



An objective rationale for the choice of regularisation parameter with application to global multiple-frequency *S*-wave tomography

C. Zaroli¹, M. Sambridge², J.-J. Lévêque¹, E. Debayle³, and G. Nolet⁴

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

²Research School of Earth Sciences, Australian National University, Canberra ACT 0200, Australia

³Laboratoire de Géologie de Lyon, UMR5276, CNRS, Université Lyon 1 et ENS de Lyon, France

⁴Géoazur, Université de Nice Sophia-Antipolis, UMR6526, CNRS, France

Correspondence to: C. Zaroli (c.zaroli@unistra.fr)

Received: 4 June 2013 – Published in Solid Earth Discuss.: 28 June 2013

Revised: 29 August 2013 – Accepted: 10 September 2013 – Published: 15 October 2013

Abstract. In a linear ill-posed inverse problem, the regularisation parameter (damping) controls the balance between minimising both the residual data misfit and the model norm. Poor knowledge of data uncertainties often makes the selection of damping rather arbitrary. To go beyond that subjectivity, an objective rationale for the choice of damping is presented, which is based on the coherency of delay-time estimates in different frequency bands. Our method is tailored to the problem of global multiple-frequency tomography (MFT), using a data set of 287 078 *S*-wave delay times measured in five frequency bands (10, 15, 22, 34, and 51 s central periods). Whereas for each ray path the delay-time estimates should vary coherently from one period to the other, the noise most likely is not coherent. Thus, the lack of coherency of the information in different frequency bands is exploited, using an analogy with the cross-validation method, to identify models dominated by noise. In addition, a sharp change of behaviour of the model ℓ^∞ -norm, as the damping becomes lower than a threshold value, is interpreted as the signature of data noise starting to significantly pollute at least one model component. Models with damping larger than this threshold are diagnosed as being constructed with poor data exploitation. Finally, a preferred model is selected from the remaining range of permitted model solutions. This choice is quasi-objective in terms of model interpretation, as the selected model shows a high degree of similarity with almost all other permitted models (correlation superior to 98 % up to spherical harmonic degree 80). The obtained tomographic model is displayed in the mid lower-mantle (660–1910 km depth), and is shown to be compatible with three other recent

global shear-velocity models. A wider application of the presented rationale should permit us to converge towards more objective seismic imaging of Earth's mantle.

1 Introduction

Until recently, ray theory (RT) formed the backbone of global seismic tomography, mainly because of its simplicity and computational efficiency (e.g. Grand et al., 1997; Van der Hilst et al., 1997; Ritsema et al., 1999, 2011; Fukao et al., 2001; Debayle et al., 2005). RT is based on the approximation that seismic waves have an infinite frequency, or a zero wavelength. It assumes that body-wave travel times, which represent onset times, are only dependent upon Earth's structure along geometrical ray paths. In reality, body-waves observed on broadband seismographs have wavelengths ranging from 10 to 1000 km, or even more. Digital instrumentation has led to the use of cross-correlation to determine travel times. However, the cross-correlation time window allows low frequencies to influence the measurements. Therefore, RT can break down when used with cross-correlation measurements for imaging small-scale heterogeneities, since diffraction effects make travel times (and amplitudes) dependent on a 3-D region around the ray path. A recent focus has been to take into account the finite-frequency (FF) behaviour of body-waves. This has been supported by continued systematic evidence for body-wave travel time dispersion due to various forms of scattering (e.g. Hung et al., 2004; Yang et al., 2006; Sigloch and Nolet, 2006; Zaroli et al.,

2010). To better constrain the structure of the Earth's interior, new theoretical developments on seismic wave propagation have emerged in recent years, and received increasing attention in tomography. Dahlen et al. (2000) developed an FF approach (hereafter referred to as banana–doughnut theory, BDT) that is efficient enough to be applied to large-scale problems and a wide range of frequencies. It is based on a ray–Born approximation that is much faster than the mode summation approaches proposed earlier by Marquering et al. (1998) and Zhao and Jordan (1998). We refer the reader to Sect. 1.6 of Nolet (2008) for more complete references to work preceding Dahlen et al. (2000). To go beyond tomographic limitations from ray-based FF kernels of BDT (e.g. Nolet, 2008), alternatives are available (e.g. Zhao et al., 2000; Tromp et al., 2005; Nissen-Meyer et al., 2007; Zhao and Chevrot, 2011a, b). The most promising ones are based on numerical techniques (e.g. Komatitsch et al., 2002) and the adjoint method (e.g. Tarantola, 1987; Tromp et al., 2005). Such large scale tomographic applications are currently limited to periods greater than ~ 50 s, where the dominant part of the signal corresponds to surface waves (e.g. Fichtner et al., 2009; Tape et al., 2010; Lekic and Romanowicz, 2011). Mercierat and Nolet (2012) investigate the accuracy of BDT delay-time predictions and conclude that, even though errors with V_S kernel predictions are larger than those for V_P , the errors remain well below typical observational uncertainty, while the kernels computed this way require two to three orders of magnitude less CPU time. In the past 10 yr, various applications of BDT have shown interesting tomographic results (e.g. Montelli et al., 2004b, 2006; Hung et al., 2004; Yang et al., 2006, 2009; Sigloch et al., 2008; Sigloch, 2011; Sigloch and Mihalynuk, 2013; Nolet, 2009; Tian et al., 2009, 2011). In parallel, several studies questioned whether the benefits from using BDT, rather than simple RT, could be smaller than the uncertainty posed by the subjective regularisation of the inverse problem (e.g. Sieminski et al., 2004; Trampert and Spetzler, 2006; Van der Hilst and de Hoop, 2005; Boschi et al., 2006). It seems pertinent that tomographers ask what are the consequences on seismic models of the subjectivity inherent to the choice of regularisation parameter. Linear tomographic inverse problems, $\mathbf{d} = \mathbf{G}\mathbf{m}$, are usually ill-posed and require a subjective degree of regularisation to deal with data errors and stabilise the solution. One usually uses Tikhonov regularisation (Tikhonov, 1963), which consists in solving the minimisation problem $\mathbf{m}^\lambda = \arg \min(\|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \lambda^2\|\mathbf{m}\|_2^2)$, where λ is a real regularisation parameter (damping) to be chosen with care. A convenient and widely used graphical tool (e.g. Aster et al., 2012) for setting the damping λ is to analyse the trade-off curve (L curve) between the model norm ($\|\mathbf{m}\|_2^2$, squared Euclidean norm) and the data misfit (χ_{red}^2 , reduced chi-square). In a tomographic experiment, data errors are a mix of observational and modelling errors. Ideally, if the statistics of data errors would perfectly be known, the op-

timal solution would be reached for $\chi_{\text{red}}^2 \simeq 1$ near the bend of the L curve. In practice this is never met, as data errors are usually just best guesses, and one faces the dilemma to choose a solution around the L curve's corner as a best compromise between minimising both the residual data misfit and the model norm. Therefore, the selection of an optimally regularised solution is to a large extent arbitrary – more akin to an art form – which can lead to different model interpretations. To converge towards more reliable tomographic models, one would prefer to lessen the inherent subjectivity of the damping choice. The goal of this study is to present a new approach aimed at objectifying the choice of regularisation parameter in the case of global multiple-frequency tomography (MFT, equivalent to BDT applied in multiple-frequency bands, as named by Sigloch et al., 2008). We will illustrate our approach using a global data set of 287 078 shear-wave delay times measured at 10, 15, 22, 34 and 51 s central periods (Zaroli et al., 2010). To summarise, our approach will consist of the following three parts: (i) identifying models with poor data exploitation (over-damped); (ii) identifying models dominated by noise (under-damped); and (iii) selecting a preferred model among the remaining ones. We will show that this final choice is of no consequence in terms of model interpretation, and thus is quasi-objective. The preferred model we obtain this way will be displayed in mid lower-mantle (660–1910 km depth), where our data coverage is at its highest. We will briefly discuss its major structural features, and show its compatibility with three other, latest generation, global shear-velocity models.

2 Global multiple-frequency S-wave tomography

2.1 A global data set of multiple-frequency S-wave delay-times

We use a globally distributed data set of 287 078 S and SS delay-times measured at 10, 15, 22, 34 and 51 s periods (Zaroli et al., 2010), as summarised in Table 1. Both single-phase (S , SS) and two-phase interference ($S+sS$, $SS+sSS$) delays are considered. Measurements are performed on the transverse component by cross-correlating the observed and synthetic waveforms. Synthetics are computed with the WKBJ method (Chapman, 1978) within the spherical IASP91 1-D shear-wave velocity model (Kennett and Engdahl, 1991) extended with PREM's 1-D attenuation (q_0) model (Dziewonski and Anderson, 1981). The global $2^\circ \times 2^\circ$ crustal model CRUST2.0 (Bassin et al., 2000) is used for incorporating most of crustal reverberations, on receiver side, in the synthetic waveforms. As seismic waves propagate through the Earth, an attenuating medium, a physical dispersion correction should be applied to cross-correlation delay times. Zaroli et al. (2010) show that “the globally averaged dispersion observed for S and SS delay-times favour a frequency-dependent 1-D attenuation model $q(\omega) \propto q_0 \times$

Table 1. Global data set of multiple-frequency shear-wave delay-times.

Period	10 s	15 s	22 s	34 s	51 s	Total
<i>S</i>	15 739	31 264	43 263	43 263	35 457	169 706
<i>SS</i>	2763	14 518	36 142	36 142	27 807	117 372
Total	18 502	45 782	79 765	79 765	63 264	287 078

$\omega^{-\alpha}$, with an α value of 0.2 for *S* and 0.1 for *SS*.” Thus, *S* and *SS* delays are corrected for physical dispersion by injecting those α values in Eq. (16) of Zaroli et al. (2010). After correction for physical dispersion due to 1-D attenuation $q(\omega)$ and crustal reverberations, the data exhibit a residual dispersion of the order of 1–2 s in the 10–51 s period range. Zaroli et al. (2010) suggest that this residual dispersion is partly related to seismic heterogeneities in the mantle. For instance, they show that wavefront-healing phenomenon (e.g. Gudmundson, 1997) is clearly observed for *S* waves having passed through negative V_S anomalies. Tian et al. (2011) show that the inclusion of additional dispersion due to 3-D attenuation structure has little influence on *S*-wave models. Sigloch et al. (2008) show that one can neglect it for *P*-waves, and, similarly, Savage et al. (2010) observe only a small effect on seismic wave travel times. Thus, we neglect the role of 3-D variations of attenuation to explain the residual dispersion in our data. Bolton and Masters (2001) extensively discuss the assignment of quantitative errors in the case of a global *S* wave data set. Following their analysis, we aim at identifying the separate contributions to the total data variance σ_T^2 in our data set. We have: $\sigma_T^2 = \sigma_{3-D}^2 + \sigma_X^2 + \sigma_N^2$, where σ_{3-D}^2 is due to 3-D seismic heterogeneities, σ_X^2 is due to earthquake location errors, and σ_N^2 is attributed to measurement errors. Bolton and Masters (2001) estimate σ_X for *S* waves to be 1.6–2.5 s, assuming a typical depth uncertainty of about 10 km, at epicentral distances of about 70°, and for mislocation vectors of length 10–20 km. In this study, we have attributed to each datum d_i a constant value for the source uncertainty $\sigma_X = 2.5$ s. Earthquake mislocation and origin time effects are not simultaneously estimated for each datum during the inversion. In addition, we assume that the source uncertainty is the same for an *S*-wave measured at either 10, 15, 22, 34, or 51 s period, so that taking it into account would not change the main results of this study. Moreover, Masters et al. (2000) report that their tomographic results of *S* wave inversions vary little when source effects are included. As for the contribution of the measurement error, we have attributed to each datum d_i an individual error $\{\sigma_N\}_i$ computed with Eq. (9) of Zaroli et al. (2010). It varies between 0.1 and 3.7 s, though its average value is 0.49, 0.57, 0.67, 0.73, and 1.08 s for data subsets at 10, 15, 22, 34, and 51 s periods, respectively. Thus, the total data uncertainty σ_i for each datum d_i is $\sigma_i = (\sigma_X^2 + \{\sigma_N\}_i^2)^{1/2}$.

