & Gubbins, D., 1980. Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media, *Geophys. J. Roy. Astr. Soc.*, **63**, 95–116.

Tarantola, A., 1987. Inversion of traveltimes and seismic waveforms, *In: Seismic Tomography, ed. Nolet G., Reidel, Dordrecht*, pp. 135–157.

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

VanDecar, J. & Crosson, R., 1990. Determination of teleseismic arrival times using multi-channel cross-correlation and least squares, *Bull. seism. Soc. Am.*, **80**, 150–159.

Woodward, R. & Masters, G., 1991. Global upper mantle structure from long-period differential travel times, *J. Geophys. Res*, **96**(6351-6377).

Woodward, R., Forte, A., Su, W.-J., & Dziewonski, A. M., 1993. Constraints on the large-scale structure of the Earth's mantle, *In: Takahashi, E. and Jeanloz, R. and Rubie, D. (Eds.), Evolution of the Earth and Planets, Geophys. Monogr. Ser., AGU*, **74**, 89–109.

Zaroli, C., Debayle, E., & Sambridge, M., 2010. Frequency-dependent effects on global S-wave traveltimes: wavefront-healing, scattering and attenuation, *Geophys. J. Int.*, **182**, 1025–1042.

Zaroli, C., Sambridge, M., Lévêque, J.-J., Debayle, E., & Nolet, G., 2013. An objective rationale for the choice of regularisation parameter with application to global multiple-frequency S-wave tomography, *Solid Earth*, **4**, 357–371.

Zhan, Z., Kanamori, H., Tsai, V., Helmberger, D., & Wei, S., 2014. Rupture complexity of the 1994 bolivia and 2013 sea of okhotsk deep earthquakes, *Earth Planet. Sci. Lett.*, **385**, 89–96.

Table 1. Notations for particular models, $\boldsymbol{m}(\lambda)$, and their associated optimal damping values, $\lambda^{\mathrm{opt}}\{\boldsymbol{m}(\lambda)\}$.

$\boldsymbol{m}(\lambda)$	data	mislocation	$\lambda^{\mathrm{opt}}\{\boldsymbol{m}(\lambda)\}$
$\boldsymbol{A}_0(\lambda)$	$\delta t_i^{\mathrm{3D+N}}$	" no misloc. "	λ^{A_0}
$\boldsymbol{D}_0(\lambda)$	$\delta t_{ij}^{\mathrm{3D+N}}$	–	λ^{D_0}
$\boldsymbol{A}_1(\lambda)$	$\delta t_i^{\mathrm{3D+N+X}}$	" $\sigma^{\mathrm{X}} = 1.4\,\mathrm{s}$ "	λ^{A_1}
$\boldsymbol{D}_1(\lambda)$	$\delta t_{ij}^{\mathrm{3D+N+X}}$	–	λ^{D_1}
$\boldsymbol{A}_2(\lambda)$	$\delta t_i^{\mathrm{3D+N+X}}$	" $\sigma^{\mathrm{X}} = 2.8\,\mathrm{s}$ "	λ^{A_2}
$\boldsymbol{D}_2(\lambda)$	$\delta t_{ij}^{\mathrm{3D+N+X}}$	–	λ^{D_2}

