A consistent multiparameter Bayesian full waveform inversion scheme for imaging heterogeneous isotropic elastic media

Li-Yu Kan, Sébastien Chevrot, Vadim Monteiller

Geophysical Journal International

Geophys. J. Int. (2023) **232,** 864–883
Advance Access publication 2022 September 16
GJI Seismology

https://doi.org/10.1093/gji/ggac363

A consistent multiparameter Bayesian full waveform inversion scheme for imaging heterogeneous isotropic elastic media

Li-Yu Kan,[1] Sébastien Chevrot [1] and Vadim Monteiller[2]

[1] *Géosciences Environment Toulouse, Observatoire Midi-Pyrénées, CNRS UMR 5563, Université Paul Sabatier, Toulouse 31400, France.*
E-mail: li-yu.kan@get.omp.eu
[2] *Aix Marseille University, CNRS, Centrale Marseille, LMA UMR 7031, Marseille 13453, France*

Accepted 2022 September 13. Received 2022 August 31; in original form 2022 June 9

SUMMARY

The inversion of complete seismic waveforms offers new perspectives to better constrain the elastic properties of Earth's interior. However, models of density and seismic velocities obtained from full waveform inversions are generally characterized by very different and uneven spatial resolutions. Because the 3-D structure of the Earth represents small deviations from average reference Earth models, the absolute values of density, V_P and V_S in the Earth are strongly correlated. Here, we exploit this strong correlation between model parameters as *a priori* information introduced into a new full waveform inversion algorithm, by considering a non-diagonal 3-D model covariance matrix in which the spatial correlations of elastic properties are described with an exponential covariance function. The inverse of such a model covariance matrix is easy to compute, and we thus have all the ingredients to construct a consistent Bayesian full waveform inversion scheme. We show that taking into account the correlations between density and seismic velocities can lead to dramatic improvements on the reconstructed models of density, seismic velocities and V_P/V_S ratio. This new imaging approach opens new perspectives for refining tomographic images of density and seismic velocities in the lithosphere and upper mantle on a regional scale by full waveform inversion of teleseismic body waves.

Key words: Inverse theory; Waveform inversion; Seismic tomography.

1 INTRODUCTION

Seismic tomography still heavily relies on the exploitation of the phase of the body and surface waves, and thus primarily constrains the shear and compressional velocities in the Earth's interior. With the continued improvement of computer power, full waveform inversion (FWI) has now become feasible, at least over regional domains of a few degrees in size, opening up new perspectives for obtaining finely resolved models of seismic velocities (Wang *et al.* 2016), density (Beller *et al.* 2017) and even seismic anisotropy (Beller & Chevrot 2020). In the inversion of complete teleseismic wavefields, the different types of seismic phases have very different contributions to the model reconstruction. Directly transmitted waves mainly constrain the long wavelengths of the velocity model, whereas later arrivals such as reflected and converted phases are sensitive to sharp velocity and density gradients. Therefore, while FWI of teleseismic *P* waves allows us to simultaneously reconstruct 3-D models of V_P, V_S and density, which are key parameters to get crucial insight into the compositional and thermal state of the lithosphere, these models are characterized by contrasting spatial resolutions.

This contribution aims at investigating the beneficial effects for FWI of a new regularization approach in model space that imposes a degree of correlation between density, V_P and V_S, the three parameters that describe isotropic elastic Earth models. Whereas these correlations are well known and already found numerous and important applications in various branches of geophysics, they have been ignored in multiparameter FWIs until now.

The correlation between compressional wave velocity and density has been particularly well studied, and several simple laws relating V_P to density have been proposed (e.g. Nafe & Drake 1957; Birch 1961; Brocher 2005). These relationships played an important role in the construction of the first reference Earth models because normal mode and dispersion data are insufficient to fully resolve the density structure. For example, the PREM model (Dziewonski & Anderson 1981) was required to satisfy Birch's law in the upper mantle. These relationships were also successfully applied to model the gravity field, predicting a density model from a tomographic model (e.g. Wang *et al.* 2016; Martin *et al.* 2021), or for joint/cooperative inversions of seismic and gravity data (e.g. Lines *et al.* 1988; Lees & VanDecar 1991). Mantle circulation models, which provide key insights into plate tectonics and the evolution of the Earth's interior, also rely on density models derived from tomographic models (e.g. Becker & O'Connell 2001; Simmons *et al.* 2006). The strong correlation of *P* and *S* station corrections (e.g.

Robertson & Woodhouse 1997) provides direct observational evidence of the correlation between the lateral variations of V_P and V_S. The global average ratio of 2.85 between the S and P corrections leads to $\nu = \partial \ln V_S / \partial \ln V_P = 1.6$ in the upper mantle, assuming a constant V_P/V_S of 1.8. This result is in fairly good agreement with the value obtained at the top of the lower mantle in a subsequent global-scale statistical analysis of P- and S-wave traveltimes (Bolton & Masters 2001). Experimental mineral physics data predict variations of ν in the upper mantle from 1.3 at low temperature to 2.2 at high temperature, taking into account the significant effect of anelasticity (Cammarano *et al.* 2003). On the other hand, the V_P/V_S ratio is more variable. For crustal rocks, it is mainly controlled by the silica content (e.g. Christensen 1996; Chevrot & van der Hilst 2000; Savard *et al.* 2018) and the fluid concentration (e.g. Watanabe 1993; Savard *et al.* 2018). For mantle rocks, this ratio varies with the Mg# (e.g. Lee 2003) or with the fraction of partial melt (e.g. Watanabe 1993).

To summarize this brief overview of the literature, the strong correlations between density and seismic velocities are now clearly established and widely exploited. However, FWIs have so far not assumed that model parameters are correlated, a simplification that can potentially result in strong artefacts in the reconstructed models, in particular of density and V_P/V_S.

The purpose of this study is first to point out that in realistic models of the Earth, the absolute values of density and seismic velocities are necessarily strongly correlated. It is then shown that taking into account the strong correlations between density and V_P and between V_P and V_S, the density, V_P, V_S and V_P/V_S models obtained by FWI are considerably improved. The paper is organized as follows. We first recall in Section 2 the Bayesian formulation of a waveform inversion problem for imaging isotropic elastic media. We then discuss how to make the inversion results consistent, that is independent of the choice of model parametrization. The key is to consider non-zero off-diagonal terms in the model covariance matrix, which describe the correlation between model parameters. Another important practical ingredient is to consider exponential kernels to describe the spatial correlation functions, as these kernels lead to an inverse covariance matrix that is a simple differential operator. The impact of considering the correlation between model parameters is then studied through a number of synthetic inversion experiments performed in a simplified continental subduction model in Section 3. These tests are performed with a hybrid numerical method that combines the frequency–wavenumber method to the spectral-element method in order to compute complete synthetic seismograms of teleseismic waves in a 3-D regional model. The results demonstrate that the correlations between model parameters compensate for the poor sensitivity to deep density and shear velocity anomalies when only teleseismic P waves are inverted. A preliminary application to long-period teleseismic P waveforms recorded along a transect deployed in southern Peru is then presented and discussed in Section 4. Finally, in Section 5, we discuss guidelines for properly defining the hyperparameters of an FWI problem, as well as the advantages of including SH waveforms in the inversion.