2.2 Setting up the inverse problem

At each period a waveform is influenced by a weighted average of Earth’s mantle through its corresponding 3-D sensitivity kernel. In principle, measuring the delay-time of a seismic phase at several periods should increase the amount of independent information in the inverse problem, and lead to improved tomographic imaging. The general form of the MFT inverse problem is

$$\delta t_i(T) = \int_{V_i(T)} K_i(\mathbf{r}; T) m(\mathbf{r}) d^3\mathbf{r}, \quad (1)$$

where $\delta t_i(T)$ is the time residual of target seismic phase i measured in a passband with dominant wave period T . The volume $V_i(T)$ is limited to the region where the amplitude of the sensitivity (Fréchet) kernel $K_i(\mathbf{r}; T)$ is significant. The model parameter $m(\mathbf{r})$ represents a velocity perturbation with respect to the 1-D reference velocity model IASP91 at each point \mathbf{r} in the medium. This linear inverse problem can be formulated as $\mathbf{d} = \mathbf{G}\mathbf{m}$, where \mathbf{d} and \mathbf{m} are vectors of data (size N) and model parameters (size M), respectively. The \mathbf{G} sensitivity matrix (size $N \times M$) is the projection of the Fréchet kernels onto the model parameterisation. In Appendices A and B, we expand on the computation of (i) the data-driven model parameterisation, and (ii) the “analytical” ray-based FF kernels. For more details, including for the computation of \mathbf{G} , we refer the reader to Zaroli (2010). In order to solve this linear inverse problem, we assume that the prior covariance matrices of the data, \mathbf{C}_d , and of the model parameters, \mathbf{C}_m , follow Gaussian probability functions. Thus, we may obtain the maximum-likelihood estimate of the model solution by minimising (e.g. Tarantola and Necessian, 1984; Tarantola, 1987): $f(\mathbf{m}) = (\mathbf{G}\mathbf{m} - \mathbf{d})^t \mathbf{C}_d^{-1} (\mathbf{G}\mathbf{m} - \mathbf{d}) + \mathbf{m}^t \mathbf{C}_m^{-1} \mathbf{m}$; which leads to solving a system of normal equations:

$$\begin{pmatrix} \mathbf{C}_d^{-1/2} \mathbf{G} \\ \mathbf{C}_m^{-1/2} \end{pmatrix} \mathbf{m} = \begin{pmatrix} \mathbf{C}_d^{-1/2} \mathbf{d} \\ \mathbf{0} \end{pmatrix}. \quad (2)$$

For each source–receiver pair, uncertainties associated with delays should partially be correlated at different periods. For instance, because of common errors on source location and origin time. So that the matrix \mathbf{C}_d should contain off-diagonal terms. However, for simplicity, we use a diagonal data covariance matrix of the form: $\mathbf{C}_d = \text{diag}(\sigma_i^2)_{1 \leq i \leq N}$. Tomographic inverse problems are usually ill-posed and require a degree of regularisation to deal with data errors and stabilise the solution. The model covariance, \mathbf{C}_m , represents our prior expectation of how model parameters are correlated, e.g. high correlation for nearby parameters. We use a simple model covariance of the form: $\mathbf{C}_m = \sigma_m^2 \mathbf{I}_M$, where \mathbf{I}_M is the identity matrix of size $M \times M$. The parameter σ_m influences the solution by damping the model norm and allows us to regularise the inverse problem. Because FF kernels

integrate over a volume as wide as several Fresnel zones, a simple regularisation parameter (damping) $\lambda = 1/\sigma_m$ is sufficient, in our experience, to obtain smoothed solutions. Setting $\mathbf{G}' = \mathbf{C}_d^{-1/2}\mathbf{G}$ and $\mathbf{d}' = \mathbf{C}_d^{-1/2}\mathbf{d}$, one can write (Eq. 2) as

$$\begin{pmatrix} \mathbf{G}' \\ \lambda \mathbf{I}_M \end{pmatrix} \mathbf{m} = \begin{pmatrix} \mathbf{d}' \\ \mathbf{0} \end{pmatrix}. \quad (3)$$

To solve the weighted, damped least-squares problem (Eq. 3), we use LSQR (Paige and Saunders, 1982), an iterative row action method that converges to solution $\mathbf{m}^\lambda = \arg \min(\|\mathbf{d}' - \mathbf{G}'\mathbf{m}\|_2^2 + \lambda^2\|\mathbf{m}\|_2^2)$. Care has been taken to ensure LSQR convergence with sufficient iterations. After a first inversion with zero damping, outliers with data misfit deviations larger than $\pm 3\sigma$ were rejected for subsequent inversions. Total amount of surviving data is $N = 287078$ (cf., Table 1). Let us define $\chi^2 = \sum_{i=1}^N (\{\mathbf{d}'\}_i - \{\mathbf{G}'\mathbf{m}^\lambda\}_i)^2 = \sum_{i=1}^N (\{\mathbf{d}\}_i - \{\mathbf{G}\mathbf{m}^\lambda\}_i)^2/\sigma_i^2$. It is a direct measure of the data misfit, in which we weight the misfits inversely with their standard errors. In the perfect case that data are on average satisfied with a misfit of one standard deviation, we find $\chi^2 \simeq N$, so that it is convenient to define the reduced chi square $\chi_{\text{red}}^2 = \chi^2/N$. As always, the crucial choice of damping λ controls the balance between both minimising the model norm $\|\mathbf{m}\|_2^2$ and data misfit χ_{red}^2 . In the following, we present an objective rationale for the choice of λ that is tailored to the MFT problem.

3 An objective rationale for the choice of damping

The MFT inverse problem (Eq. 1) is ill-posed in the sense that, in absence of regularisation, small changes in the data can lead to large changes in the computed model solution. This forces us to impose some sort of regularisation of the problem to filter out the noise effects, which leads to solving (Eq. 3). The big question is how to choose the regularisation parameter (damping λ) that gives an appropriate balance between filtering out enough noise without losing structural information. To converge towards more reliable tomographic images, there is a need to significantly lessen the subjectivity associated to the damping choice. Not doing so could prevent us from correctly mapping into the model the weak *S*-wave delay-time dispersion observed in our global data set. First, let us define some useful notations. Let \mathbf{d}_T be the data subset with central period T , \mathbf{MB}^λ the multi-band model from inversion of $\mathbf{d}_{10,15,22,34,51}$ with damping λ , $\mathbf{SB}_{34}^{\lambda'}$ the single-band model from inversion of \mathbf{d}_{34} with damping λ' , and $\chi_{\text{red}}^2(\mathbf{m}, \mathbf{d}_T)$ the misfit of \mathbf{d}_T with model \mathbf{m} . Perhaps the most convenient and widely used graphical tool for setting the damping is the *L* curve analysis. It usually consists in analysing the curve of trade-off between the size of the model (e.g. measured by $\|\mathbf{m}\|_2^2$) and the data misfit (e.g. measured by χ_{red}^2). One usually gets a characteristic L-shaped curve, with a (not often distinct) corner

separating “vertical” and “horizontal” parts of the curve. Figure 1 shows the curve of trade-off between $\|\mathbf{MB}^\lambda\|_2^2$ and $\chi_{\text{red}}^2(\mathbf{MB}^\lambda, \mathbf{d}_{10,15,22,34,51})$. One sees that the model with data misfit $\chi_{\text{red}}^2(\mathbf{MB}^\lambda, \mathbf{d}_{10,15,22,34,51}) = 1$ is probably under-damped as it lies on the quasi-vertical leg of the *L* curve. This indicates that this model is dominated by noise, as will be confirmed in Sect. 3.2, and is a reminder of our imperfect knowledge of the statistics of data errors (σ_i). Our estimates of σ_i appear to be slightly under-estimated, which may be caused by non-Gaussian distributions of source and measurement uncertainties. It is possible to estimate observational errors of teleseismic waves using the concept of summary rays (e.g. Morelli and Dziewonski, 1987; Gudmundsson et al., 1990). However, the problem is larger for MFT, since very little experience so far is at hand on the observational error sources (e.g. crustal reverberations) that affect narrow-band cross-correlation estimates. Hansen and O’Leary (1993) suggest that the optimal damping could be the one maximising the *L* curve’s curvature (*maximum curvature criterion*). Because of some intrinsic arbitrariness in this criterion, it should be applied with caution (e.g. Boschi et al., 2006). For instance, the *L* curve has to be a plot of un-dimensional quantities, by scaling both data and model to prior uncertainty. We compute the curvature as $\kappa(\lambda) = (\rho'\eta'' - \rho''\eta')/(\rho'^2 + \eta'^2)^{3/2}$, where $\eta = \|\mathbf{MB}^\lambda\|_2^2$, and $\rho = \chi_{\text{red}}^2(\mathbf{MB}^\lambda, \mathbf{d}_{10,15,22,34,51})$. We normalise η and ρ by their extrema values. The prime and double prime mean the first and second derivative with respect to λ , respectively. Figure 1 shows the plot of curvature κ in function of data misfit, superimposed to the *L* curve plot. The multi-band model maximising κ is shown with a black diamond marker, and is named $\mathbf{MB}^{\kappa_{\text{max}}}$. In Sect. 3.2, we will identify $\mathbf{MB}^{\kappa_{\text{max}}}$ as part of models dominated by noise. Hansen and O’Leary (1993) also suggest that the maximum curvature criterion may yield an under-damped solution if the *L* curve is characterised by a “smooth” corner, as it is in our case. Finally, there is a wide range of models with acceptable data misfit around the *L* curve bend. Picking one of them is to a large extent arbitrary. This motivates us to develop, for the MFT problem, a more objective approach to determine an “optimal range” of damped solutions. It will primarily consist in identifying models with “poor data exploitation” or “dominated by noise”, as summarised in Fig. 1 and explained below.

3.1 Identifying multi-band models with poor data exploitation

We aim at identifying over-damped models for which the structural information of data is poorly exploited. This range of models will be referred to as *poor data exploitation*. Within this range, decreasing the value of damping λ should lead to a significant decrease of data misfit while the model size, as measured by a given norm, should not increase by too much. Given our regularisation method, it seems natural to measure the size of model \mathbf{m}^λ with the ℓ^2 -norm:

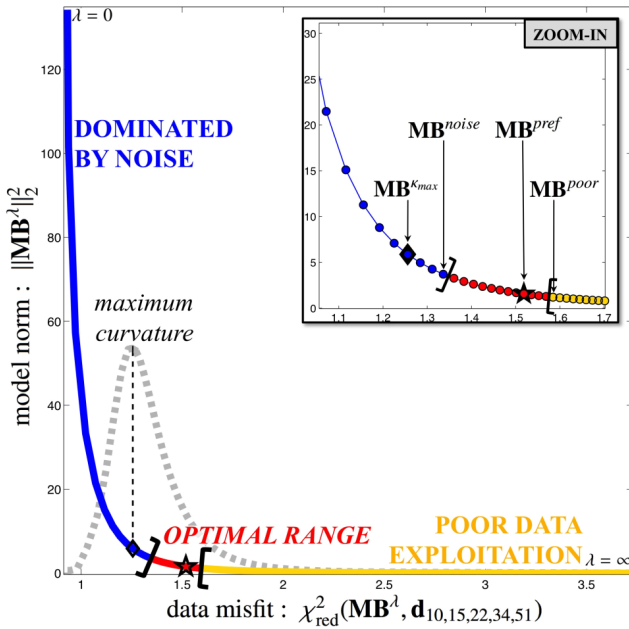


Fig. 1. Summary of our approach for objectifying the choice of regularisation parameter (damping). We plot the curve of trade-off (L curve) between the model norm, $\|\mathbf{MB}^\lambda\|_2$, and its residual data misfit, $\chi_{\text{red}}^2(\mathbf{MB}^\lambda, \mathbf{d}_{10,15,22,34,51})$, as damping λ varies. The L curve is displayed with solid-line parts in blue, red and light orange, corresponding to different model ranges. Its curvature κ is shown with a grey dashed line (with arbitrary amplitude scaling). The model $\mathbf{MB}^{k_{\text{max}}}$ (black diamond) corresponds to the maximum curvature of the L curve. The two extreme models $\mathbf{MB}^{\text{noise}}$ and $\mathbf{MB}^{\text{poor}}$ delimit the range of models “dominated by noise” or “with poor data exploitation”, respectively. In Sect. 3.3, we show that the selection of our preferred model $\mathbf{MB}^{\text{pref}}$ (black star), from the remaining “optimal range” of permitted models, can be qualified as quasi-objective.