Figure 1. Spatial distribution of receivers (triangles) and earthquakes (stars). Black dashed line: tectonic plates.
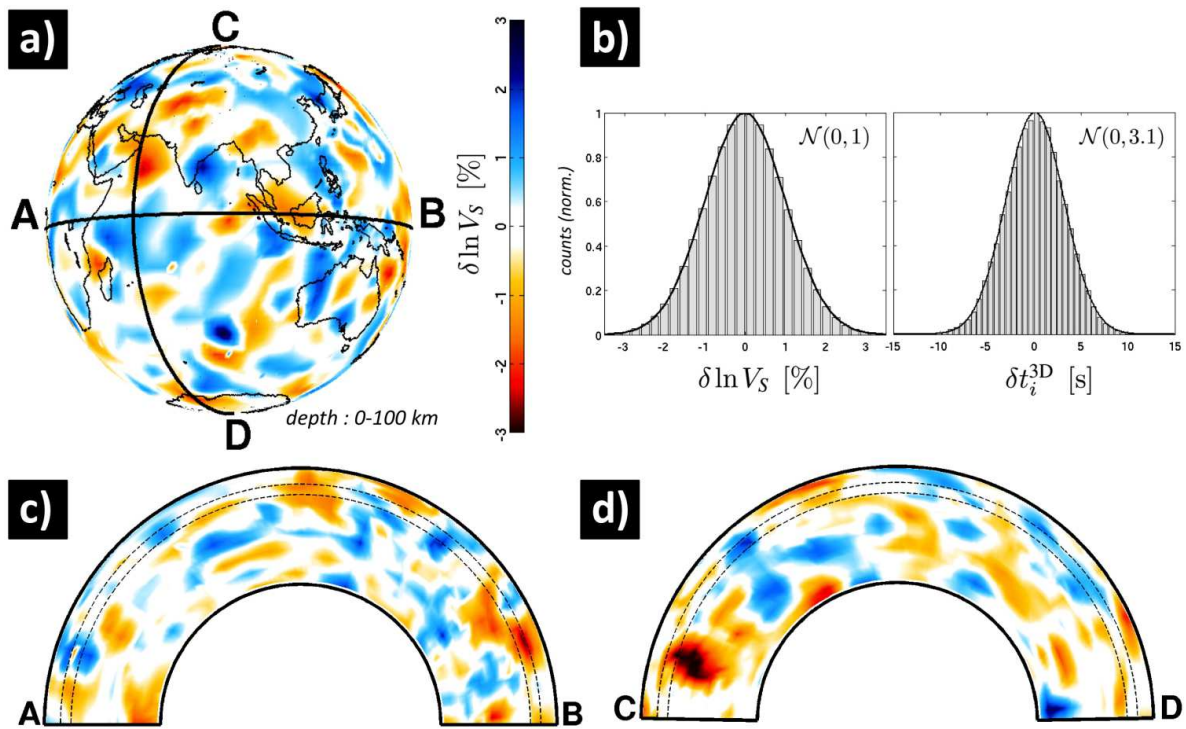
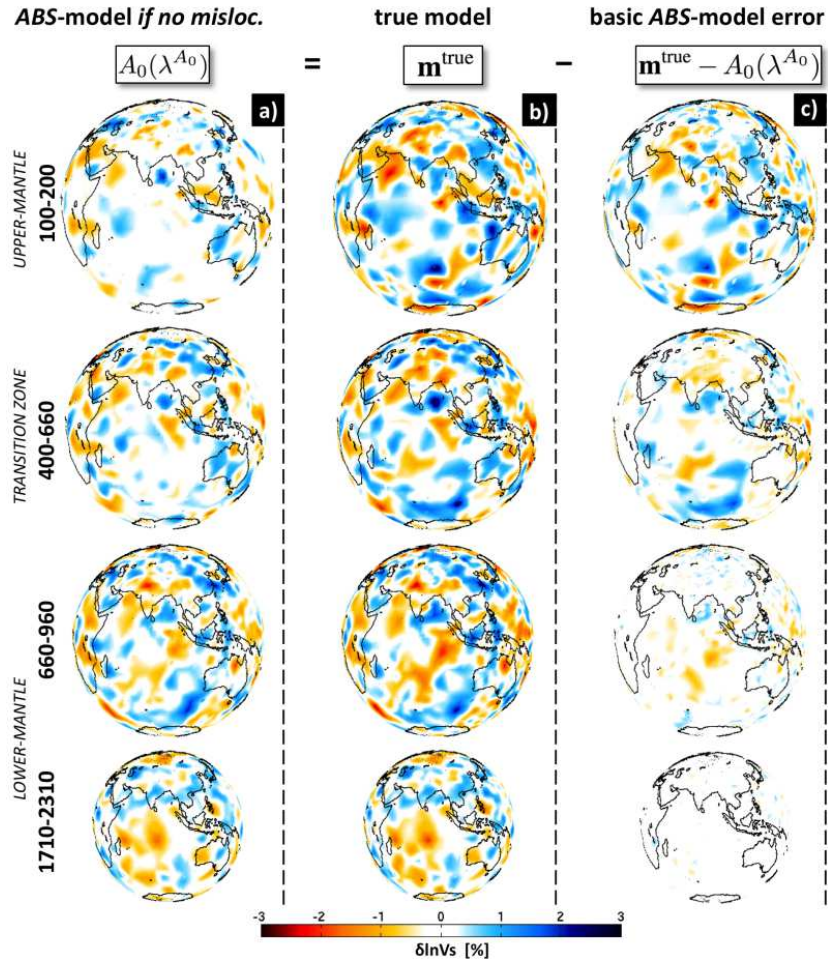**Figure 2.** First true-model input, m^{true} (Gaussian Random Field).
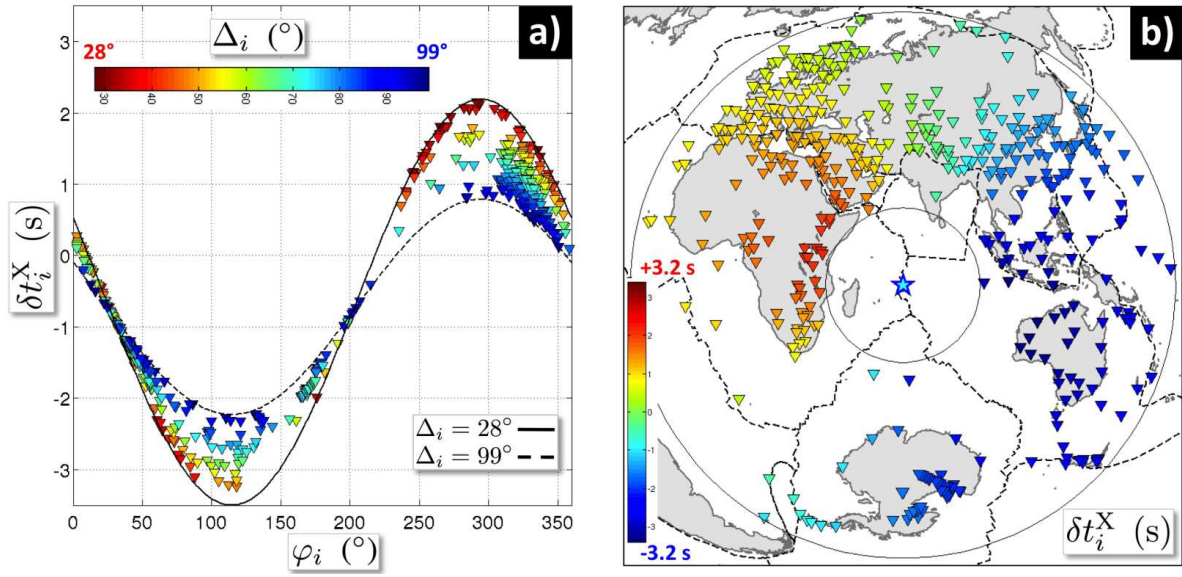
Figure 3. An example of mislocation residual-times, δt_i^{X}, for direct *S* phases measured at receivers i (cf., Eq. 6). Mislocation parameters are: $(\delta x, \delta y, \delta z) = (18.3, -8.3, 2.5)\,\mathrm{km}$, corresponding to an horizontal mislocation $\delta \ell \sim 20\,\mathrm{km}$. a) Solid and dashed black lines are theoretical mislocation residual-times computed using Eqs. 7 and 8, at $28\,^\circ$ and $99\,^\circ$ epicentral distance (Δ_i), respectively. φ_i