2 FORMULATION OF A CONSISTENT BAYESIAN FWI PROBLEM

We briefly recall the basic principles of inverse problem theory, following Tarantola (2005), to which the reader is referred for a more thorough presentation. Using the Bayes theorem, the conditional probability of a model $\mathbf{m}$ given a data set $\mathbf{d}$ is

$$\rho(\mathbf{m}|\mathbf{d}) = \frac{\rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m})}{\rho(\mathbf{d})}, \tag{1}$$

where $\rho(\mathbf{d}|\mathbf{m})$ is the probability density of $\mathbf{d}$ given the model $\mathbf{m}$, and $\rho(\mathbf{m})$ is the probability density of the model parameters. For a particular realization of the observed data $\mathbf{d}^{\rm obs}$, $\rho(\mathbf{d}^{\rm obs})$ is constant and therefore

$$\rho(\mathbf{m}|\mathbf{d}^{\rm obs}) \propto \rho(\mathbf{d}^{\rm obs}|\mathbf{m})\rho(\mathbf{m}). \tag{2}$$

If we now assume that there is no modeling error and that the data uncertainties are Gaussian, then

$$\rho(\mathbf{d}^{\rm obs}|\mathbf{m}) = \text{const.} \times \exp\left[-\frac{1}{2}(\mathbf{d}^{\rm obs} - \mathbf{g}(\mathbf{m}))^t \mathbf{C_D}^{-1}(\mathbf{d}^{\rm obs} - \mathbf{g}(\mathbf{m}))\right], \tag{3}$$

where $\mathbf{C_D}$ is the data covariance matrix. If we assume that the *a priori* information on model parameters is also Gaussian, then

$$\rho(\mathbf{m}) = \text{const.} \times \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_{\rm prior})^t \mathbf{C_M}^{-1}(\mathbf{m} - \mathbf{m}_{\rm prior})\right], \tag{4}$$

where $\mathbf{C_M}$ is the *a priori* model covariance matrix. When the dimensionality of the model space is low (less than a few thousand parameters), stochastic methods (e.g. Monte-Carlo, simulated annealing) can be used to explore $\rho(\mathbf{m}|\mathbf{d})$. For the waveform inversion problems that we will consider, with about 20 million free parameters, we have thus to consider the optimization method that finds the most probable model that minimizes the objective function (Tarantola 2005)

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\rm obs})^t \mathbf{C_D}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\rm obs}) + \frac{1}{2}(\mathbf{m} - \mathbf{m}_{\rm prior})^t \mathbf{C_M}^{-1}(\mathbf{m} - \mathbf{m}_{\rm prior}), \tag{5}$$

where $\mathbf{g}$ is the forward wave equation operator, and $\mathbf{m}_{\rm prior}$ is the mean of the Gaussian distribution of model parameters. In practice, $\mathbf{m}_{\rm prior}$ will be a smooth 1-D reference Earth model. The rationale for using a smooth 1-D reference is to avoid having sharp discontinuities, which can create artefacts in the reconstructed 3-D model and slow down the convergence of the iterative inversion algorithm (Monteiller *et al.* 2015).

Although the eq. (5) is mathematically correct, in practice it may not converge numerically to its true minimum if one of the two terms of the cost function is much larger than the other. This can happen for example in the case of a very large number of model parameters compared to the small number of data, as it is usually the case in FWI, where a limited number of carefully selected waveforms are inverted. For this reason, we define the new objective function

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\rm obs})^t \mathbf{C_D}^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{u}_{\rm obs}) + \frac{\lambda}{2}(\mathbf{m} - \mathbf{m}_{\rm prior})^t \mathbf{C_M}^{-1}(\mathbf{m} - \mathbf{m}_{\rm prior}), \tag{6}$$

where λ is a scaling parameter that is introduced to balance the contribution of the two terms of the cost function. Note that the scaling parameter could have been applied to the first term of the cost function and that therefore the definition of the model covariance matrix does not depend on it.

In this study, we will illustrate the methodology by focusing on the inversion of complete teleseismic waveforms to obtain 3-D models of the isotropic elastic structure of the Earth. However, its domain of application is more general and the method could be applied to any type of multiparameter inversion.

The minimum of function (6) can be found with an iterative algorithm (e.g. Tarantola & Valette 1982; Tarantola 2005), in which the model $\mathbf{m_k}$ at iteration k is perturbed by the quantity

$$\Delta \mathbf{m}_k = -\alpha_k \mathbf{H}_k^{-1} \cdot \gamma_\mathbf{k} \tag{7}$$

where α is the step length, $\mathbf{H}^{-1}$ the inverse Hessian, and $\gamma_\mathbf{k}$ the gradient of the cost function (6) at iteration k. This gradient is given by

$$\gamma_\mathbf{k} = \frac{\partial \chi_k}{\partial \mathbf{m}} = \mathbf{G}_k^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{g}(\mathbf{m_k}) - \mathbf{u}_{\rm obs}) + \lambda \mathbf{C}_\mathbf{M}^{-1}(\mathbf{m_k} - \mathbf{m}_{\rm prior}), \tag{8}$$

with $\mathbf{G_k}$ the derivative or Jacobian operator given by

$$\mathbf{G}_k = \frac{\partial \mathbf{g}}{\partial \mathbf{m}}(\mathbf{m_k}). \tag{9}$$

In tomographic applications based upon complete seismic waveforms, the first contribution to the gradient can be computed with the adjoint method (Tromp *et al.* 2005) whereas the second term is usually discarded.

The L-BFGS method (Nocedal & Wright 2006) is a quasi-Newton method which has rapidly gained popularity for solving FWI problems. Its main advantages are that it behaves like a preconditioned steepest descent method during the first iterations and that it has a fast convergence comparable to that of a Newton method. Moreover, the L-BFGS method does not require the computation and storage of the inverse Hessian. Instead, the L-BFGS method iteratively builds a finite-difference approximation of the inverse Hessian from the stored gradients and models obtained in previous iterations. For a more detailed description of the L-BFGS method and its application to FWI, the reader is referred to Monteiller *et al.* (2015).

2.1 Model parametrization and consistency of the inverse problem

Isotropic elastic media are described by three parameters, the density ρ, the compressional velocity $V_{\rm P}$ and the shear velocity $V_{\rm S}$. Alternatively, one may also consider other parametrizations such as (ρ, $I_{\rm P}$, $I_{\rm S}$), with $I_{\rm P}$ and $I_{\rm S}$ respectively the P and S impedances, (ρ, λ, μ), with λ and μ the two Lamé parameters, (ρ, $V_{\rm P}$, $V_{\rm P}/V_{\rm S}$), or (ρ, κ, μ), with κ the bulk modulus. In principle, these different parametrizations are equivalent, because any set of parameters can be recombined to reconstruct say the (ρ, $V_{\rm P}$, $V_{\rm S}$) model. However, in practice, owing to the overlap of scattering diagrams for the different parameters (e.g. Wu & Aki 1985; Gholami *et al.* 2013; Operto *et al.* 2013), the strong nonlinearity of FWI, and the presence of noise contamination in the data, the reconstructed tomographic models strongly depend on the chosen parametrization and regularization scheme.

Another issue stems from the physical units chosen for the model parameters. To define a norm in the model space, model parameters need to be normalized by their standard deviation, which are stored on the diagonal of the model covariance matrix. Using reduced centred model parameters, defined by

$$\hat{\mathbf{m}}_\mathbf{k} = \mathbf{C}_\mathbf{M}^{-\frac{1}{2}}(\mathbf{m}_k - \mathbf{m}_{\rm prior}), \tag{10}$$

allows us to work in an dimensionless model space, which makes the inverse problem better conditioned. Using this definition of the auxiliary model $\hat{\mathbf{m}}_\mathbf{k}$, and the chain rule for differentiation, the gradient becomes

$$\frac{\partial \chi_k}{\partial \hat{\mathbf{m}}} = \frac{\partial \chi_k}{\partial \mathbf{m}} \frac{\partial \mathbf{m}}{\partial \hat{\mathbf{m}}}. \tag{11}$$

We deduce from formula (10) that

$$\mathbf{m}_k = \mathbf{C}_\mathbf{M}^{\frac{1}{2}} \hat{\mathbf{m}}_\mathbf{k} + \mathbf{m}_{\rm prior}, \tag{12}$$

which gives

$$\frac{\partial \mathbf{m}}{\partial \hat{\mathbf{m}}} = \mathbf{C}_\mathbf{M}^{\frac{1}{2}}. \tag{13}$$

Therefore, with the new model metric, the normalized gradient becomes

$$\frac{\partial \chi_k}{\partial \hat{\mathbf{m}}} = \mathbf{C}_\mathbf{M}^{\frac{1}{2}} \mathbf{G}_k^t \mathbf{C}_\mathbf{D}^{-1}(\mathbf{g}(\mathbf{m_k}) - \mathbf{u}_{\rm obs}) + \lambda \hat{\mathbf{m}}_\mathbf{k}. \tag{14}$$

This formulation leads to a stable regularization of the inverse problem by smoothing the gradient with the model covariance matrix. Another advantage of this approach is that we can determine directly the model perturbations at each node of the spectral-element mesh, thereby avoiding back-and-forth projections in a Cartesian tomographic grid, as in Wang *et al.* (2016). This results in a simplified and versatile inversion algorithm that can be applied in arbitrarily complex 3-D spectral-element meshes. For example, to account for the effects of surface topography we just need to deform the upper edge of the mesh accordingly, which greatly simplifies the inversion algorithm.