$\|\mathbf{m}^\lambda\|_2 = (\sum_{j=1}^M |m_j^\lambda|^2)^{1/2}$. However, it corresponds to an L curve with a very smooth corner, as shown in Fig. 1. To render this smooth corner sharper, one may analyse a new L curve based on the model ℓ^∞ -norm: $\|\mathbf{m}^\lambda\|_\infty = \max(|\{m_j^\lambda\}_{1 \leq j \leq M}|)$. We further refer to this new trade-off curve as the L^∞ curve. It simply is an alternative evaluation of the size of regularised models plotted in function of data misfit. Figure 2 shows this curve of trade-off between $\|\mathbf{MB}^\lambda\|_\infty$ and $\chi_{\text{red}}^2(\mathbf{MB}^\lambda, \mathbf{d}_{10,15,22,34,51})$. As expected, the corner of the L^∞ curve has been sharpened by the ℓ^∞ -norm, but it is interesting to see that two regimes are now apparent on both sides of a breaking point. Let λ_{poor} be the damping value for which this change of regime occurs, and $\mathbf{MB}^{\text{poor}}$ be the corresponding multi-band model. Figure 2 shows that as the damping λ decreases down to the critical value λ_{poor} , the ℓ^∞ -norm (i.e. the largest model component) increases very gently. One sees that as soon as the damping gets smaller than this critical value, then at least one model component starts to behave differently: its amplitude is suddenly more

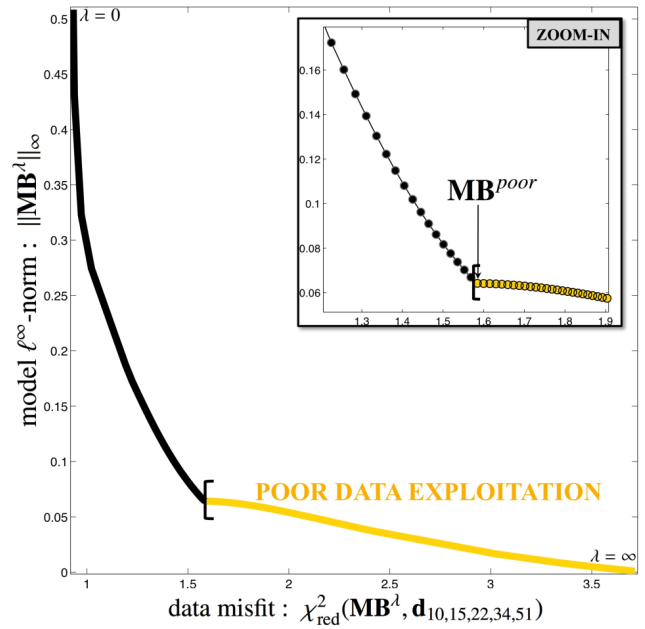


Fig. 2. Plot of the curve of trade-off between the model ℓ^∞ -norm, $\|\mathbf{MB}^\lambda\|_\infty$, and its residual data misfit, $\chi_{\text{red}}^2(\mathbf{MB}^\lambda, \mathbf{d}_{10,15,22,34,51})$. A sharp change of behaviour of the model ℓ^∞ -norm occurs as the damping λ gets lower than a threshold value λ_{poor} , which is interpreted as the signature of data noise starting to significantly pollute at least one model component. Models with damping $\lambda \geq \lambda_{\text{poor}}$ are identified as with “poor data exploitation”.

increased while the data misfit continues to decrease. Since the ℓ^∞ -norm is equal to the model parameter with largest amplitude, it easily detects the presence of at least one noisy model component with unrealistic high amplitude. In contrast, Fig. 1 shows a smooth increase of the ℓ^2 -norm of models slightly less damped than the critical model $\mathbf{MB}^{\text{poor}}$. This is due to the fact that the ℓ^2 -norm reflects integrated information over all model parameters, and is less sensitive to a single noisy parameter. Therefore, we interpret λ_{poor} as being the smallest damping value for which no model component is significantly contaminated by noise. The damping range corresponding to poor data exploitation is then defined by $\lambda \geq \lambda_{\text{poor}}$ (where poor stands for “poor data exploitation”). We have visually checked that tomographic models within this range of damping do not seem to be deteriorated by noise. Finally, as one aims at exploiting the most of our data set (e.g. structural dispersion due to wavefront healing), it may be rewarding to consider models with a better fit to data than for $\mathbf{MB}^{\text{poor}}$, which motivates us to identify models dominated by noise.

3.2 Identifying multi-band models dominated by noise

We aim at identifying a class of multi-band models referred to as *dominated by noise*. The key idea is to take advantage of the expected mutual coherency of delay-times measured

in different frequency bands (10, 15, 22, 34, and 51 s central periods). That is, for each ray path, the structural information should be coherent from one period to the other, but the noise information most likely not. In the following, using a form of cross-validation method, we show that the lack of coherency of the information in different frequency bands can be exploited to identify models dominated by noise.

3.2.1 Identifying single-band models dominated by noise

The first step is to identify a range of single-band models, $\mathbf{SB}_{34}^{\lambda'}$ (inversion of 34 s data subset, \mathbf{d}_{34} , with damping λ'), which are “dominated by noise”. Our approach consists in testing how well single-band models fit the 10 s data subset (\mathbf{d}_{10}), as damping varies. The philosophy of this approach is somewhat analogous to the cross-validation (CV) method, in that we aim at estimating the fit of a model ($\mathbf{SB}_{34}^{\lambda'}$) to a data set (\mathbf{d}_{10}) that is similar but not identical to the data (\mathbf{d}_{34}) that were used to derive the model. However, our data partitioning approach markedly differs from the random subdivision of data inherent in CV. Partition into frequency bands, based on similar sensitivities to mantle structure, seems to make more sense in the context of our multi-frequency data set.

Figure 3 shows the corresponding curve of trade-off between the model norm, $\|\mathbf{SB}_{34}^{\lambda'}\|_2^2$, and its 10 s data misfit, $\chi_{\text{red}}^2(\mathbf{SB}_{34}^{\lambda'}, \mathbf{d}_{10})$. This trade-off curve is not L shaped anymore, but is characterised by a remarkable “reversal”. Let λ'_{noise} be the damping value corresponding to this reversal, and $\mathbf{SB}_{34}^{\text{noise}}$ be the associated model. One sees that the more the damping λ' is decreased beyond the critical value λ'_{noise} , the more the single-band model gets complex (increase of $\|\mathbf{SB}_{34}^{\lambda'}\|_2^2$) while its 10 s data misfit gets poorer (increase of $\chi_{\text{red}}^2(\mathbf{SB}_{34}^{\lambda'}, \mathbf{d}_{10})$). We interpret as dominated by noise the range of single-band models, $\mathbf{SB}_{34}^{\lambda'}$, with damping $\lambda' \leq \lambda'_{\text{noise}}$ (where noise stands for “dominated by noise”). Indeed, for large values of damping λ' one expects smooth single-band models giving similar data predictions at all periods. As one decreases the damping λ' into the domain where we are fitting the noise more than the structure-imposed trend in the 34 s data, one expects that single-band models will fail to satisfy the 10 s data, because the 10 and 34 s data noise should not be coherent. Therefore, we consider $\mathbf{SB}_{34}^{\text{noise}}$ as the most damped single-band model among those obviously dominated by noise. As a consequence, the range of single-band models with damping “slightly larger” than λ'_{noise} must somehow be significantly polluted by noise. This may intuitively be understood when looking at the inset in Fig. 3. The more the damping decreases towards λ'_{noise} , the more the trade-off curve becomes vertical, i.e. decreasing the 10 s data misfit costs very much in terms of model norm increase. A priori, one ignores what is the most expensive cost that is affordable, so that one cannot objectively identify this range of models with significant noise contamination, i.e. before

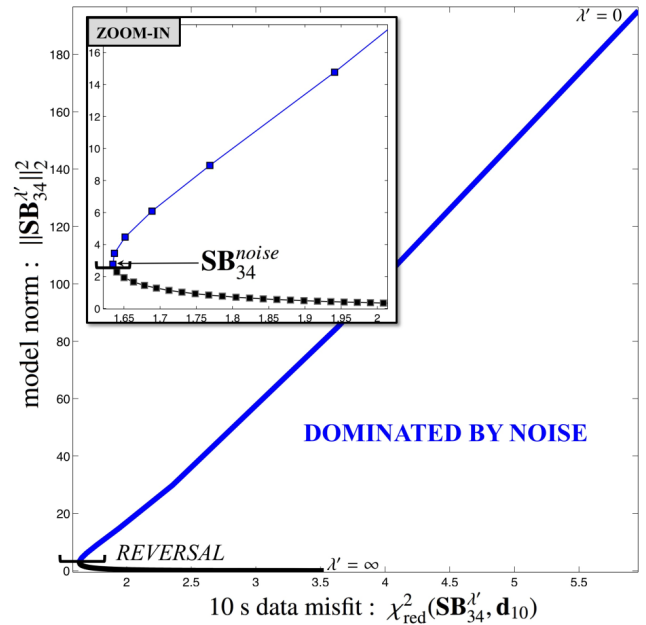


Fig. 3. Illustration of the coherency analysis. It consists in testing how well single-band models, $\mathbf{SB}_{34}^{\lambda'}$ (inversion of 34 s data), fit the 10 s data (\mathbf{d}_{10}) as damping λ' varies. The curve of trade-off between the single-band model norm, $\|\mathbf{SB}_{34}^{\lambda'}\|_2^2$, and its 10 s data misfit, $\chi_{\text{red}}^2(\mathbf{SB}_{34}^{\lambda'}, \mathbf{d}_{10})$, is characterised by a “reversal” which occurs for damping λ'_{noise} . Single-band models with damping $\lambda' \leq \lambda'_{\text{noise}}$ are identified as “dominated by noise”.

the noise predominant regime. However, its existence will be taken into account for the choice of the preferred model (Sect. 3.3).