denotes the earthquake-receiver azimuth. b) Circles denote an epicentral distance of $28\,^\circ$ and $99\,^\circ$, and the blue star represents the earthquake epicenter.
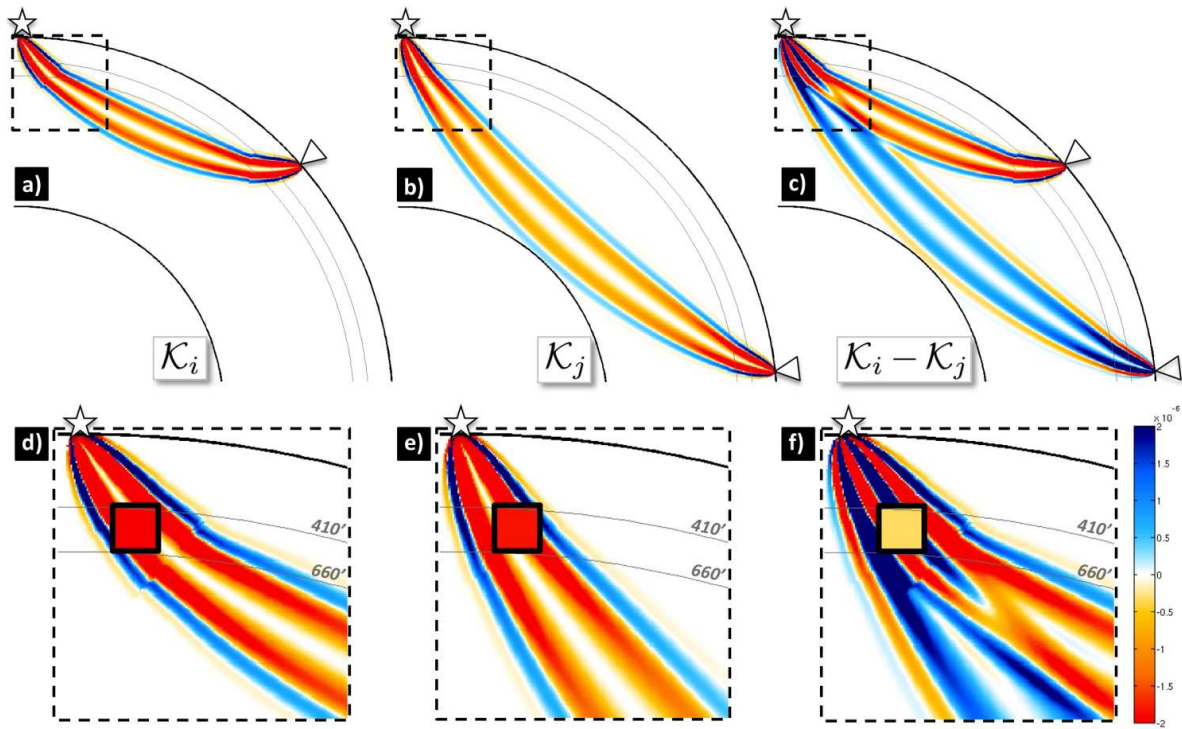
Figure 4. Illustration of the sensitivity corresponding to the difference of finite-frequency kernels, for direct *S*-waves, in an "ideal" case where: $\{\Delta_i = 50°, \Delta_j = \Delta_i + 35°, \varphi_i = \varphi_j\}$. a–c) $\mathcal{K}_i$, $\mathcal{K}_j$, $\mathcal{K}_i - \mathcal{K}_j$, d–f) corresponding zoom-in near the source region, respectively. One sees that $\mathcal{K}_i - \mathcal{K}_j$ can become less sensitive to the structure on the source side, as illustrated with the averaged kernel sensitivity to a square-shaped velocity anomaly of 200 km length inside the transition zone. Kernels are computed at $T = 20$ s period; units are s.km^{-3}.