For convenience, it is often assumed that the model covariance matrix is purely diagonal. This is a source of inconsistencies in multiparameter inversions. This problem has been clearly illustrated by Babuška & Cara (1991) with the inversion of the dispersion of the fundamental and higher modes of both Love and Rayleigh waves to determine a vertical transversely isotropic model of the upper mantle. They performed the first inversion using a transversely isotropic model parametrization in terms of β_V and $\xi = (\beta_H/\beta_V)^2$, with β_V and β_H the velocity of horizontally propagating S waves polarized horizontally and vertically respectively. However, a new inversion using parameters β_H and β_V lead to a drastically different model of the β_H/β_V ratio. To make the inversion results consistent, they pointed out that it is necessary to introduce the matrix $\mathbf{T}$ that transforms model perturbations ($\delta\beta_V$, $\delta\xi$) to ($\delta\beta_V$, $\delta\beta_H$), defined by

$$\begin{bmatrix} \delta\beta_V \\ \delta\beta_H \end{bmatrix} = \mathbf{T} \cdot \begin{bmatrix} \delta\beta_V \\ \delta\xi \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \sqrt{\xi} & \frac{\beta_V}{2\sqrt{\xi}} \end{bmatrix} \begin{bmatrix} \delta\beta_V \\ \delta\xi \end{bmatrix}. \tag{15}$$

If the model covariance matrix $\mathbf{C_M}$ is diagonal for the set of parameters (β_V, ξ) then we need to use the covariance matrix $\mathbf{C'_M}$ for parameters (β_V, β_H) given by

$$\mathbf{C'_M} = \mathbf{T} \cdot \mathbf{C_M} \cdot \mathbf{T}^t \tag{16}$$

to ensure the consistency of the inversion.

For isotropic FWI, this means that if we want to transform model parametrization (ρ, $V_{\rm P}$, $V_{\rm S}$) with the model covariance matrix

$$\mathbf{C}_{(\rho, V_{\rm P}, V_{\rm S})} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_{\rm P}}^2 & 0 \\ 0 & 0 & \sigma_{V_{\rm S}}^2 \end{bmatrix} \tag{17}$$

to parametrization (ρ, $V_{\rm P}$, $V_{\rm P}/V_{\rm S}$), according to

$$\begin{bmatrix} \delta\rho \\ \delta V_{\rm P} \\ \delta \frac{V_{\rm P}}{V_{\rm S}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{V_{\rm S}} & -\frac{V_{\rm P}}{V_{\rm S}^2} \end{bmatrix} \begin{bmatrix} \delta\rho \\ \delta V_{\rm P} \\ \delta V_{\rm S} \end{bmatrix} \tag{18}$$

then we need to define the new model covariance matrix as

$$\mathbf{C}_{(\rho, V_{\rm P}, \frac{V_{\rm P}}{V_{\rm S}})} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_{\rm P}}^2 & \frac{1}{V_{\rm S}}\sigma_{V_{\rm P}}^2 \\ 0 & \frac{1}{V_{\rm S}}\sigma_{V_{\rm P}}^2 & \frac{1}{V_{\rm S}^2}\sigma_{V_{\rm P}}^2 + \frac{V_{\rm P}^2}{V_{\rm S}^4}\sigma_{V_{\rm S}}^2 \end{bmatrix} \tag{19}$$

in order to get consistent inversion results. From these simple considerations, we thus see that the non-diagonal terms of the model covariance matrix, which describe the correlations between model parameters, are key ingredients for the formulation of a consistent inverse problem. Whereas model parametrization can be chosen at will, the assumptions that we make (implicitly or not) on the degree of correlation between model parameters can potentially strongly impact the results of inversions.

2.2 The model covariance matrix for a 3-D isotropic Earth model

The main impediment to introduce a complete non-diagonal model covariance matrix into an inversion algorithm stems from much increased computational cost and algorithmic complexity, which can indeed become prohibitive for very large 3-D problems.

Let us consider a model covariance matrix of the form

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2\mathbf{C} & r_{\rho,V_\mathrm{P}}\sigma_\rho\sigma_{V_\mathrm{P}}\mathbf{C} & r_{\rho,V_\mathrm{S}}\sigma_\rho\sigma_{V_\mathrm{S}}\mathbf{C} \\ r_{\rho,V_\mathrm{P}}\sigma_\rho\sigma_{V_\mathrm{P}}\mathbf{C} & \sigma_{V_\mathrm{P}}^2\mathbf{C} & r_{V_\mathrm{P},V_\mathrm{S}}\sigma_{V_\mathrm{P}}\sigma_{V_\mathrm{S}}\mathbf{C} \\ r_{\rho,V_\mathrm{S}}\sigma_\rho\sigma_{V_\mathrm{S}}\mathbf{C} & r_{V_\mathrm{P},V_\mathrm{S}}\sigma_{V_\mathrm{P}}\sigma_{V_\mathrm{S}}\mathbf{C} & \sigma_{V_\mathrm{S}}^2\mathbf{C} \end{bmatrix} \tag{20}$$

where the $r_{i,j}$ are the correlation coefficients between parameters i and j, $\mathbf{C}$ the spatial covariance kernel, σ_ρ the standard deviation of density, σ_{V_P} the standard deviation of V_P, and σ_{V_S} the standard deviation of V_S.

This model covariance matrix can be rewritten

$$\mathbf{C_M} = \Sigma\mathbf{S}\mathbf{R}\mathbf{S}\Sigma, \tag{21}$$

where

$$\Sigma = \begin{bmatrix} \sigma_\rho\mathbf{I} & 0 & 0 \\ 0 & \sigma_{V_\mathrm{P}}\mathbf{I} & 0 \\ 0 & 0 & \sigma_{V_\mathrm{S}}\mathbf{I} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{C}^{\frac{1}{2}} & 0 & 0 \\ 0 & \mathbf{C}^{\frac{1}{2}} & 0 \\ 0 & 0 & \mathbf{C}^{\frac{1}{2}} \end{bmatrix}, \tag{22}$$

and

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & r_{\rho,V_\mathrm{P}}\mathbf{I} & r_{\rho,V_\mathrm{S}}\mathbf{I} \\ r_{\rho,V_\mathrm{P}}\mathbf{I} & \mathbf{I} & r_{V_\mathrm{P},V_\mathrm{S}}\mathbf{I} \\ r_{\rho,V_\mathrm{S}}\mathbf{I} & r_{V_\mathrm{P},V_\mathrm{S}}\mathbf{I} & \mathbf{I} \end{bmatrix}. \tag{23}$$

The matrix $\mathbf{R}$ encodes the correlations between model parameters. The decomposition of this matrix when the correlations between the different pairs are equal is detailed in Appendix A. In the general case, that is when the correlation coefficients are not equal, the singular value decomposition of $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{V}\Lambda\mathbf{V}^t, \tag{24}$$

with $\mathbf{V}$ the matrix containing the orthogonal eigenvectors of $\mathbf{R}$. Using the decomposition of $\mathbf{C_M}$, we can derive the expression of the inverse covariance matrix

$$\mathbf{C_M}^{-1} = \Sigma^{-1}\mathbf{S}^{-1}\mathbf{V}\Lambda^{-1}\mathbf{V}^t\mathbf{S}^{-1}\Sigma^{-1}, \tag{25}$$

from which we can deduce that

$$\mathbf{C_M}^{\frac{1}{2}} = \mathbf{V}\Lambda^{\frac{1}{2}}\mathbf{V}^t\mathbf{S}\Sigma = \Sigma\mathbf{S}\mathbf{V}\Lambda^{\frac{1}{2}}\mathbf{V}^t, \tag{26}$$

and

$$\mathbf{C_M}^{-\frac{1}{2}} = \Sigma^{-1}\mathbf{S}^{-1}\mathbf{V}\Lambda^{-\frac{1}{2}}\mathbf{V}^t = \mathbf{V}\Lambda^{-\frac{1}{2}}\mathbf{V}^t\mathbf{S}^{-1}\Sigma^{-1}. \tag{27}$$

The operator $C(\mathbf{r}, \mathbf{r}')$ describes the correlation of model parameters at positions $\mathbf{r}$ and $\mathbf{r}'$. Correlation kernels can be seen as smoothing operators and their inverse as roughening operators (Oliver 1998). As can be deduced from eqs (10) and (14), both the $\mathbf{C_M}^{\frac{1}{2}}$ and $\mathbf{C_M}^{-\frac{1}{2}}$ operators are needed for the implementation of a consistent FWI algorithm. The application of the model covariance matrix to a model vector thus involves three steps: First, normalization by the standard deviation with the diagonal matrix Σ, as in eq. (10), followed by smoothing with the spatial filter $\mathbf{S}$, and finally recombination of model parameters to account for their physical correlation with the matrix $\mathbf{R}$.