3.2.2 Inferring multi-band models dominated by noise from single-band models

We aim at identifying multi-band models dominated by noise using our knowledge of the previously identified range of single-band models dominated by noise. The situation is illustrated in Fig. 4. We know that the single-band model $\mathbf{SB}_{34}^{\text{noise}}$, which results from an inversion of the 34 s data set only, is dominated by noise. Therefore, it means that one should not try to fit the 34 s data set (\mathbf{d}_{34}) better than does the model $\mathbf{SB}_{34}^{\text{noise}}$. This implies that an acceptable multi-band model, which results from an inversion of data at all periods including at 34 s, should not give a better fit to \mathbf{d}_{34} than $\mathbf{SB}_{34}^{\text{noise}}$. Let $\mathbf{MB}^{\text{noise}}$ be the multi-band model (with damping λ_{noise}) defined as having the “same 34 s data misfit” as the single-band model $\mathbf{SB}_{34}^{\text{noise}}$, that is $\chi_{\text{red}}^2(\mathbf{MB}^{\text{noise}}, \mathbf{d}_{34}) = \chi_{\text{red}}^2(\mathbf{SB}_{34}^{\text{noise}}, \mathbf{d}_{34})$. Thus, multi-band models \mathbf{MB}^{λ} with 34 s data misfit such as $\chi_{\text{red}}^2(\mathbf{MB}^{\lambda}, \mathbf{d}_{34}) \leq \chi_{\text{red}}^2(\mathbf{MB}^{\text{noise}}, \mathbf{d}_{34})$ are identified as dominated by noise.

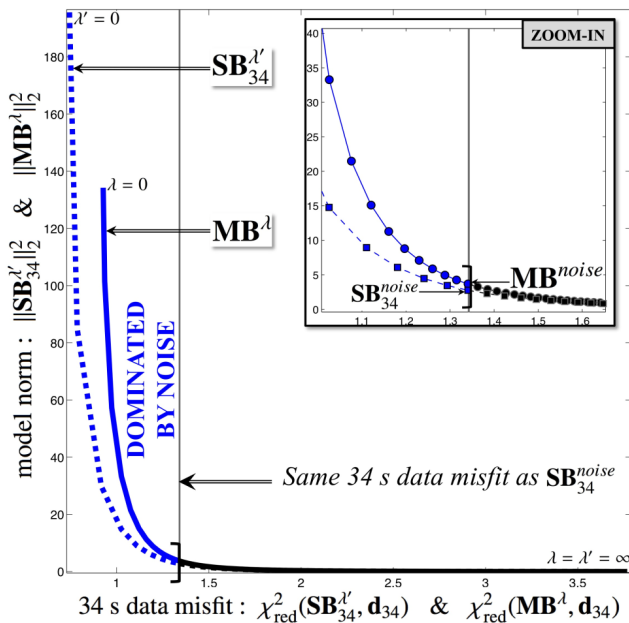


Fig. 4. Illustration showing how to infer multi-band models dominated by noise from single-band models. The curve of trade-off, between model norm and 34 s data misfit, is plotted for single- ($\mathbf{SB}_{34}^{\lambda'}$, dashed line) and multi-band (\mathbf{MB}^{λ} , solid line) models. An acceptable multi-band model should not give a better fit to the 34 s data (\mathbf{d}_{34}) than the single-band model $\mathbf{SB}_{34}^{\text{noise}}$. Thus, multi-band models with $\chi_{\text{red}}^2(\mathbf{MB}^{\lambda}, \mathbf{d}_{34}) \leq \chi_{\text{red}}^2(\mathbf{MB}^{\text{noise}}, \mathbf{d}_{34})$ are identified as “dominated by noise”. The multi-band model $\mathbf{MB}^{\text{noise}}$ is defined as having the same 34 s data misfit as $\mathbf{SB}_{34}^{\text{noise}}$.

For completeness, the coherency analysis should be applied to all combinations of single-band models and fit of data subsets. The relatively low number of 10, 15, and 51 s data (cf. Table 1) restricted us to analyse single-band models derived from either 22 or 34 s data. To do this, we imposed the same source–receiver geometry for multi- and single-band (22 or 34 s) models, so that differences between models cannot be due to this extraneous factor. We analysed single-band models, $\mathbf{SB}_{34}^{\lambda'}$ (or $\mathbf{SB}_{22}^{\lambda'}$), with their fit to the 10, 15, 22 (or 34), and 51 s data subsets, respectively. Corresponding L curve reversals were reached for different values of damping, λ' , which mainly reflects, at each period, the data noise level and the sensitivity kernel. We observed that the analysis of models derived from the 34 s data, and their fit to the 10 s data subset, lead to identify the narrowest range of acceptable multi-band models. Thus, for clarity, Sect. 3.2 has been focussed on this case.

3.3 Quasi-objective choice of preferred multi-band model

We have identified an *optimal range* (cf. Fig. 1) of multi-band models \mathbf{MB}^{λ} with damping $\lambda_{\text{noise}} < \lambda < \lambda_{\text{poor}}$. The next step

consists in choosing a “preferred” model $\mathbf{MB}^{\text{pref}}$ among those candidates, i.e. in between the two extreme models $\mathbf{MB}^{\text{noise}}$ and $\mathbf{MB}^{\text{poor}}$ as illustrated in Fig. 5a. Picking a particular solution within this range is a matter of compromise between exploiting most of the structural information (small damping) and minimising noise influence (large damping). One cannot avoid some subjectivity concerning this ultimate choice. Thus, one needs to estimate if this degree of subjectivity can still lead to significant model differences in terms of structural interpretation. We do this by looking at the level of correlation between the candidate solutions.

Figure 5b and d shows the correlation of models $\mathbf{MB}^{\lambda}|_{[960,1510]}$ from the optimal range with respect to the two extreme models $\mathbf{MB}^{\text{noise}}|_{[960,1510]}$ and $\mathbf{MB}^{\text{poor}}|_{[960,1510]}$, respectively, where $|_{[960,1510]}$ means that we only consider a vertical average of a model over depth range 960–1510 km. The correlation is displayed for spherical harmonic degrees $l = 1-80$. The characteristic horizontal length associated with degree $l = 80$, in this particular depth range, is about 400 km. We checked that all correlation results were similar if considering any other depth range through the mantle. From hereon, for ease of notation, we drop the $|_{[960,1510]}$ subscript on models. The correlation is computed as follows: let $S_A(l) = (\sum_{m=-l}^l A_l^m A_l^{m*})^{1/2}$ be the spectrum of a given square-integrable function A defined on the unit sphere, with A_l^m the spherical harmonic coefficients at degree l and azimuthal order m , and $*$ the complex conjugate. The correlation between two such functions A and B is defined as $\text{Corr}(A, B; l) = \sum_{m=-l}^l A_l^m B_l^{m*} / (S_A(l) S_B(l))$. Figure 5b and d shows that every model within the optimal range has a correlation coefficient larger than $\sim 91\%$, up to degree $l = 80$, with respect to the two extreme models. This indicates that this optimal range contains highly similar models, in terms of shear-velocity anomaly patterns. Thus, selecting any of them would lead to a similar structural interpretation. However, it seems to us wiser to choose the preferred model $\mathbf{MB}^{\text{pref}}$ as better correlated to $\mathbf{MB}^{\text{poor}}$ (not significantly influenced by noise) than to $\mathbf{MB}^{\text{noise}}$ (dominated by noise). To support this choice, note that models slightly more damped than $\mathbf{MB}^{\text{noise}}$, though not quantitatively identified, are certainly polluted by noise as suggested at the end of Sect. 3.2.1. Figure 5c shows that our preferred model $\mathbf{MB}^{\text{pref}}$ has a correlation coefficient larger than $\sim 98\%$, up to degree $l = 80$, with respect to almost every model within the optimal range (excepted those too close from $\mathbf{MB}^{\text{noise}}$). Obviously, such small model differences are not relevant, so that our choice of a preferred solution can be qualified as quasi-objective.

4 Looking at mid lower-mantle through multiple-frequency S-wave delay-times

We present here our preferred multi-band model, $\mathbf{MB}^{\text{pref}}$, at selected depths in mid lower-mantle. It consists in 3-D shear-wave velocity anomalies ($\delta \ln V_S$) resulting from the

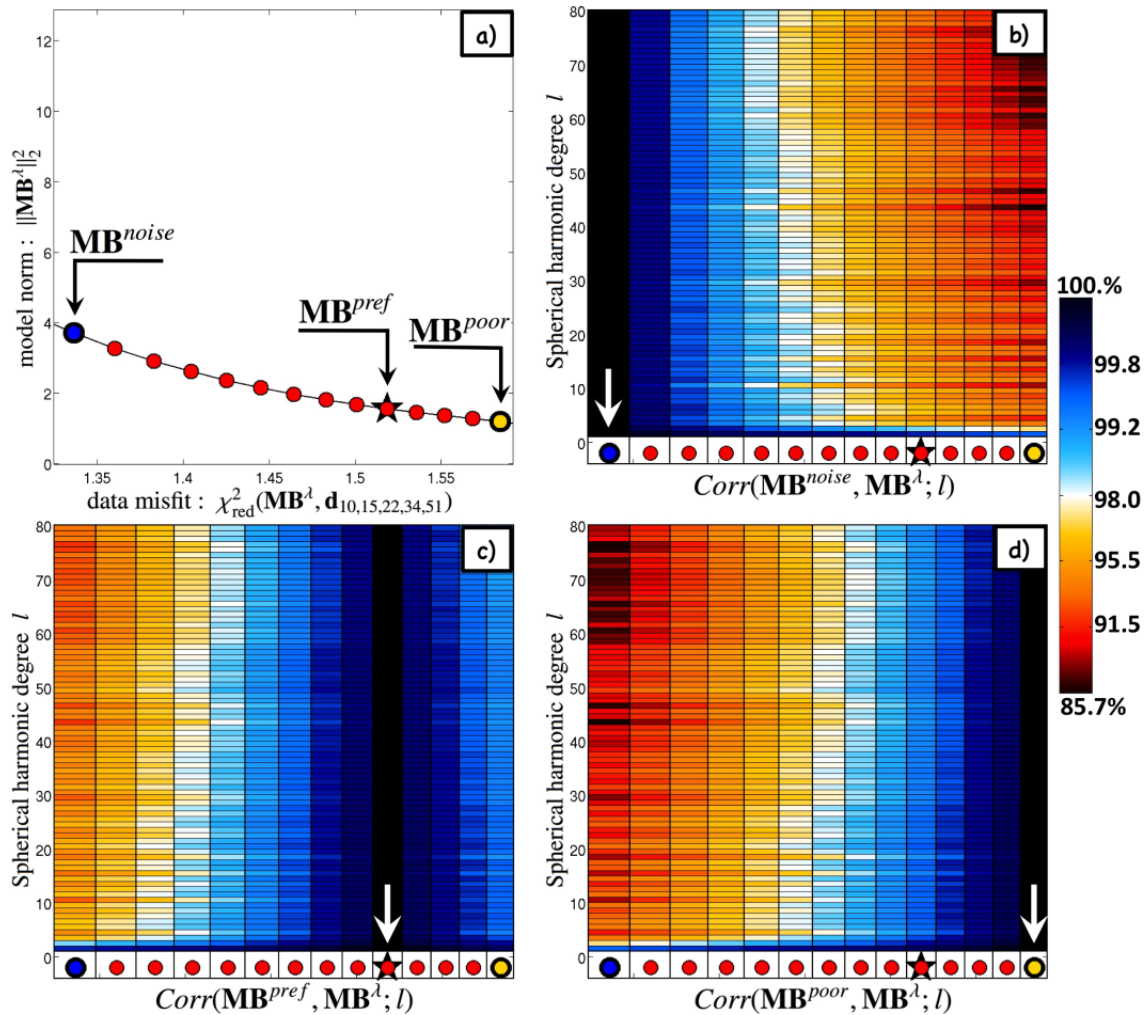


Fig. 5. Illustration showing that the selection of our preferred multi-band model $\mathbf{MB}^{\text{pref}}$ (black star) is quasi-objective in terms of model interpretation. **(a)** Trade-off curve corresponding to the “optimal range” of models, after identification of those “dominated by noise” ($\mathbf{MB}^{\text{noise}}$) or with “poor data exploitation” ($\mathbf{MB}^{\text{poor}}$). **(b–d)** Plot of the correlation of three specific models ($\mathbf{MB}^{\text{noise}}$, $\mathbf{MB}^{\text{pref}}$, $\mathbf{MB}^{\text{poor}}$, as indicated by white arrows) with several models \mathbf{MB}^λ spanning the “optimal range” of models (respectively). Correlation is shown for spherical harmonic degrees $l = 1–80$, and is displayed using the colour scale on the right (black is 100 %, and white is 98 %). One sees that $\mathbf{MB}^{\text{pref}}$ has a correlation coefficient greater than 98 %, up to degree $l = 80$, with respect to almost all models \mathbf{MB}^λ within the optimal range. The characteristic horizontal length associated with $l = 80$ is ~ 400 km. Here, the correlation is shown for models averaged over 960–1510 km depth, but we checked that it holds true at other depths.