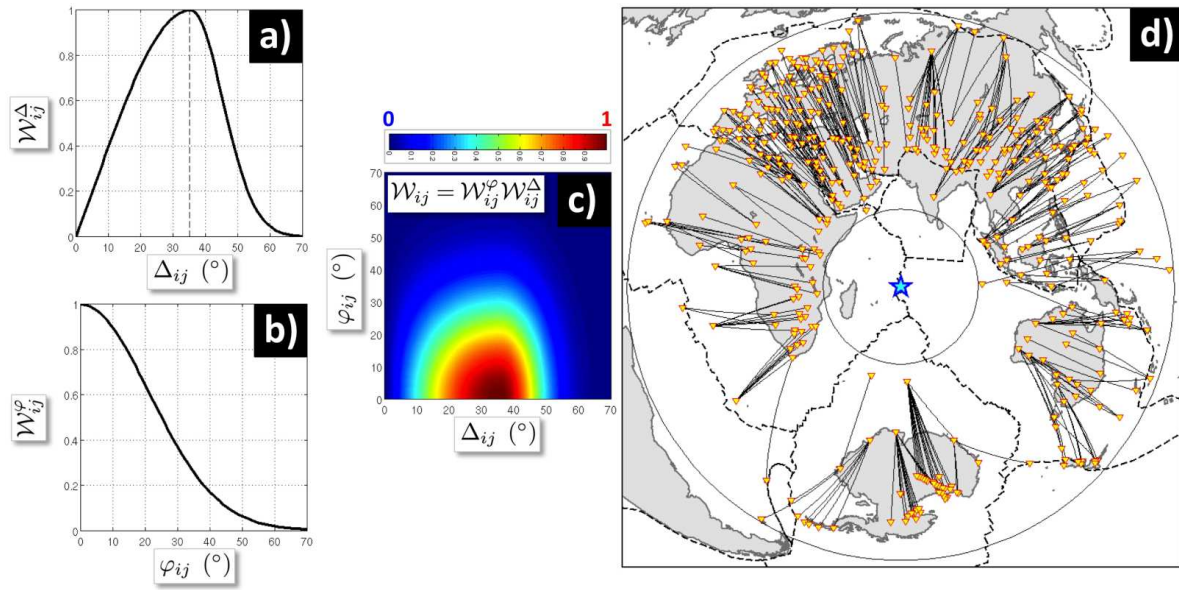
Figure 5. a–c) Illustration of the weight-functionals, $\mathcal{W}_{ij}(\varphi_{ij}, \Delta_{ij}) = \mathcal{W}_{ij}^{\varphi}(\varphi_{ij}) \times \mathcal{W}_{ij}^{\Delta}(\Delta_{ij})$, used for selecting the optimised receiver pairs (i, j), for each event, where $\varphi_{ij} = |\varphi_i - \varphi_j|$, and $\Delta_{ij} = |\Delta_i - \Delta_j|$. d) The best-partner receiver pairs are shown with black solid lines.
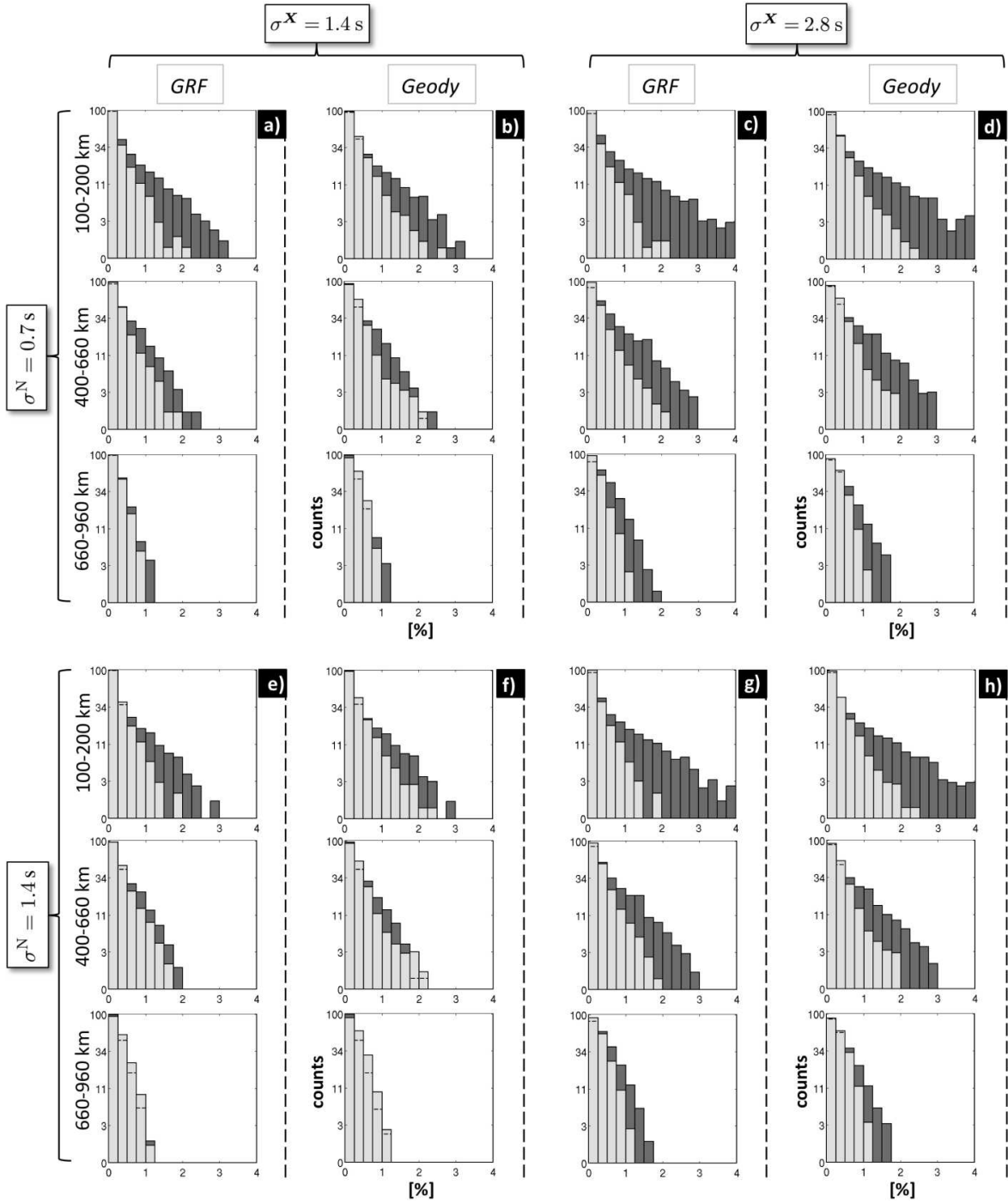
Figure 6. Comparison of the first true-model input (Gaussian Random Field) with its corresponding reference absolute-model. a) $A_0(\lambda^{A_0})$, b) m^{true}, c) $m^{\text{true}} - A_0(\lambda^{A_0})$. Note