In tomographic inversions, Gaussian kernels are often considered (e.g. Nataf *et al.* 1986) but they strongly filter out the short wavelengths of the model. Since the main motivation of FWI is to refine the spatial resolution of tomographic images, we must therefore use correlation kernels that decay less rapidly at high wavenumbers, such as exponential (or Laplacian) kernels (e.g. Oliver 1998; Trinh *et al.* 2017; Araujo *et al.* 2021). Since Gaussian or exponential kernels have broad support, convolving these filters on a large 3-D grid can be time-consuming and difficult to implement when the computational grid is distributed with domain decomposition. Furthermore, as pointed out by Oliver (1998), computing (and storing) the full inverse covariance matrix is a challenge. However, inverse exponential kernels can be defined analytically with finite-difference operators (e.g. Oliver 1998; Tarantola 2005; Trinh *et al.* 2017), an interesting property that has been already exploited to regularize tomographic inversions (e.g. Potin 2016; Araujo *et al.* 2021).

A 3-D exponential kernel is defined as (Oliver 1998; Trinh *et al.* 2017)

$$C(\mathbf{r}, \mathbf{r}') = e^{-\sqrt{\frac{(x-x')^2}{L_x^2}+\frac{(y-y')^2}{L_y^2}+\frac{(z-z')^2}{L_z^2}}}, \tag{28}$$

with L_x, L_y and L_z the correlation (or smoothing) lengths along the x, y and z directions, respectively. The 3-D exponential kernel admits the simple inverse operator (Trinh *et al.* 2017)

$$C^{-1}(\mathbf{r}, \mathbf{r}') = \frac{1}{8\pi L_x L_y L_z}(\mathbf{I} - \Delta_L)^2, \tag{29}$$

with

$$\Delta_L = L_x^2\partial_x^2 + L_y^2\partial_y^2 + L_z^2\partial_z^2 \tag{30}$$

the Laplacian operator (Oliver 1998; Trinh *et al.* 2017). It is easy to verify that the application of this inverse operator to the exponential covariance kernel gives the identity operator (Oliver 1998; Trinh *et al.* 2017). From eq. (29), we deduce that

$$C^{-\frac{1}{2}} = \frac{1}{\sqrt{8\pi L_x L_y L_z}}(\mathbf{I} - \Delta_L). \tag{31}$$

The $C^{-\frac{1}{2}}$ is a local differential operator which involves only the identity and the Laplacian operators. It is therefore easy to compute the results of applying this operator to a model vector, for example with a spectral-element method (SEM), which has the advantage that this of being able to be implemented on the same mesh that is used for the resolution of the weak form of the wave equation. By doing so, we also avoid having to project the model back and forth on a regular tomographic Cartesian grid, as in our previous implementations of the inversion algorithm (Monteiller *et al.* 2015; Wang *et al.* 2016). By using exponential kernels, it is therefore straightforward to introduce a non-diagonal inverse model covariance operator into the FWI. As pointed out by Trinh *et al.* (2017), for large problems involving domain decomposition, it may be simpler and more efficient to compute the filtered model $\mathbf{m_f} = C^{\frac{1}{2}}\mathbf{m}$ by solving the linear system

$$C^{-\frac{1}{2}}\mathbf{m_f} = \mathbf{m}. \tag{32}$$

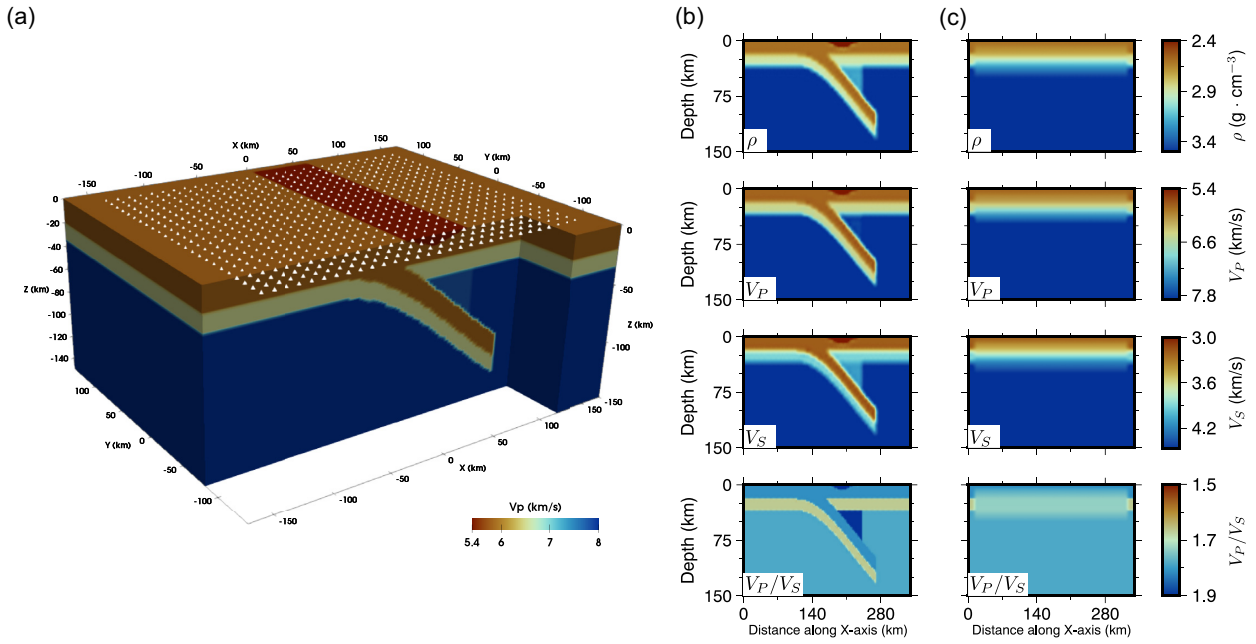

Figure 1. The continental subduction model used in the synthetic tests. The model is parametrized by the values of density, V_P and V_S at each node of the spectral-element mesh. We show here the V_P model, expressed in km s^{-1}, and the positions of the receivers located on the surface (white triangles). The slab extends down to 130 km depth, and the width of the forearc basin is 50 km.

Table 1. Values of model parameters inside each structural unit of the model shown in Fig. 1. Density ρ is in g cm^{-3}, V_P and V_S in km s^{-1}, P and S impedances I_P and I_S in g cm^{-3} km s^{-1}, and the Lamé parameters λ and μ in GPa.

Unit	ρ	V_P	V_S	V_P/V_S	I_P	I_S	λ	μ
Forearc basin	1.80	4.80	2.40	2.00	8.64	4.32	20.74	10.37
Upper crust	2.60	5.80	3.20	1.81	15.08	8.32	34.22	26.62
Lower crust	2.90	6.50	3.90	1.67	18.85	11.31	34.31	44.11
Mantle wedge	3.20	7.80	4.10	1.90	24.96	13.12	87.10	53.79
Upper mantle	3.50	8.00	4.48	1.79	28.00	15.68	83.51	70.25

Because $C^{-\frac{1}{2}}$ is sparse, only the non-zero elements need to be stored, and this system can be solved iteratively with a conjugate gradient method (Trinh *et al.* 2017). However, as mentioned by Trinh *et al.* (2017), a simpler matrix-free conjugate gradient algorithm can be implemented with the SEM. In this case, we can simply compute the matrix–vector products with the weak form of the Laplacian, which is already stored by the SEM. Therefore, the vector $\mathbf{m_f}$ is easy to compute, for negligible additional memory and computational costs.