simultaneous inversion of a global data set of 287 078 S and SS delay-times measured by cross-correlation in five frequency bands (10, 15, 22, 34, and 51 s central periods). Technical tomographic details are given in Sect. 2 and in Appendices A and B. The construction of such a model was motivated by the recent discovery of significant structural dispersion in this data set, after correction for crustal and 1-D attenuation effects. Taking into account this new observable global tomography holds the promise to better constrain the structure of Earth’s interior. Our current data coverage is too sparse below ~ 2000 km depth, owing to the drastic criteria applied by Zaroli et al. (2010) for cross-correlating fully isolated S and SS waveforms only. For instance, S waves

with epicentral distance $\Delta \geq 75^\circ$ were rejected to avoid interference between S and ScS . This was the price to pay for single-phase cross-correlation delay-time measurements; although BDT can handle multiple phases, to model closely spaced phases as upper-mantle triplications or S - ScS beyond 80° , good amplitudes are needed and full waveform techniques are easier to use than ray-based FF kernels. Besides, as the paraxial approximation is not valid near antipodal epicentral distances, we did not use SS delays with $\Delta \geq 140^\circ$ (Tian et al., 2007). Thus, we focus on the mid lower-mantle, where our current data coverage is the most adequate for high-resolution imaging, in the three depth ranges: 660–960, 960–1510, and 1510–1910 km. In the following, we discuss

major structural features in our model, and show its compatibility with three other recent global shear-wave velocity models. We postpone a detailed structural interpretation of our tomographic results, which is beyond the scope of this paper.

Figure 6 displays our model $\mathbf{MB}^{\text{pref}}$ between 660 and 1910 km depth. One can see long wavelength, high-velocity anomalies concentrated in the circum-Pacific and regions under Asia, as widely documented in the past (e.g. Li and Romanowicz, 1996; Masters et al., 1996; Grand et al., 1997; Van der Hilst et al., 1997; Masters et al., 2000; Ritsema and Heijst, 2000; Megnin and Romanowicz, 2000; Romanowicz, 2003; Montelli et al., 2004a, b, 2006; Houser et al., 2008; Ritsema et al., 2011). Geodynamicists now widely agree upon a link between these broad-scale anomalies and cold downwellings from ancient subductions driving an important part of the mantle circulation (e.g. Schubert et al., 2009). For instance, one sees the seismic signature of the ancient Farallon slab beneath North America (e.g. Grand et al., 1997), and of the remnants of Tethys beneath the Mediterranean/southern Eurasia (e.g. Van der Hilst and Karason, 1999; Houser et al., 2008). They are usually associated with cold material sinking into the lower mantle. Concerning the Farallon slab, it is worth noting that our model features at 960–1510 km depth a detached slab beneath the western quarter of North America (Fig. 6d), as recently discussed by Sigloch and Mihalynuk (2013). The presence of two large low-velocity regions at the core–mantle boundary, beneath Africa and the Pacific, is a common feature found in previously cited whole-mantle S -wave models. These two regions are often referred to as superplumes, and could be feeding up smaller low-velocity anomalies in the mid lower-mantle (e.g. Davaille, 1999; Davaille et al., 2005; Courtillot et al., 2003; Lay, 2005; Montelli et al., 2006; Schubert et al., 2009). Figure 6 shows that some of them are apparent in our model, most often beneath Africa and the Pacific and generally located near to a known hotspot (Tahiti, Samoa, Horn of Africa, Afar, South Africa, etc).

Figures 6 and 7 show that long-wavelength structures, which are very similar across recent tomographic models (e.g. Romanowicz, 2003), are well retrieved in this study. More precisely, Fig. 7 shows for visual comparison our model $\mathbf{MB}^{\text{pref}}$ with three other, latest generation, global shear-velocity models: S40RTS (Ritsema et al., 2011), PRI-S05 (Montelli et al., 2006), and TX2007 (Simmons et al., 2007). S40RTS results from Rayleigh wave phase velocity, teleseismic shear-wave travel time and normal mode splitting function measurements, and is parameterised by spherical harmonics up to degree 40 and by 21 vertical spline functions. PRI-S05 and $\mathbf{MB}^{\text{pref}}$ result from cross-correlation shear-wave delay times measured in one or several frequency band(s), respectively. They are irregularly parameterised according to data coverage. TX2007 results from a joint inversion of seismic data (teleseismic shear-wave travel times) and geodynamic constraints (global free-air gravity, tectonic

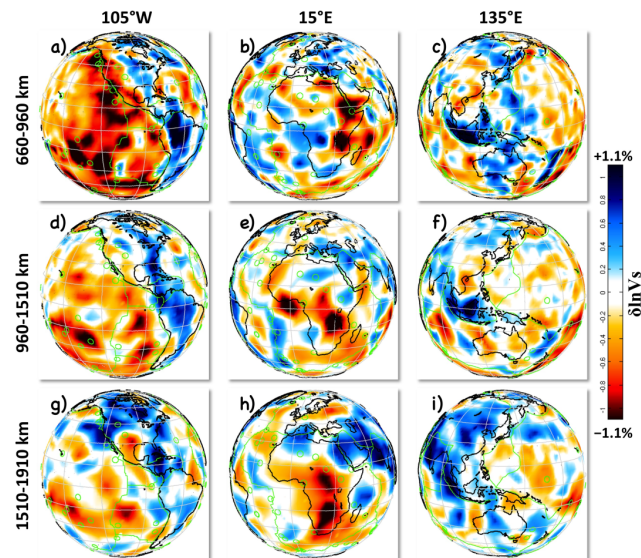


Fig. 6. Tomographic maps of shear-velocity anomalies of our model $\mathbf{MB}^{\text{pref}}$ at selected depths in mid lower-mantle: 660–960, 960–1510, and 1510–1910 km depth. Black lines: continents; green lines: tectonic plates; green circles: hotspots (Anderson and Schramm, 2005).

plate divergence, excess ellipticity of crust–mantle boundary, dynamic surface topography), and is parameterised with an irregular grid. S40RTS and TX2007 are based on RT, while PRI-S05 and $\mathbf{MB}^{\text{pref}}$ rely on BDT. Model differences are expected to come from (i) data sets (type of data, sources–receivers geometry, reference model, correction of attenuation and crust, etc), (ii) model parameterisation, (iii) model regularisation, and (iv) modelling of wave propagation. The four models exhibit an overall good agreement, in mid lower-mantle, at long wavelengths. As expected, model differences are more significant at short wavelengths. Figure 8 is an attempt to better characterise those differences. We show the spectra of the four models using spherical harmonics decomposition (Fig. 8a–c), and the correlation of our preferred model with respect to the three others (Fig. 8d–f). We find that the correlation is statistically significant up to a degree $l \sim 30$ in the mid lower-mantle, which corresponds to anomalies of ~ 1000 km horizontal length. Note that all spectra have very little energy for a degree l larger than 30. Our model is thus well compatible with other recent global tomographic studies. However, we also report some important differences at short and intermediate wavelengths. PRI-S05 and $\mathbf{MB}^{\text{pref}}$ tend to show slightly larger amplitudes, in mid lower-mantle, than S40RTS and TX2007. These model differences may partly be explained by the theoretical frameworks (e.g. BDT can compensate for wavefront-healing effect but RT cannot) or by the different nature of the data sets. However, a significant part of these differences likely comes from the used tomographic recipe for mapping the data into the model. We think an important ingredient is the

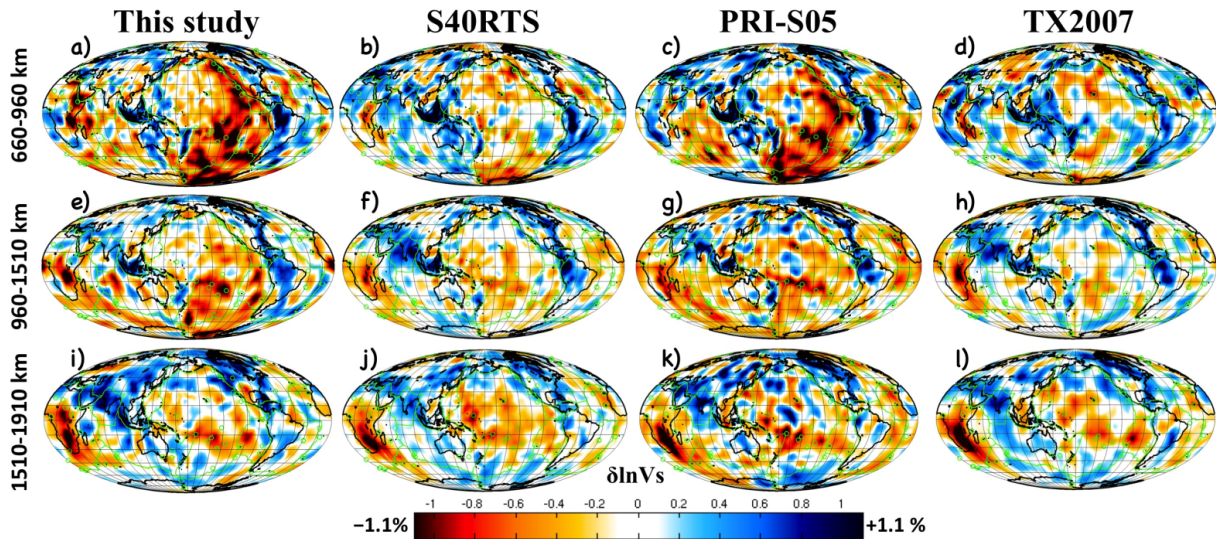


Fig. 7. Visual comparison of our model ($\mathbf{MB}^{\text{pref}}$, left column) with three other global shear-velocity anomaly tomographic models (S40RTS, PRI-S05, TX2007), displayed in mid lower-mantle with respect to IASP91 and after projection onto our model parameterisation.