that in order to visually enlarge the illustrative results of our approach, we do not use a whole Earth projection in Figs. 6–8 (geographical considerations are not within our scope), but we show the statistics of our results over the entire mantle in Fig. 9.
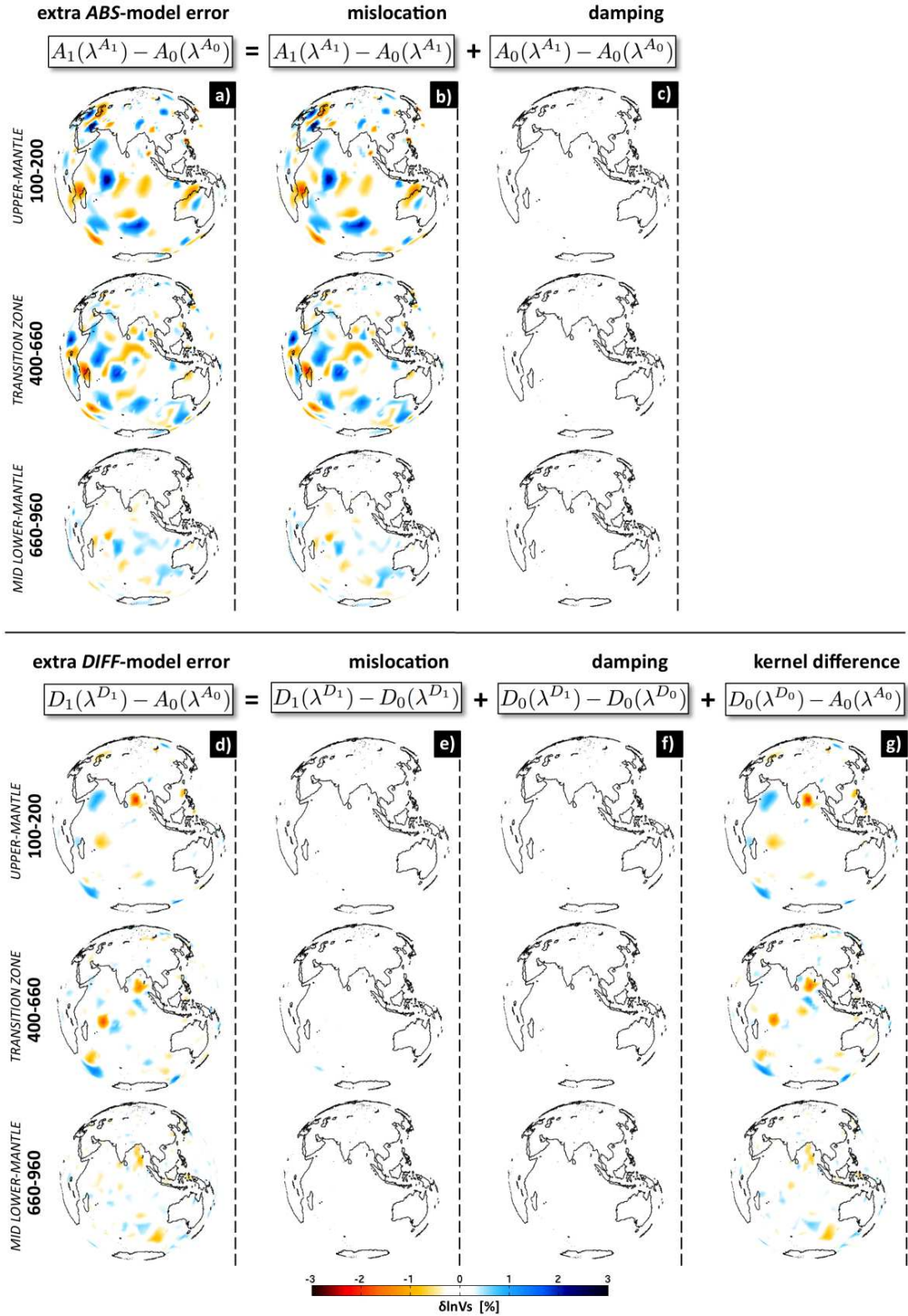
Figure 7. Case of low mislocation-bias ($\sigma^X = 1.4$ s): comparison of the extra errors for absolute- and differential models (cf., Eqs. 24–25). a) $\{A_1(\lambda^{A_1}) - A_0(\lambda^{A_0})\}$, b) $\{A_1(\lambda^{A_1}) - A_0(\lambda^{A_1})\}$, c) $\{A_0(\lambda^{A_1}) - A_0(\lambda^{A_0})\}$, d) $\{D_1(\lambda^{D_1}) - A_0(\lambda^{A_0})\}$, e) $\{D_1(\lambda^{D_1}) - D_0(\lambda^{D_1})\}$, f) $\{D_0(\lambda^{D_1}) - D_0(\lambda^{D_0})\}$, g) $\{D_0(\lambda^{D_0}) - A_0(\lambda^{A_0})\}$.