3 NUMERICAL EXPERIMENTS

We now perform several numerical tests to assess the influence of using non-diagonal model covariance matrices in FWIs.

3.1 Description of the model

We build and mesh a 2.5-D model of continental subduction (Fig. 1) composed of five distinct units: upper crust, lower crust, upper mantle, forearc basin, and mantle wedge. This configuration is similar to the one considered in Monteiller *et al.* (2021). Inside each domain the values of density, V_P, and V_S are constant (see Table 1). These values were chosen such as to be representative of the different geological units and standard continental crust and mantle. We impose a large V_P/V_S ratio in the forearc basin and in the mantle wedge, where hydration of mantle rocks is expected to strongly decrease the shear velocity.

The computational domain is discretized in a regular Cartesian grid with dimensions 340 × 280 × 150 km along the x, y and z directions, respectively. The model is meshed with ∼8.5 km hexahedral elements. We use Lagrange polynomials of degree 7 in the three orthogonal directions so that the geometry of each element is defined by 8 x 8 x 8 =512 nodes. The grid is composed of 19 840 spectral elements and 6 922 154 nodes or grid points. Since the minimum shear wave velocity is 2.4 km s^{-1} (Table 1), the minimum period resolved is around 2.5 s. Our goal is to invert the values of density, V_P, and V_S at each node of the spectral-element mesh, which will also be our tomographic grid. The number of free parameters in our FWI experiments is therefore $3 \times 6\,922\,154 = 20\,766\,462$.

3.2 The degree of correlation between absolute and relative values of elastic parameters

Table 1 lists alternative parametrizations that can be used to describe our 3-D elastic subduction model. These parameters can be derived from the values of density and seismic velocities, according to the following formula: $I_P = \rho V_P$, $I_S = \rho V_S$, $\lambda = \rho(V_P^2 - 2V_S^2)$, and $\mu = \rho V_S^2$. Obviously, since we consider a model composed of only

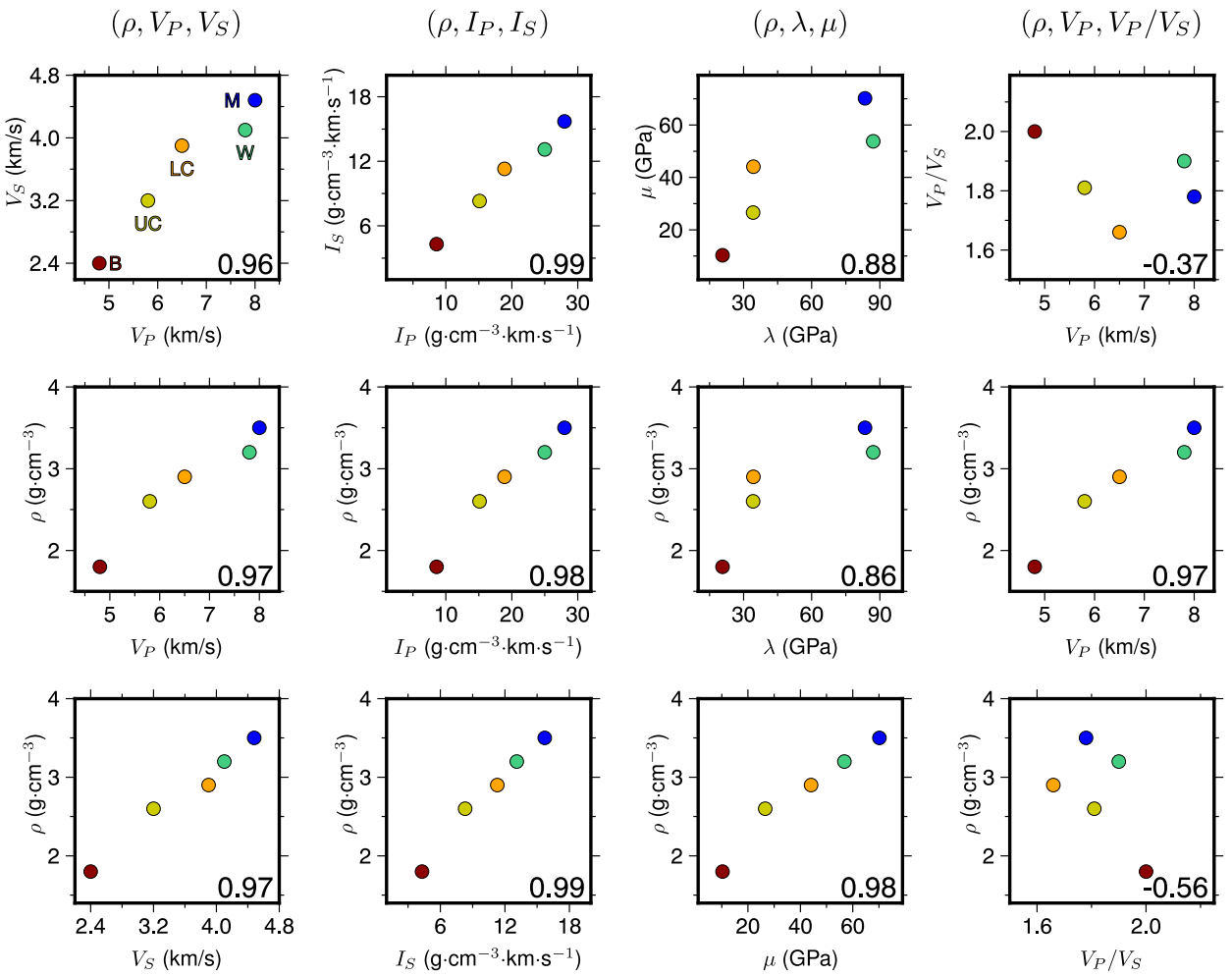

Figure 2. Scatter plots of joint variations of the three pairs of parameters for the four parametrizations (ρ, V_P, V_S), (ρ, I_P, I_S), (ρ, λ, μ) and (ρ, V_P, V_P/V_S). The values of the parameters for the five geological units are indicated by coloured circles: forearc basin (B: Brown circles), upper crust (UC: yellow circles), lower crust (LC: orange circles), mantle wedge (W: green circles) and upper mantle (M: blue circles). The parameters' units are the same as those given in the caption of Table 1. The correlation coefficient for each pair of parameters is indicated in the lower right corner of the diagrams.

five distinct homogeneous units, the variability of each parameter is rather limited. Nevertheless, the examination of the joint variations of the different pairs of parameters gives an insight into their degree of correlation and thus into the *a priori* information that could be incorporated in the inversions.

We consider the four parametrizations (ρ, V_P, V_S), (ρ, I_P, I_S), (ρ, λ, μ) and (ρ, V_P, V_P/V_S). Fig. 2 illustrates the joint variations of the three pairs of parameters corresponding to these parametrizations, with the corresponding correlation coefficients given at the bottom right of each diagram. The first three parametrizations are characterized by large (>0.86) correlation coefficients. The correlations between model parameters are particularly strong (>0.96) for the (ρ, V_P, V_S) and (ρ, I_P, I_S) parametrizations. In general, these plots suggest that imposing a diagonal model covariance matrix to regularize the inversion, that is assuming that model parameters are uncorrelated, is clearly a poor assumption that can potentially bias the results of FWIs. The correlation coefficients for the three pairs of parameters are very similar to the first three parametrizations, which suggests that building the model covariance matrix with equal correlation coefficients is a reasonable assumption. In that case, the decomposition of the model covariance matrix is very simple (see Appendix A). The last parametrization (ρ, V_P, V_P/V_S) displays a very different behaviour. Whereas the correlation between ρ and V_P is strong, with a correlation coefficient of 0.97, the V_P/V_S ratio is poorly correlated to either ρ or V_P, with an even negative correlation coefficient. For this parametrization, assuming a diagonal model covariance matrix is thus probably less prejudicial.