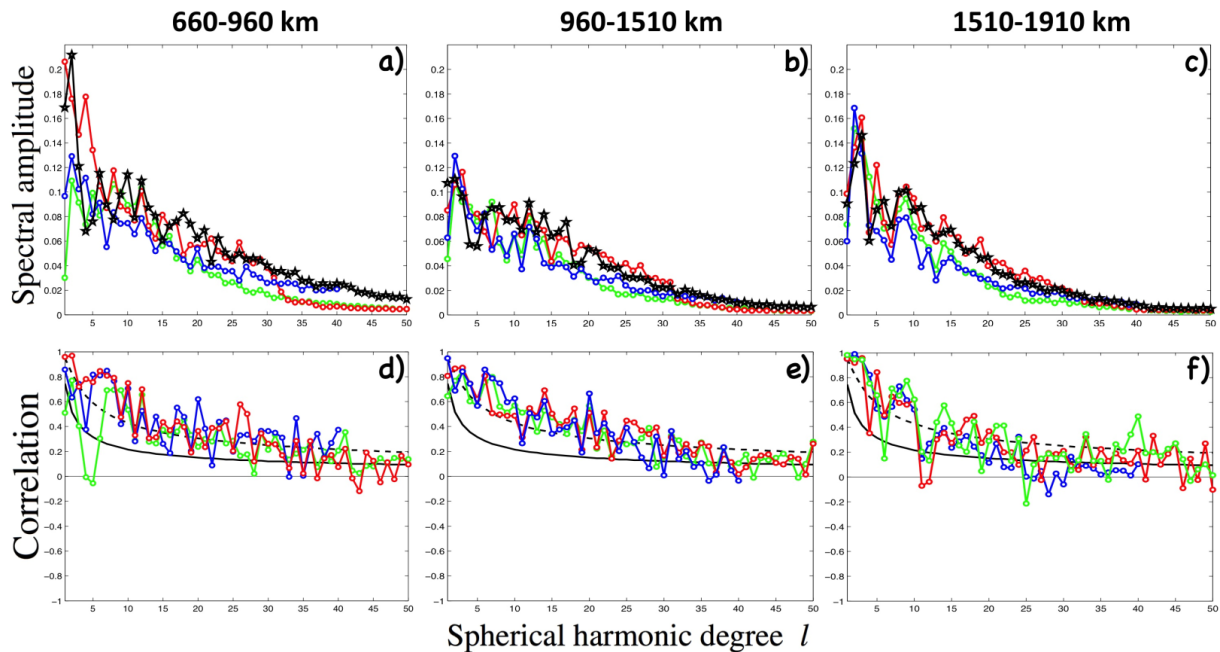


Fig. 8. (a–c) Spectral amplitude as a function of spherical harmonic degree $l = 1-50$ for four tomographic models in mid lower-mantle: our model $\mathbf{MB}^{\text{pref}}$ (black), S40RTS (blue, plotted up to $l = 40$), PRI-S05 (red), and TX2007 (green). (d–f) Correlation as a function of degree l between our model $\mathbf{MB}^{\text{pref}}$ and the three other models: S40RTS (blue), PRI-S05 (red), and TX2007 (green). The black dashed- and solid-lines indicate the 95 and 66 % significance levels, respectively.

subjective amount of regularisation used in the inversion. In the case of global MFT, we have shown how to strongly lessen this subjectivity, and hence counteract our poor knowledge of data uncertainties. A broader use of this philosophy could help to converge towards more objective seismic imaging of the mantle, which could strengthen the comparison of different tomographic models.

5 Conclusions

A recent focus in global tomography has been to exploit multiple-frequency body-wave delay-times (referred to as MFT) for better constraining the 3-D velocity structure of Earth's mantle. Solving such a linearised ill-posed inverse problem requires the determination of an optimal value of the

regularisation parameter (damping). Poor knowledge of data uncertainties often makes the selection of damping rather arbitrary, which leaves the choice of optimum model to the investigator. This issue is even more important for MFT, since to date there exists little understanding of the observational error sources that affect narrow-band cross-correlation estimates. To go beyond that subjectivity, an objective rationale for the choice of damping in the case of MFT has been presented and validated using a large data set of 287 078 *S*-wave delay-times measured in five frequency bands (10, 15, 22, 34, and 51 s central periods). First, we have shown how to take advantage of the expected coherency of delay-time estimates in different frequency bands. That is, whereas for each ray path the delay-time estimates should vary coherently from one period to the other, the noise most likely is not coherent. Thus, we have shown that the lack of coherency of the information in different frequency bands could be exploited, using an analogy with the cross-validation method, to identify models dominated by noise (under-damped). In addition, we have shown that analysing the behaviour of the model ℓ^∞ -norm, as damping varies, could give us access to the signature of data noise starting to significantly pollute at least one model component, and thus could help us to identify models constructed with poor data exploitation (over-damped). We have shown that the selection of a preferred model from the remaining range of permitted solutions, i.e. models neither dominated by noise nor with poor data exploitation, could be qualified as quasi-objective. Indeed, the selected model shows a high degree of similarity with almost all other permitted models (correlation superior to 98 % up to spherical harmonic degree 80). For completeness, the resulting tomographic model has been presented at selected depths in mid lower-mantle (660–1910 km depth), and shown to be compatible with three other recent global shear-velocity models. In conclusion, we believe that the presented rationale, for objectifying the choice of damping, could benefit various styles of the inverse problem. For instance, the coherency analysis could easily be applied to other data types, provided that data could be subdivided into subsets with similar sensitivity to model parameters – as we did when subdividing our multi-band data set into single-band subsets. A wider application of this rationale should help to converge towards more objective seismic imaging of Earth's mantle, and thus to better understand its dynamics.

Appendix A

Multi-resolution model parameterisation based on spherical triangular prisms

We assume that the whole-mantle shear-wave model is described by a finite number of parameters $(m_j)_{1 \leq j \leq M}$. Thus, the continuous model is given by $m(\mathbf{r}) = \sum_{j=1}^M m_j b_j(\mathbf{r})$, where the basis functions, b_j , represent the model parameterisation. A major constraint in global seismic tomography is the irregular illumination of the mantle, due to the non-uniform geographical distribution of earthquakes and receivers. In most tomographic studies (e.g. Houser et al., 2008; Ritsema et al., 1999, 2011), the spatial variability in the data's resolving power is largely ignored by using uniform basis functions (e.g. cubic cells or spherical harmonics). An alternative is to use a multi-resolution model grid that is adapted to the spatially varying constraints of the data (e.g. Michelini, 1995; Spakman and Bijwaard, 2001; Montelli et al., 2006; Simmons et al., 2009). A brief review on irregular parameterisation can be found in Sambridge and Rawlinson (2005) and Rawlinson et al. (2010). In this study, we have built a model parameterisation that attempts to be adapted to the scale of the smallest structure one may expect to resolve with the data. The whole mantle is divided into 18 spherical layers, whose thickness varies between 100 and 200 km (Fig. A1a). Each layer is divided into a set of spherical triangular prisms, whose spatial distribution is laterally irregular. The vertices (nodes) of the spherical triangle at the top of each prism represent the model parameters (red dots in Fig. A1c). We follow the strategy of Nolet and Montelli (2005) to find an optimal node distribution in each layer, based on the ray density, that attempts to maximise the extraction of structural information from the data. The major difference of our approach is that we do not use tetrahedra to fill in the mantle. We prefer to use spherical triangular prisms (Fig. A1c), spanning constant-depth spherical layers, because this makes the refinement of the model grid simpler. The total number of nodes we use to parameterise the mantle is $M = 38\,125$. Node spacing ranges from about 200 to 800 km. The current coverage of seismic stations generally allows us to have closely spaced nodes in the Northern Hemisphere; node spacing is coarser in the Southern Hemisphere. Figure A1b shows an example of the node distribution obtained in one layer (530–660 km depth). Note how the model grid has been refined beneath seismically active regions where the ray density is relatively high. The entire set of nodes (latitude, longitude, depth) of our global parameterisation is available upon request from the first author.

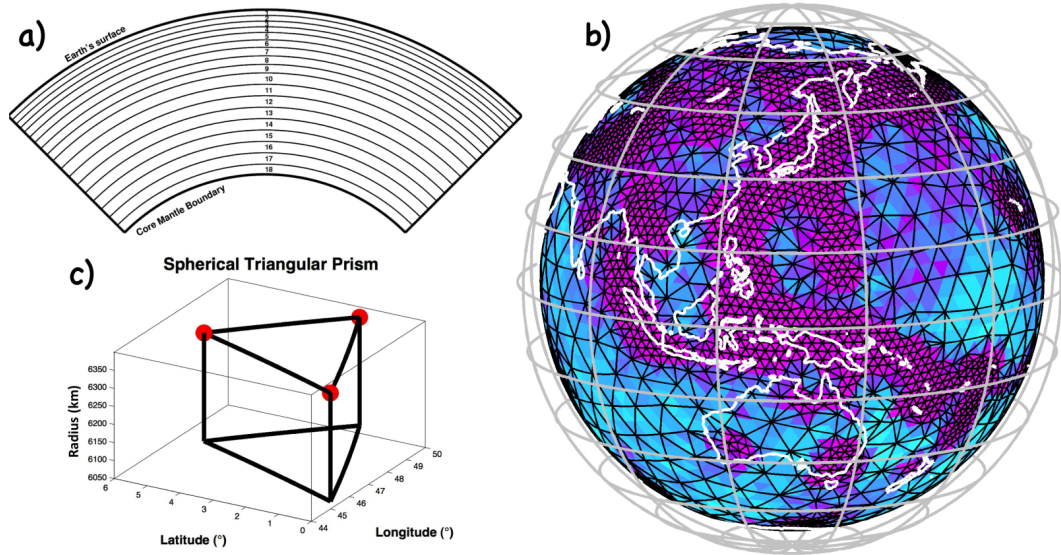


Fig. A1. Illustration of the model parameterisation. **(a)** 18 constant-depth spherical layers used for radially parameterising the mantle. **(b)** Optimised node distribution (black triangles mesh) obtained in one layer (530–660 km depth). The nodes agency attempts to fit a “resolution length” function, ranging in this case from 210 (magenta) to 850 km (cyan), that is driven by data coverage. **(c)** Representation of a spherical triangular prism, where the red dots represent three nodes of the model grid.

Appendix B

Analytical expressions for finite-frequency ray-based sensitivity kernels

We aim at deriving analytical expressions for single-phase (K_S , K_{SS}) and two-phase interference (K_{S+sS} , K_{SS+sSS}) kernels, in order to decrease the computational cost of banana-doughnut kernels (Dahlen et al., 2000). This is valuable as we aim at computing hundreds of thousands of kernels on a very fine grid (regular cells with edges of 20 km); and it may be useful for double-checking the validity of numerical calculations. Examples of such analytical kernels are displayed in Fig. B1. The finite-frequency delay-time sensitivity kernel of a single-phase, with respect to velocity perturbation ($\delta c/c$), is (Dahlen et al., 2000)

$$\begin{cases} K(\mathbf{r}_x) = -\frac{1}{2\pi c(\mathbf{r}_x)} \frac{R_{rs}}{c_r R_{xr} R_{xs}} \frac{N(\Delta\Phi)}{D} \\ N(\Delta\Phi) = \int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin[\omega\Delta T - \Delta\Phi] d\omega \\ D = \int_0^\infty \omega^2 |\dot{m}(\omega)|^2 d\omega, \end{cases} \quad (\text{B1})$$

where \mathbf{r}_x is the spatial position of scatterer x , $\Delta\Phi$ is the phase shift due to passage through caustics or super critical reflection, R_{rs} , R_{xr} , R_{xs} are the geometrical spreading factors, $|\dot{m}(\omega)|^2$ is the source power spectrum, ΔT is the detour time of the scattered wave, c_r and $c(\mathbf{r}_x)$ are the velocities at receiver and scatterer position, respectively. Earthquake catalogues contain a large majority of shallow earthquakes, which implies to take into account the interference of a di-