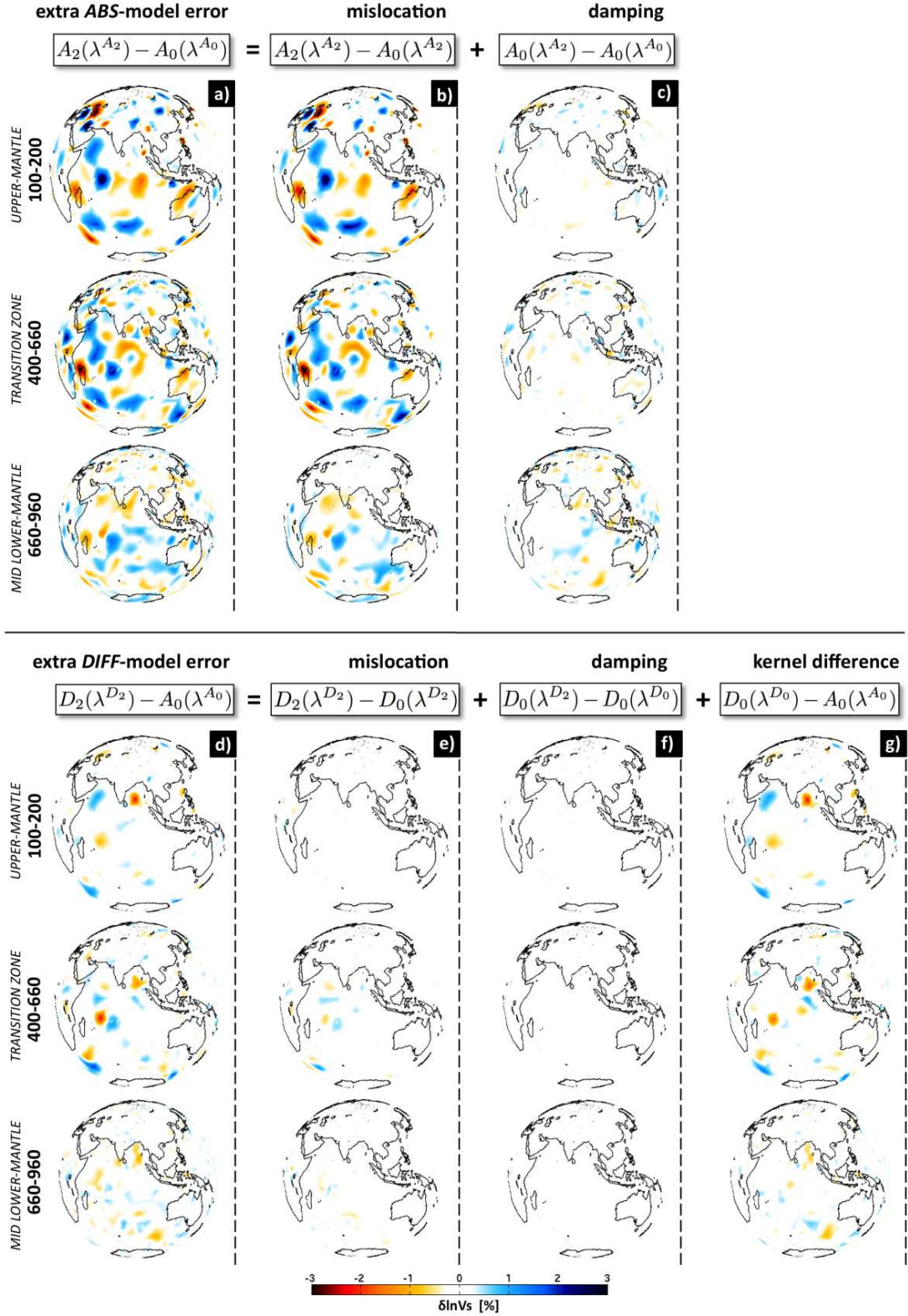
Figure 8. Case of high mislocation-bias ($\sigma^X = 2.8$ s): comparison of the extra errors for absolute- and differential models (cf., Eqs. 24–25). a) $\{\boldsymbol{A}_2(\lambda^{A_2}) - \boldsymbol{A}_0(\lambda^{A_0})\}$, b) $\{\boldsymbol{A}_2(\lambda^{A_2}) - \boldsymbol{A}_0(\lambda^{A_2})\}$, c) $\{\boldsymbol{A}_0(\lambda^{A_2}) - \boldsymbol{A}_0(\lambda^{A_0})\}$, d) $\{\boldsymbol{D}_2(\lambda^{D_2}) - \boldsymbol{A}_0(\lambda^{A_0})\}$, e) $\{\boldsymbol{D}_2(\lambda^{D_2}) - \boldsymbol{D}_0(\lambda^{D_2})\}$, f) $\{\boldsymbol{D}_0(\lambda^{D_2}) - \boldsymbol{D}_0(\lambda^{D_0})\}$, g) $\{\boldsymbol{D}_0(\lambda^{D_0}) - \boldsymbol{A}_0(\lambda^{A_0})\}$.