The correlation coefficients measured between the three parameter pairs in the first two parametrizations are all larger than 0.96. Such large values may be surprising at first glance, but they are a true reflection of our knowledge of the Earth's interior (and not a bias in the methodology followed to build this model). To demonstrate this, we performed a simple statistical experiment with the (ρ, V_P, V_S) parametrization. We generate 1000 random models by first selecting 10 values of V_P drawn at random between 4.5 and 8.5 km s^{-1}. We then derive the values of V_S from those of V_P, using a V_P/V_S ratio that varies randomly between 1.62 and 1.98. These values are representative of the variations in our synthetic model. Fig. 3 presents the histogram of the correlation coefficients obtained, which shows a clear peak at 0.96, that is at the same value obtained with our subduction model. We again obtain large correlation coefficients (>0.80) if we randomly vary the V_P/V_S ratio from 1.44 to 2.16, that is over an extreme and most likely unrealistic range of variations. We would obtain similar numbers with the pair

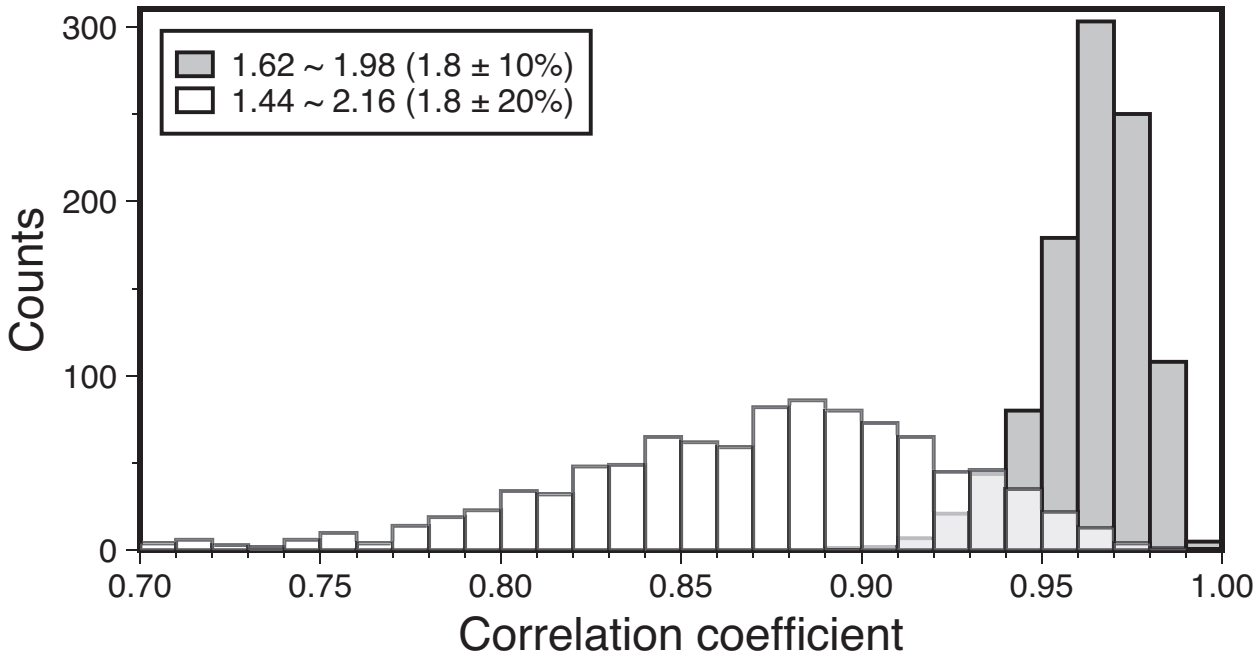

Figure 3. Histogram of the correlation coefficients measured in random models of V_P and V_S, with V_P/V_S ratio ranging from 1.62 to 1.98 (grey bars), and from 1.44 to 2.16 (white bars).

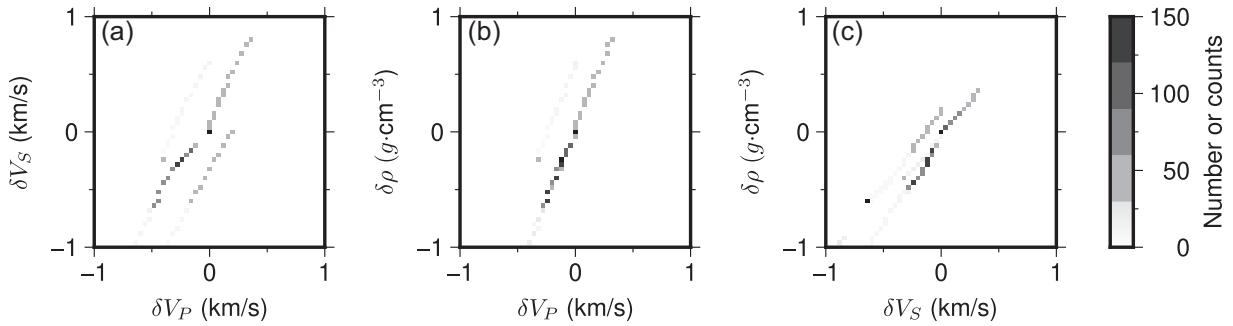

Figure 4. Scatter plots showing the joint variations of model perturbations for the three pairs of parameters for the (ρ, V_P, V_S) parametrization: (a) δV_S versus δV_S, (b) $\delta\rho$ versus δV_P, and (c) $\delta\rho$ versus δV_S. The model perturbations are defined by the difference between the value in the true model and in the initial 1-D background model. The computed correlation coefficients are about 0.97 for these three distributions.

of parameters (ρ, V_P). From this simple experiment, we therefore conclude that because heterogeneities in the Earth are moderate, of the order of a few percent with respect to a reference 1-D model, the absolute values of density and seismic velocities are strongly correlated.

Fig. 4 presents the distribution of the model perturbations $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_{\text{prior}}$, where $\mathbf{m}_{\text{prior}}$ is the smooth 1-D model. The scatter plots clearly show that the model perturbations are not randomly distributed. On the contrary, they seem to be strongly correlated, as quantified by the correlation coefficients which are all around 0.97. This strong correlation observed between the different perturbations of the model (or relative parameters) simply results from the large correlations that exist between the absolute values of parameters. For example, variations of model perturbations inside a homogeneous unit such as the lower crust will change with the depth position of the tomographic node considered, and will therefore exhibit the same correlation as the one observed in the 1-D reference model. Similarly, at a given depth, the spatial variations of the elastic parameters between the different units will also follow on average the same scaling relationships observed on the absolute elastic parameters. Therefore, these simple considerations suggest that assuming that the elastic parameters vary independently in the Earth is generally a poor assumption that should be abandoned.

3.3 Computation of synthetic seismograms with the FK/SEM hybrid method

We use a numerical hybrid method to compute the complete regional wavefield produced by an incoming teleseismic body wave (Monteiller *et al.* 2013, 2021). Owing to the small size of our regional grid, the curvature of both the wave fronts and of the Earth can be neglected (Monteiller *et al.* 2021). We can thus approximate the incident wavefields with plane waves, and use the FK/SEM hybrid modeling method as described in Monteiller *et al.* (2021). The FK method is particularly efficient because for teleseismic applications we only need to compute the incident wavefield corresponding to a single wavenumber. Another advantage of the FK method is that, in contrast to AxiSEM, the amount of computations scales linearly with the maximum frequency, and thus it is possible to reach

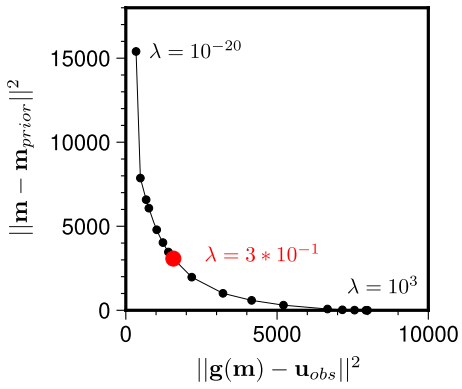

Figure 5. L-curve showing the variations of data and model terms of the cost function for different values of the regularization parameter λ. The optimal regularization coefficient (λ = 0.3) is chosen at the kink of the L-curve.

high frequencies with modest computational resources. By limiting the time-consuming 3-D computations inside the regional domain, hybrid methods make FWIs feasible on moderate-size clusters.