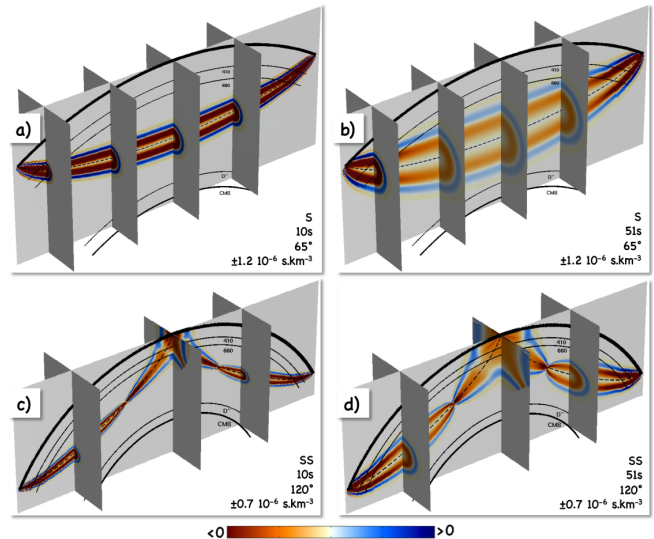


Fig. B1. Three-dimensional view of “analytical” ray-based finite-frequency delay-time sensitivity kernels. Each subfigure provides the phase name (S or SS), the dominant wave period (10 or 51 s), the epicentral distance (65° for S , 120° for SS), and the colour scale bounds (s km^{-3}). The source is at the surface, and the black dashed line represents the geometrical ray path.

rect phase (e.g. S , SS) with its associated depth phase (e.g. sS , sSS) within the cross-correlation time window. The kernel associated with two-phase interference is (Dahlen et al., 2000)

$$\left\{ \begin{aligned} K_{1+2}(\mathbf{r}_x) &= -\frac{1}{2\pi c(\mathbf{r}_x)} \frac{N_{1+2}(\Delta\Phi)}{D_{1+2}} \\ N_{1+2}(\Delta\Phi) &= \int_0^\infty \omega^3 |\dot{m}(\omega)|^2 [\dots]_{1+2} d\omega \\ [\dots]_{1+2} &= [\dots]_{11} + [\dots]_{12} + [\dots]_{21} + [\dots]_{22} \\ [\dots]_{11} &= \sqrt{|\det(\mathbf{H}_1)|} \sin(\omega\Delta T_1 - \Delta\Phi) \\ [\dots]_{12} &= \sqrt{|\det(\mathbf{H}_1)|} \sin(\omega(\Delta T_1 + t_1 - t_2) - \Delta\Phi) \\ [\dots]_{21} &= \sqrt{|\det(\mathbf{H}_2)|} \sin(\omega(\Delta T_2 + t_2 - t_1) - \Delta\Phi) \\ [\dots]_{22} &= \sqrt{|\det(\mathbf{H}_2)|} \sin(\omega\Delta T_2 - \Delta\Phi) \\ D_{1+2} &= \int_0^\infty 2(1 + \cos(\omega(t_1 - t_2)))\omega^2 |\dot{m}(\omega)|^2 d\omega, \end{aligned} \right. \quad (\text{B2})$$

where “1” and “2” stand for the direct and depth phases, and where t_i , ΔT_i , and \mathbf{H}_i are the predicted arrival times, detour times, and Hessian matrices of the direct ($i = 1$) and depth ($i = 2$) phases, respectively. Provided the filtered source time function $m(t)$ has a known analytical expression, one sees that it is possible to calculate analytical expressions for both single- and two-phase interference kernels. We only need to calculate the ratios of integrals over ω in Eqs. (B1) and (B2); the remaining terms can easily be computed using the software by Tian et al. (2007), available at: <https://www.geoazur.net/GLOBALSEIS/Soft.html>. Thus, following Hung et al. (2001), we assume a Gaussian filtered source time function whose spectrum is given by $|\dot{m}(\omega)|^2 = \frac{\omega^2 T^2}{2\pi} e^{-\left(\frac{\omega T}{2\pi}\right)^2}$, with T the dominant wave period. This assumption is appropriate for S -waves within the 10–51 s period range. Unless the wave is super-critically reflected, with an angle-dependent phase shift, $\Delta\Phi$ takes three possible values: 0, $-\pi/2$, and $-\pi$. For S waves, $\Delta\Phi$ is always equal to 0, while for SS waves $\Delta\Phi = -\pi/2$ in between the two caustics and $\Delta\Phi = 0$ elsewhere. Equations (B1) and (B2) neglect the effects of transmission coefficients as well as directional effects in the scattering, which are small as long as the sensitivity kernel is narrow. Thus, the maximum detour time ΔT to which we compute the kernel is equal to T (wave dominant period), which neglects third and higher order Fresnel zones. The resulting analytical expressions we obtain for single-phase kernels (K_S , K_{SS}) are

$$\left\{ \begin{aligned} N(\Delta\Phi=0) &= 16\pi^{10} \sqrt{\pi} \frac{\Delta T}{T^9} e^{-\left(\frac{\Delta T}{T}\right)^2} \left(\Delta T^4 - 5\left(\frac{T\Delta T}{\pi}\right)^2 + \frac{15}{4}\left(\frac{T}{\pi}\right)^4 \right) \\ N(\Delta\Phi=-\frac{\pi}{2}) &= \frac{T^2}{2\pi} \left(2^5 \left(\frac{\pi}{T}\right)^{10} \left(\Delta T^4 - \frac{9}{2}\left(\frac{T\Delta T}{\pi}\right)^2 + 2\left(\frac{T}{\pi}\right)^4 \right) \right. \\ &\quad \left. - 2^5 \sqrt{\pi} \left(\frac{\pi}{T}\right)^{11} \Delta T e^{-\left(\frac{\pi\Delta T}{T}\right)^2} \left(\Delta T^4 - 5\left(\frac{T\Delta T}{\pi}\right)^2 + \frac{15}{4}\left(\frac{T}{\pi}\right)^4 \right) \text{erfi}\left(\frac{\pi\Delta T}{T}\right) \right) \\ D &= 6\pi^{\frac{3}{2}} \left(\frac{\pi}{T}\right)^3. \end{aligned} \right. \quad (\text{B3})$$

For two-phase interference kernels (K_{S+sS} , K_{SS+sSS}), we find

$$\left\{ \begin{aligned} N_{1+2}(\Delta\Phi=0) &= \sqrt{|\det(\mathbf{H}_1)|} (I(\Delta T_1) + I(\Delta T_1 + t_1 - t_2)) \\ &\quad + \sqrt{|\det(\mathbf{H}_2)|} (I(\Delta T_2) + I(\Delta T_2 + t_2 - t_1)) \\ I : x &\rightarrow e^{-\left(\frac{x\pi}{T}\right)^2} H_5\left(\frac{x\pi}{T}\right) \sqrt{\pi} \left(\frac{\pi}{T}\right)^6 \\ H_5 : x &\rightarrow 32x^5 - 160x^3 + 120x \quad (\text{Hermite polynomial of order 5}) \\ N_{1+2}(\Delta\Phi=-\frac{\pi}{2}) &= \sqrt{|\det(\mathbf{H}_1)|} (J(\Delta T_1) + J(\Delta T_1 + t_1 - t_2)) \\ &\quad + \sqrt{|\det(\mathbf{H}_2)|} (J(\Delta T_2) + J(\Delta T_2 + t_2 - t_1)) \\ J : x &\rightarrow 2^5 \left(\frac{\pi}{T}\right)^{10} \left(x^4 - \frac{9}{2}\left(\frac{xT}{\pi}\right)^2 + 2\left(\frac{T}{\pi}\right)^4 \right) \\ &\quad - 2^5 \sqrt{\pi} \left(\frac{\pi}{T}\right)^{11} x e^{-\left(\frac{x\pi}{T}\right)^2} \left(x^4 - 5\left(\frac{xT}{\pi}\right)^2 + \frac{15}{4}\left(\frac{T}{\pi}\right)^4 \right) \text{erfi}\left(\frac{x\pi}{T}\right) \\ D_{1+2} &= 12\pi^{\frac{3}{2}} \left(\frac{\pi}{T}\right)^3 + 2\gamma(t_1 - t_2) \\ \gamma : x &\rightarrow 8\pi^{\frac{3}{2}} \left(\frac{\pi}{T}\right)^7 e^{-\left(\frac{x\pi}{T}\right)^2} \left(x^4 - 12\left(\frac{xT}{2\pi}\right)^2 + 12\left(\frac{T}{2\pi}\right)^4 \right). \end{aligned} \right. \quad (\text{B4})$$

We use the derived Maclaurin series expansion for the evaluation of the imaginary error function, which is

$$\text{erfi}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{t^2} dt = \frac{2}{\sqrt{\pi}} \sum_{n=1}^{\infty} \frac{z^{2n-1}}{(2n-1)((n-1)!)} \quad (\text{B5})$$

The numerical approach of Simpson et al. (2003) is used to estimate the number n of terms needed for the series to converge. We find that $n = 10$ is appropriate in our case. For completeness, note that for compound rays (sS , SS , sSS), one scatterer can, sometimes, be associated to more than one perpendicular projection point on the geometrical ray path (e.g. Hung et al., 2000; Tian et al., 2007; Nolet, 2008). Thus, for scatterers near a discontinuous interface, such as the surface, we made sure to take into account incoming rays that hit the scatterer directly as well as those that first visit a boundary.

Acknowledgements. The authors thank the Iris and Geoscope data centres for providing seismological data, the young researcher ANR TOMOGLOB no. ANR-06-JCJC-0060 and the SEISGLOB ANR 2011 Blanc SIMI 5-6-016-01, the ERC (Advanced grant 226837) and a Marie Curie Re-integration grant (project 223799). Discussions with A. Maggi and B. S. A. Schuberth have been stimulating at various stages of this research. The authors thank M. Obayashi and K. Yoshizawa for helpful reviews that improved the original paper.

Edited by: K. Liu

References

- Anderson, D. L. and Schramm, K. A.: Global hotspot maps, in: Plates, Plumes and Paradigm, edited by: Presnall, D. and Anderson, D., G. S. of America, 2005.
- Aster, R. C., Borchers, B., and Thurber, C.: Parameter Estimation and Inverse Problems, Elsevier, 2012.
- Bassin, C., Laske, G., and Masters, G., The Current Limits of Resolution for Surface Wave Tomography in North America, EOS Trans AGU, 81, F897, 2000.
- Bolton, H. and Masters, G.: 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, 13527–13540, 2001.