Figure 9. Normalised histograms of the extra errors for absolute- and differential-models, computed over the whole-mantle and shown for three depth ranges. Dark-grey corresponds to $|\boldsymbol{A}_k(\lambda^{A_k}) - \boldsymbol{A}_0(\lambda^{A_0})|$, and light-grey to $|\boldsymbol{D}_k(\lambda^{D_k}) - \boldsymbol{A}_0(\lambda^{A_0})|$, where k $= \{1; 2\}$ denotes the two mislocation regimes, $\sigma^X = \{1.4; 2.8\}$ s, respectively. Two different true-model inputs are considered, Gaussian Random Field (GRF) and Geodynamical (Geody), as well as two regimes of noise-related errors, $\sigma^N = \{0.7; 1.4\}$ s. A logarithmic scale is used to facilitate the comparison of extra errors.
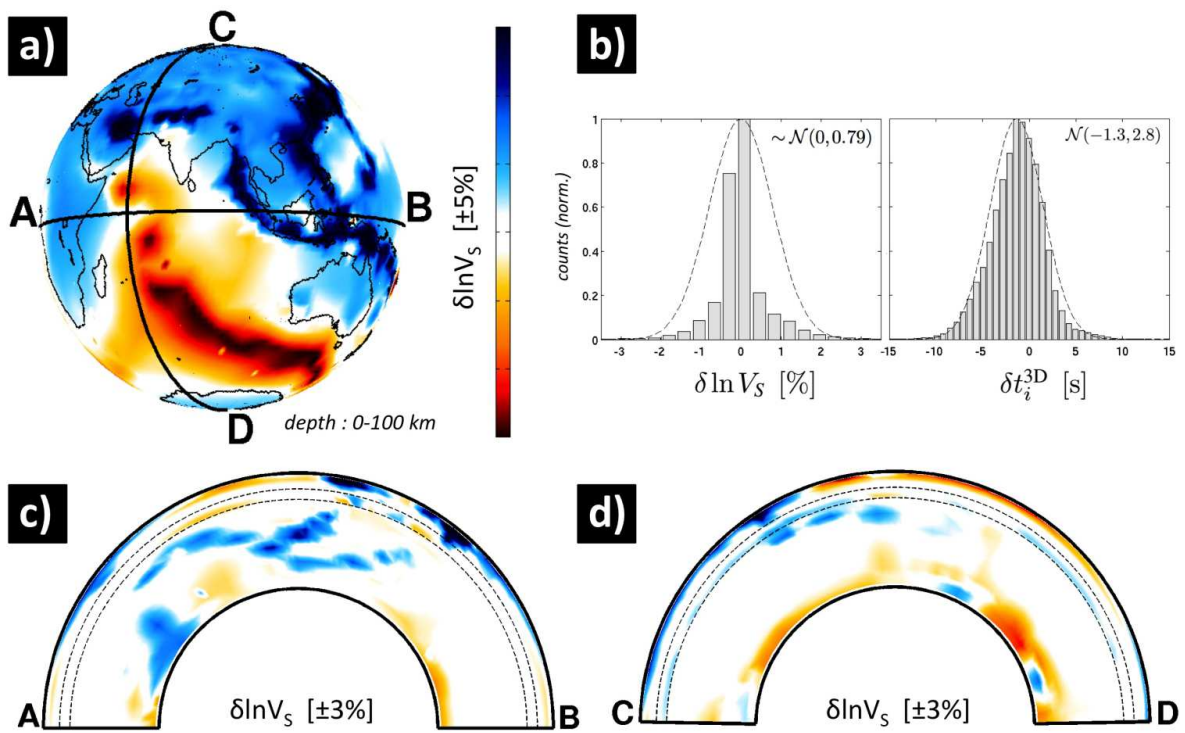
Figure 10. Second true-model input, $m_{\text{Geody}}^{\text{true}}$ (Geodynamical).