We compute the synthetic seismograms of teleseismic *P* waves coming from four different backazimuths: 0°, 70°, 180°, and 300°. The incident ray parameters of the four incoming plane waves correspond to epicentral distances of 60° for the first two, and 30° for the last two. The wavefields are recorded by a 2-D array of 720 receivers located on the surface, with a regular 10 km spacing. We consider a Gaussian source wavelet with a dominant period of 2.5 s, which is slightly longer than the resolution of the mesh to avoid numerical noise. To investigate the impact of noise in the inversion, we also generate noisy synthetic seismograms by adding real noise time-series recorded by station MLS from the French RLBP network (RESIF 1995). The amplitude of the noise is tuned to get an S/N ratio of 6, a value that may seem high but that reflects the strict criteria applied to data selection in FWI applications.

3.4 Results

3.4.1 Reference case: (ρ, V_P, V_S) inversion with a diagonal model covariance matrix

We perform a first FWI of the noisy data set using model parametrization (ρ, V_P, V_S), assuming that the parameters are uncorrelated. We thus basically reproduce the approach followed in our previous implementations of FWI on synthetic (Monteiller *et al.* 2015) or real data (Wang *et al.* 2016) but with a regularization scheme relying on the model covariance matrix, as detailed in the previous section. This first inversion will constitute our reference test case. We briefly describe the inversion strategy here and refer the reader to Monteiller *et al.* (2015) for more details. We use a hierarchical iterative L-BFGS inversion, starting with a long period inversion of data filtered with a Butterworth bandpass filter between 0.04 and 0.1 Hz in order to obtain a long-wavelength velocity model. For this first, inversion we use a smooth 1-D model derived from the ak135 reference Earth model (Kennett *et al.* 1995). The value of the regularization parameter λ is chosen at the kink of the L-curve obtained by plotting the joint variations of the data and model misfit contributions to the cost function (6) (Fig. 5). We then perform successive inversions in which we progressively decrease the cut-off period to 8, 6, 4 and finally 2.5 s, using at each iteration the model obtained at the previous iteration as the starting model.

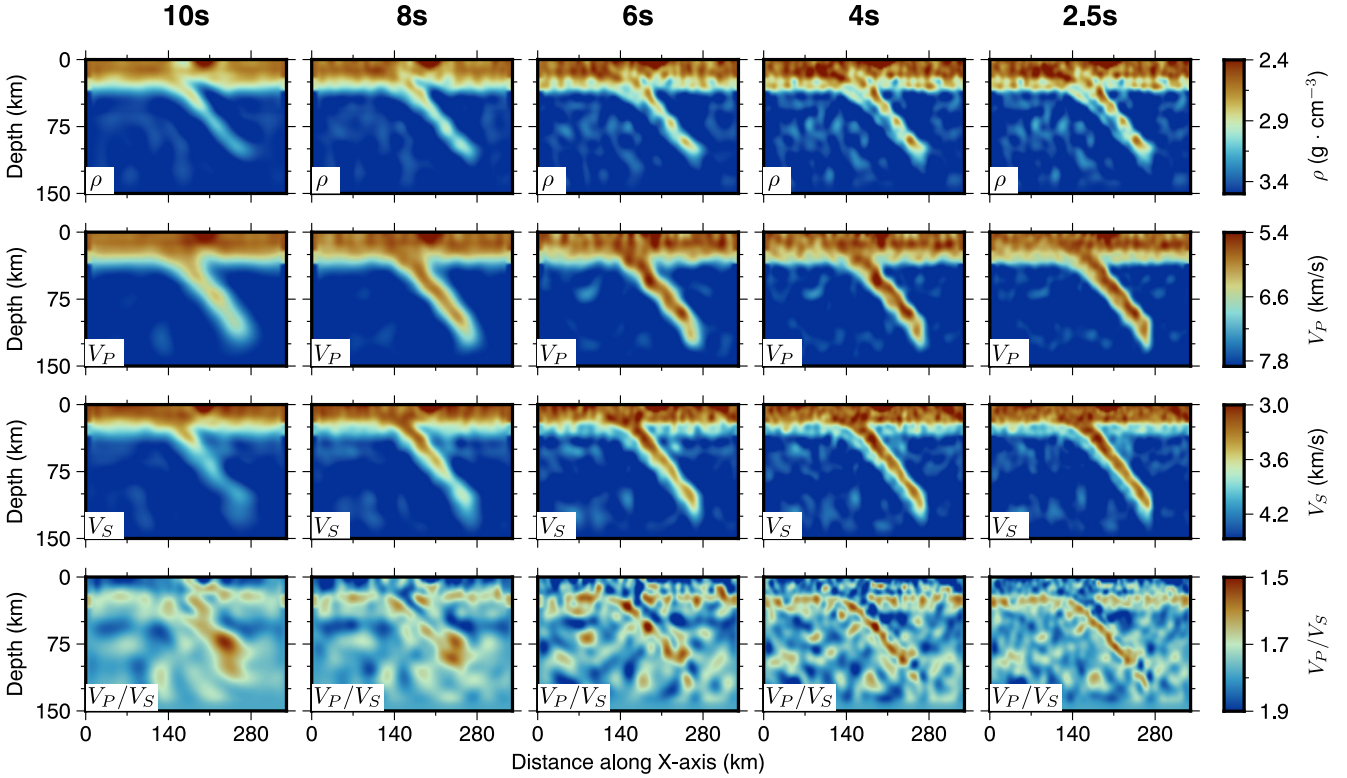

Figure 6. Results of multiscale FWI on noisy data with (ρ, V_P, V_S) parametrization, assuming a diagonal model covariance matrix and a Laplacian smoothing length of 5 km. The first inversion starts from long period filtered in 25 s to 10 s, from which we take the model as the input of the next run and decrease the cut-off period progressively to 8, 6, 4 and finally 2.5 s.

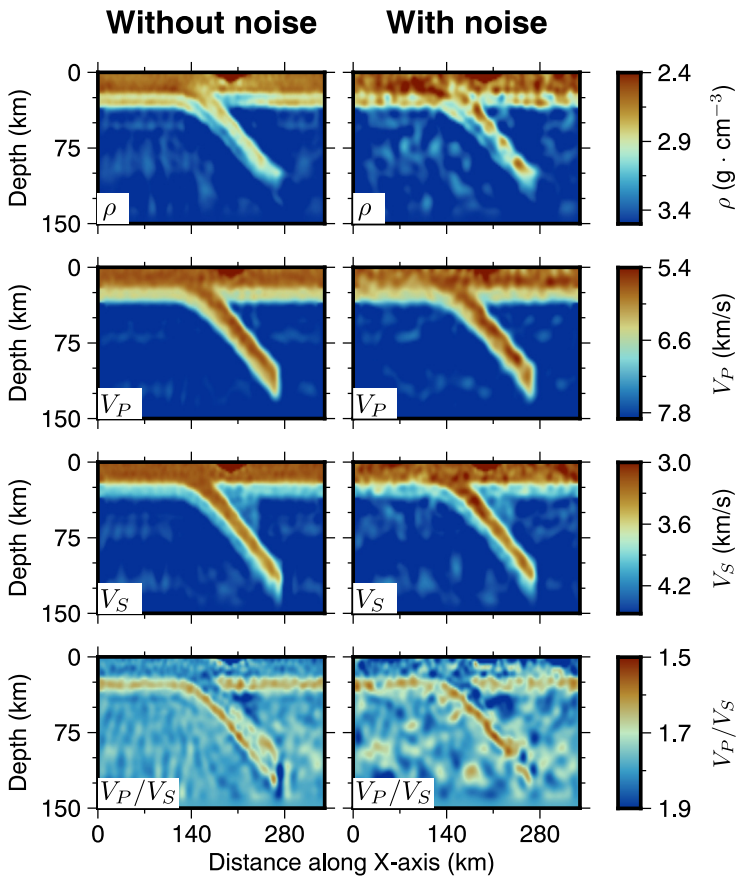

Figure 7. Results of FWI obtained on noise-free (left-hand column) and noisy (right-hand column) data using a diagonal model covariance matrix.