- Boschi, L., Becker, T. W., Soldati, G., and Dziewonski, A. M.: On the relevance of Born theory in global seismic tomography, *Geophys. Res. Lett.*, 33, L06302, doi:10.1029/2005GL025063, 2006.
- Chapman, C.: A new method for computing synthetic seismograms, *Geophys. J. Roy. Astr. Soc.*, 54, 481–518, 1978.
- Courtillot, V., Davaille, A., Besse, J., and Stock, J.: Three distinct types of hotspots in the Earth's mantle, *Earth Planet. Sc. Lett.*, 205, 295–308, 2003.
- Dahlen, F. A., Hung, S.-H., and Nolet, G.: Fréchet kernels for finite-frequency traveltimes – 1. Theory, *Geophys. J. Int.*, 141, 157–174, 2000.
- Davaille, A.: Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle, *Nature*, 402, 756–760, 1999.
- Davaille, A., Stutzmann, E., Silveira, G., Besse, J., and Courtillot, V.: Convective patterns under the Indo-Atlantic box, *Earth Planet. Sc. Lett.*, 239, 233–252, 2005.
- Debayle, E., Kennett, B. L. N., and Priestley, K.: Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia, *Nature*, 433, 509–512, 2005.
- Dziewonski, A. M. and Anderson, D.: Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, 25, 297–356, 1981.
- Fichtner, A., Kennett, B. L. N., and Igel, H.: Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, 179, 1703–1725, 2009.
- Fukao, Y., Widiyantoro, S., and Obayashi, M.: Stagnant slabs in the upper and lower mantle transition region, *Rev. Geophys.*, 39, 291–323, 2001.
- Grand, S. P., Van der Hilst, R. D., and Widiyantoro, S.: Global seismic tomography: a snapshot of convection in the Earth, *GSA Today*, 7, 1–7, 1997.
- Gudmundson, O.: On the effect of diffraction on traveltime measurement, *Geophys. J. Int.*, 124, 304–314, 1997.
- Gudmundsson, O., Davies, J. H., and Clayton, R. W.: Stochastic analysis of global travel time data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, 102, 25–44, 1990.
- Hansen, C. and O'Leary, D.: The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.*, 14, 1487–1503, 1993.
- Houser, C., Masters, G., Shearer, P. M., and Laske, G.: Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.*, 174, 195–212, 2008.
- Hung, S.-H., Dahlen, F. A., and Nolet, G.: Fréchet kernels for finite-frequency traveltimes – 2. Examples, *Geophys. J. Int.*, 141, 175–203, 2000.
- Hung, S.-H., Dahlen, F. A., and Nolet, G.: Wavefront-healing: a banana-doughnut perspective, *Geophys. J. Int.*, 146, 289–312, 2001.
- Hung, S.-H., Shen, Y., and Chiao, L.-Y.: Imaging seismic velocity structure beneath the Iceland hotspot: a finite-frequency approach, *J. Geophys. Res.*, 109, B08305, doi:10.1029/2003JB002889, 2004.
- Kennett, B. and Engdahl, E.: Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, 105, 429–465, 1991.
- Komatitsch, D., Ritsema, J., and Tromp, J.: The spectral-element method, Beowulf computing and global seismology, *Science*, 298, 1737–1742, 2002.
- Lay, T.: The deep mantle thermo-chemical boundary layer: the putative mantle plume source, Geological Society of America, Special Paper 388, 2005.
- Lekic, V. and Romanowicz, B.: Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.*, 185, 799–831, 2011.
- Li, X. D. and Romanowicz, B.: Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. Geophys. Res.*, 101, 22245–22272, 1996.
- Marquering, H., Nolet, G., and Dahlen, F. A.: Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.*, 132, 521–534, 1998.
- Masters, G., Johnson, S., Laske, G., Bolton, H., and Davies, J. H.: A shear-velocity model of the mantle, *Philos. T. Roy. Soc. A*, 354, 1385–1411, 1996.
- Masters, G., Laske, G., Bolton, H., and Dziewonski, A. M.: The relative behaviour of shear velocity, bulk sound speed and compressional velocity in the mantle: implications for chemical and thermal structure, in: *Earth's Deep Interior*, edited by: Karato, S., Forte, A., Liebermann, R. C., Masters, G., and Stixrude, L., AGU, 63–88, 2000.
- Megnin, C. and Romanowicz, B.: The 3-D velocity structure of the mantle from the inversion of body, surface, and higher mode waveforms, *Geophys. J. Int.*, 143, 709–728, 2000.
- Mercerat, E. D. and Nolet, G.: Comparison of ray- and adjoint-based sensitivity kernels for body-wave seismic tomography, *Geophys. Res. Lett.*, 39, doi:10.1029/2012GL052002, 2012.
- Micheline, A.: An adaptive-grid formalism for traveltime tomography, *Geophys. J. Int.*, 121, 489–510, 1995.
- Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R., and Hung, S.-H.: Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, 303, 338–343, 2004a.
- Montelli, R., Nolet, G., Masters, G., Dahlen, F. A., and Hung, S.-H.: Global P and PP traveltime tomography: rays versus waves, *Geophys. J. Int.*, 158, 636–654, 2004b.
- Montelli, R., Nolet, G., Dahlen, F. A., and Masters, G.: A catalogue of deep mantle plumes, new results from finite-frequency tomography, *Geochem. Geophys. Geosy.*, 7, Q11007, doi:10.1029/2006GC001248, 2006.
- Morelli, A. and Dziewonski, A. M.: Topography of the core-mantle boundary and lateral homogeneity of the liquid core, *Nature*, 325, 678–683, 1987.
- Nissen-Meyer, T., Dahlen, F., and Fournier, A.: Spherical-earth Fréchet sensitivity kernels, *Geophys. J. Int.*, 168, 1051–1066, 2007.
- Nolet, G.: *A Breviary of Seismic Tomography*, Cambridge University Press, Cambridge, UK, 2008.
- Nolet, G.: Slabs do not go gently, *Science*, 324, 1152–1153, 2009.
- Nolet, G. and Montelli, R.: Optimum parameterization of tomographic models, *Geophys. J. Int.*, 161, 365–372, 2005.
- Paige, C. C. and Saunders, M.: LSQR: An algorithm for sparse, linear equations and sparse least squares, *ACM T. Math. Softw.*, 8, 43–71, 1982.
- Rawlinson, N., Pozgay, S., and Fishwick, S.: Seismic tomography: a window into deep Earth, *Phys. Earth Planet. Inter.*, 178, 101–135, 2010.
- Ritsema, J. and Heijst, H. J. V.: Seismic imaging of structural heterogeneity in Earth's mantle: evidence for large-scale mantle flow, *Sci. Prog.*, 83, 243–259, 2000.

- Ritsema, J., van Heijst, H.-J., and Woodhouse, J. H.: Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, 286, 1925–1928, 1999.
- Ritsema, J., Deuss, A., van Heijst, H.-J., and Woodhouse, J. H.: S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltimes and normal-mode splitting function measurements, *Geophys. J. Int.*, 184, 1223–1236, 2011.
- Romanowicz, B.: Global mantle tomography: progress status in the past 10 years, *Annu. Rev. Earth Planet Sci.*, 31, 303–328, 2003.
- Sambridge, M. and Rawlinson, N.: Seismic tomography with irregular meshes, in: *Seismic Earth: Array Analysis of Broadband Seismograms*, edited by: Levander, A. and Nolet, G., AGU, 157, 49–65, 2005.
- Savage, B., Komatitsch, D., and Tromp, J.: Effects of 3-D attenuation on seismic wave amplitude and phase measurements, *Bull. Seism. Soc. Am.*, 100, 1241–1251, 2010.
- Schuberth, B. S. A., Bunge, H.-P., and Ritsema, J.: Tomographic filtering of high-resolution mantle circulation models: can seismic heterogeneity be explained by temperature alone?, *Geochem. Geophys. Geosy.*, 10, Q05W03, doi:10.1029/2009GC002401, 2009.
- Sieminski, A., L  v  que, J.-J., and Debayle, E.: Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.*, 31, L24614, doi:10.1029/2004GL021402, 2004.
- Sigloch, K.: Mantle provinces under North America from multi-frequency P wave tomography, *Geochem. Geophys. Geosy.*, 12, Q02W08, doi:10.1029/2010GC003421, 2011.
- Sigloch, K. and Mihalynuk, M.: Intra-oceanic subduction shaped the assembly of Cordilleran North America, *Nature*, 496, 50–56, doi:10.1038/nature12019, 2013.
- Sigloch, K. and Nolet, G.: Measuring finite-frequency body-wave amplitudes and traveltimes, *Geophys. J. Int.*, 167, 271–287, 2006.
- Sigloch, K., McQuarrie, M., and Nolet, G.: Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.*, 1, 458–462, 2008.
- Simmons, N. A., Forte, A. M., and Grand, S. P.: Thermochemical structure and dynamics of the African superplume, *Geochem. Geophys. Geosy.*, 12, L02301, doi:10.1029/2006GL028009, 2007.
- Simmons, N. A., Myers, S. C., and Ramirez, A.: Multi-resolution seismic tomography based on recursive tessellation hierarchy, in: *Proceedings of the 2009 Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, 1, 211–220, 2009.
- Simpson, M. J., Clement, T. P., and Yeomans, F. E.: Analytical model for computing residence times near a pumping well, *Ground Water*, 41, 351–354, 2003.
- Spakman, W. and Bijwaard, H.: Optimization of cell parameterizations for tomographic inverse problems, *Pure Appl. Geophys.*, 158, 1401–1423, 2001.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J.: Seismic tomography of the Southern California crust based upon spectral-element and adjoint methods, *Geophys. J. Int.*, 180, 433–462, 2010.
- Tarantola, A.: Inversion of traveltimes and seismic waveforms, in: *Seismic Tomography*, edited by: Nolet G., Reidel, Dordrecht, 135–157, 1987.
- Tarantola, A. and Nercessian, A.: Three-dimensional inversion without blocks, *Geophys. J. Roy. Astr. Soc.*, 76, 299–306, 1984.
- Tian, Y., Montelli, R., Nolet, G., and Dahlen, F. A.: Computing traveltimes and amplitude sensitivity kernels in finite-frequency tomography, *J. Comp. Phys.*, 226, 2271–2288, 2007.
- Tian, Y., Sigloch, K., and Nolet, G.: Multiple-frequency SH-tomography of the western US upper mantle, *Geophys. J. Int.*, 178, 1384–1402, 2009.
- Tian, Y., Zhou, Y., Sigloch, K., Nolet, G., and Laske, G.: Structure of North American mantle constrained by simultaneous inversion of multiple-frequency SH, SS, and Love waves, *J. Geophys. Res.*, 116, B02307, doi:10.1029/2010JB007704, 2011.
- Tikhonov, A.: Solution of incorrectly formulated problems and the regularization method, *Dokl. Akad. Nauk. SSSR*, 151, 501–504, 1963.
- Trampert, J. and Spetzler, J.: Surface wave tomography: finite frequency effects lost in the null space, *Geophys. J. Int.*, 164, 394–400, 2006.
- Tromp, J., Tape, C., and Liu, Q.: Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, 160, 195–216, 2005.
- Van der Hilst, R. D. and de Hoop, M. V.: Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.*, 163, 956–961, 2005.
- Van der Hilst, R. D. and Karason, H.: Compositional heterogeneity in the bottom 1000 km of Earth’s mantle: toward a hybrid convection model, *Science*, 283, 1885–1888, 1999.
- Van der Hilst, R. D., Widiyantoro, S., and Engdahl, E. R.: Evidence for deep mantle circulation from global tomography, *Nature*, 386, 578–584, 1997.
- Yang, T., Shen, Y., van der Lee, S., Solomon, S., and Hung, S.-H.: Upper mantle beneath the Azores hotspot from finite-frequency seismic tomography, *Earth Planet. Sci. Lett.*, 250, 11–26, 2006.
- Yang, T., Grand, S., Wilson, D., Guzman-Speziale, M., Gomez-Gonzalez, J., Dominguez-Reyes, T., and Ni, J.: Seismic structure beneath the Rivera subduction zone from finite-frequency seismic tomography, *J. Geophys. Res.*, 114, B01302, doi:10.1029/2008JB005830, 2009.
- Zaroli, C.: Global multiple-frequency S-wave tomography of the Earth’s mantle, Ph.D. thesis, Strasbourg University, 2010.
- Zaroli, C., Debayle, E., and Sambridge, M.: Frequency-dependent effects on global S-wave traveltimes: wavefront-healing, scattering and attenuation, *Geophys. J. Int.*, 182, 1025–1042, 2010.
- Zhao, L. and Chevrot, S.: An efficient and flexible approach to the calculation of three-dimensional full-wave Fr  chet kernels for seismic tomography – 1: Theory, *Geophys. J. Int.*, 185, 922–938, 2011a.
- Zhao, L. and Chevrot, S.: An efficient and flexible approach to the calculation of three-dimensional full-wave Fr  chet kernels for seismic tomography – 2: Numerical results, *Geophys. J. Int.*, 185, 939–954, 2011b.
- Zhao, L. and Jordan, T.: Sensitivity of frequency dependent traveltimes to laterally heterogeneous, anisotropic structure, *Geophys. J. Int.*, 133, 683–704, 1998.
- Zhao, L., Jordan, T., and Chapman, C.: Three-dimensional Fr  chet differential kernels for seismic delay times, *Geophys. J. Int.*, 141, 558–576, 2000.