APPENDIX A: SOURCE PROPAGATION, CROSS-CORRELATION TIME-RESIDUALS, AND RECEIVER PAIRS

Since the beginning of the era of digital instrumentation, a popular way to efficiently measure seismic time-residuals has been to apply cross-correlation techniques (e.g., VanDecar & Crosson 1990). Cross-correlation measurements are affected by source propagation effects that should be dealt with, no matter whether the data inversion is based on ray-theoretical or finite-frequency approaches. A natural way to deal with source propagation is to include it in the computation of synthetic seismograms. However, it is often not possible to do so because source kinematics is poorly known in most cases. It turns out that the way we deal with source mislocations in this study (through the use of receiver pairs) could also be efficient for rejecting a large part of unilateral rupture propagation effects in teleseismic cross-correlation body-wave delay-times. The purpose of this appendix is to illustrate this point.

First, let us consider for a given earthquake the observed and synthetic waveforms of a direct S phase recorded at receiver i, denoted by $u_i^{\mathrm{obs}}(t)$ and $u_i^{\mathrm{syn}}(t)$, respectively. In finite-frequency tomography (e.g., Dahlen et al. 2000), for instance, the correlation delay-time is defined as the time-lag maximising the cross-correlation function between $u_i^{\mathrm{obs}}(t)$ and $u_i^{\mathrm{syn}}(t)$ over some time-window (cf., Eq. 1). The cross-correlation function γ of two signals $s_1(t)$ and $s_2(t)$ is defined in the spectral domain such that:

$$\gamma[\mathcal{S}_1; \mathcal{S}_2](\omega) = \mathcal{S}_1(\omega)\mathcal{S}_2^*(\omega), \tag{A.1}$$

where $\mathcal{S}(\omega) = \mathcal{FT}\{s(t)\}$ denotes the Fourier Transform of signal $s(t)$, and * denotes the complex conjugate. If one considers a rupture history $m(t)$ giving rise to a source time function $\dot{m}(t)$ in the far-field, with spectrum denoted as $\dot{m}(\omega)$, we have:

$$\begin{cases} \mathcal{U}_i^{\mathrm{syn}}(\omega) = \dot{m}(\omega)\mathrm{e}^{\mathrm{i}\omega T_i^{\mathrm{syn}}} \\ \mathcal{U}_i^{\mathrm{obs}}(\omega) = \dot{m}(\omega)\mathrm{e}^{\mathrm{i}\omega T_i^{\mathrm{obs}}}, \end{cases} \tag{A.2}$$

where $T_i^{\mathrm{obs}} = t_i^{\mathrm{1D}} + \delta t_i$ and $T_i^{\mathrm{syn}} = t_i^{\mathrm{1D}}$ are the observed and predicted travel times, and $\delta t_i = \delta t_i^{\mathrm{3D}} + \delta t_i^{\mathrm{N}} + \delta t_i^{\mathrm{X}}$ is the simple correlation delay-time due to 3–D shear-velocity anomalies, noise-measurement, and source mislocation, respectively (cf., Sect. 2.2). However, even in a very smooth Earth model with no significant diffraction, the shape of the waveform can be influenced by the kinematics of the source if the fault length L is large. In the case of unilateral rupture propagation (e.g., Aki & Richards 2002), one can write the observed seismic waveform in the spectral domain such that:

$$\begin{cases} \widehat{\mathcal{U}}_i^{\mathrm{obs}}(\omega) = \mathcal{U}_i^{\mathrm{obs}}(\omega)\frac{\sin X_i(\omega)}{X_i(\omega)}\mathrm{e}^{\mathrm{i}X_i(\omega)} \\ X_i(\omega) = \omega\frac{L}{2}(\frac{1}{v} - \frac{\cos \Psi_i}{c}), \end{cases} \tag{A.3}$$

where Ψ_i is the angle between the ray direction to receiver i and the direction of rupture propagation, v

is the rupture velocity, and c is the local shear-velocity within the source region. For such an unilateral rupture, one can write the cross-correlation function of the observed and synthetic waveforms in the spectral domain as:

$$\gamma[\widehat{\mathcal{U}}_i^{\text{obs}}; \mathcal{U}_i^{\text{syn}}](\omega) = |\dot{m}(\omega)|^2 \frac{\sin X_i(\omega)}{X_i(\omega)} \mathrm{e}^{\mathrm{i}\omega(\delta t_i + \delta t_i^{\text{R}})}, \tag{A.4}$$

where $\delta t_i^{\text{R}} = \frac{L}{2}\left(\frac{1}{v} - \frac{\cos\Psi_i}{c}\right)$. Since the phase shift in the exponential term of the cross-correlation is proportional to ω, it is a pure travel time error, independent of frequency. Therefore, the actual cross-correlation delay-time $\widehat{\delta t}_i$ at receiver i can be expressed as:

$$\widehat{\delta t}_i = \delta t_i + \delta t_i^{\text{R}}. \tag{A.5}$$

The angle Ψ_i can be viewed as analogous to the angle Φ_i in Eq. 8, if one replaces the mislocation vector by the vector of rupture propagation. Similarly to source mislocation effects, the unilateral rupture propagation residual-times, δt_i^{R}, thus depend to first order on the earthquake-receiver azimuth, and to second order on the epicentral distance – assuming that the source propagation is predominantly horizontal (cf., Sect. 2.3), which should be verified for a certain number of events with $m_{\text{b}} \sim 6$. The differential delay-times for a couple of receivers (i, j) can therefore be written as:

$$\{\widehat{\delta t}_i - \widehat{\delta t}_j\} = \{\delta t_i - \delta t_j\} + \{\delta t_i^{\text{R}} - \delta t_j^{\text{R}}\}, \tag{A.6}$$

where the differential effect of rupture propagation becomes:

$$\{\delta t_i^{\text{R}} - \delta t_j^{\text{R}}\} = -\frac{L}{2c} \times (\cos\Psi_i - \cos\Psi_j). \tag{A.7}$$

Note that Eq. A7 is highly similar to Eq. 11. Therefore, using optimised receiver pairs (i, j), as defined in Sect. 3 to deal with mislocations, would also lead to minimise the term $(\cos\Psi_i - \cos\Psi_j)$, and thus to almost cancel out the effect of unilateral rupture propagation in differential cross-correlation delay-times $\{\widehat{\delta t}_i - \widehat{\delta t}_j\}$, in addition to get rid of a large part of errors in origin-time and location. Determining whether this azimuthal part is dominant in source propagation effects would require to elaborate tomographic tests similar to those conducted in Sect. 4. However, to do so would require a good knowledge of the statistics of rupture propagation for a significant set of worldwide earthquakes, which is not currently available.