The correlation (or smoothing) length of the exponential correlation operator is set to 5 km along the three spatial dimensions for all the periods. At each iteration, we compute synthetic seismograms in the current model which are compared to the data to be fitted, here the data is computed in the true subduction model. The time windows considered for the measurements of waveform misfits start 5 s before and end 70 s after the *P*-wave arrival. The waveform residuals are injected at the position of the receivers to compute the adjoint wavefields. The gradients corresponding to each parameter are then computed by correlating the forward wavefield with the adjoint wavefield, following the method described in Tromp *et al.* (2005). The standard deviations for density, V_P and V_S are set to 0.27 g cm^{-3}, 0.65 km s^{-1} and 0.37 km s^{-1}, respectively. These standard deviations were simply estimated from the distributions of model parameters in the input model. The inversion algorithm is iterated until the inversion fails to meet the Wolfe conditions during the line search, or when the reduction of the cost function goes below 0.1 per cent.

Fig. 6 shows cross-sections along the x axis in the ρ, V_P, and V_S models obtained at each stage of the hierarchical inversions. We can observe that the spatial resolution improves and that the interfaces get sharper when the cut-off period decreases. The V_P model seems to be the best resolved, especially at depth. The deep part of the slab is well reconstructed only in the V_P model. The negative anomaly related to the forearc basin has a clear signature on the density and V_S models. In general, the ρ and V_S models are characterized by a finer resolution, but only in the shallow part of the model. For a given frequency content, the wavelengths of compressional waves are about twice that of shear waves. In addition, because we only consider teleseismic *P* waves here, it is the phase (and amplitude) of directly transmitted *P* waves that contribute to the V_P model and thus mainly constrain the long wavelength V_P model. In contrast, the V_S model is mainly controlled by *P*-to-*S* reflections and/or conversions which are produced on sharp gradients or discontinuities in the underlying medium. This basically explains the main differences observed between the V_P and V_S models. Not surprisingly, the V_P/V_S model derived from the V_P and V_S models shows many artefacts that simply results from the different spatial resolutions in these two models. The sensitivity to density mainly comes from the reverberations between the free surface and internal discontinuities such as the Moho or the slab edges. This sensitivity decreases with depth, and almost vanishes beneath 75 km depth. The density model is also contaminated by significant artefacts in the upper crust. Therefore, whereas seismic waveforms do have some sensitivity to density, in particular at shallow depth, seismic data alone are not sufficient to fully recover the density structure.

For comparison, the results of the hierarchical inversion obtained on noise-free and noisy data are shown in Fig. 7. The degradation of spatial resolution and the contamination by small-scale artefacts are especially visible in the density and V_S models.

3.4.2 Consistent inversion results with different model parametrizations

As noted by Nataf *et al.* (1986) and Babuška & Cara (1991), changing the parametrization without changing the covariance matrix accordingly leads to inconsistent inversion results. We now perform

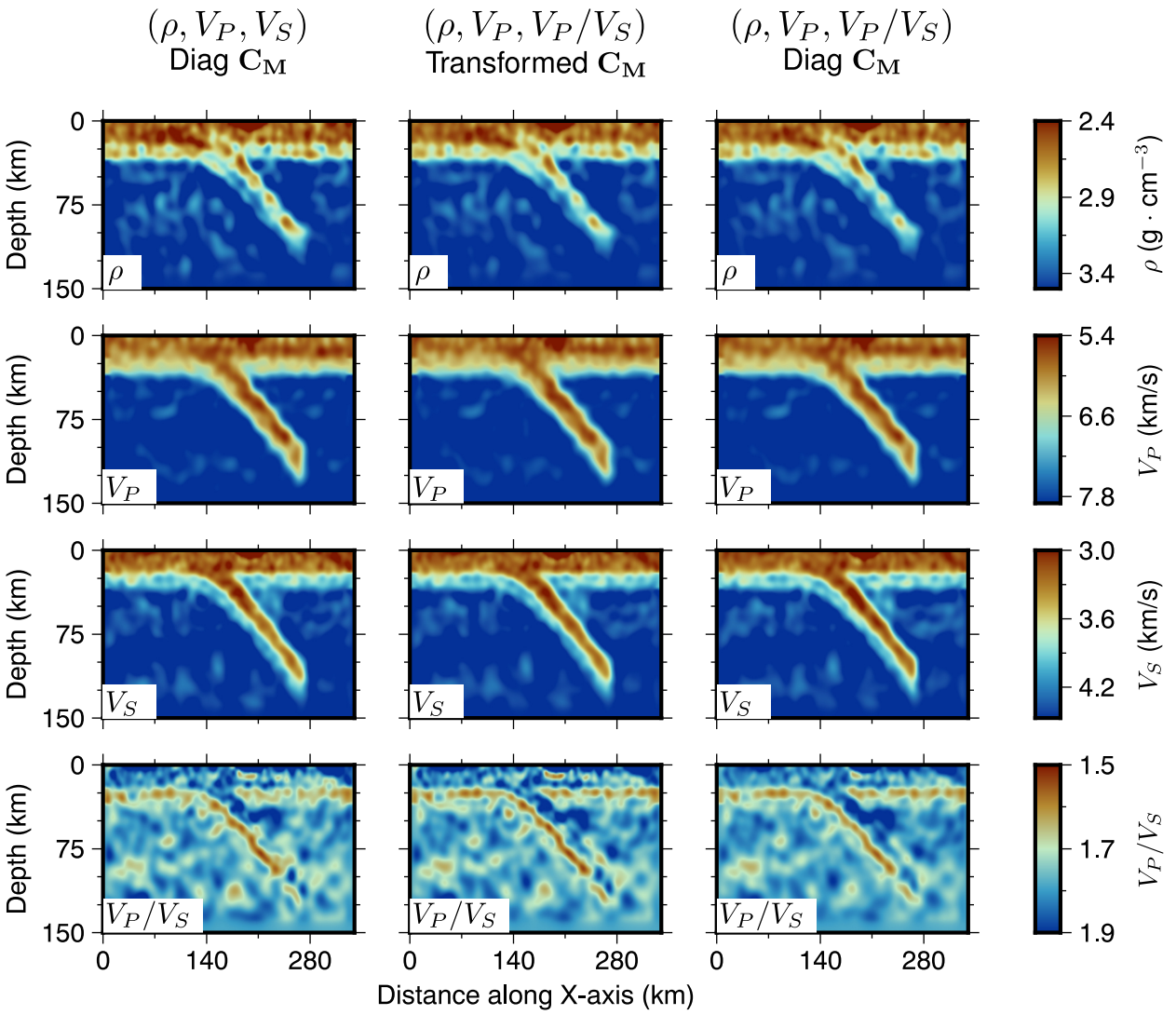

Figure 8. FWI results obtained with (a) the (ρ, V_P, V_S) parametrization and a diagonal model covariance matrix, (b) the (ρ, V_P, V_P/V_S) parametrization and a transformed $\mathbf{C_M}$ defined in eq. (19), and (c) the (ρ, V_P, V_P/V_S) and a diagonal $\mathbf{C_M}$. In all cases, the data are band-pass filtered between 0.04 and 0.40 Hz. Note the similarity of models (a) and (b) obtained with different parametrizations.

the same inversion as in the reference case, but using this time the (ρ, V_P, V_P/V_S) parametrization. The results of the hierarchical inversion for this parametrization are shown in the third column of Fig. 8 whereas the results obtained with the (ρ, V_P, V_S) parametrization are shown in the first column of Fig. 8. The reconstructed models obtained with the two parametrizations and diagonal model covariance matrices clearly differ. With the new parametrization (ρ, V_P, V_P/V_S), the V_P and V_S models seem to be slightly better reconstructed, especially in the deep part of the subduction. The general improvement in the V_P and V_S models results from the more realistic *a priori* assumption that V_P and V_P/V_S are independent, as suggested by the last column of Fig. 2.

We now consider the model covariance matrix given by (19), that is we assume that V_P and V_S are independent, but keep the same inversion parameters (ρ, V_P, V_P/V_S). The results, shown in the middle column of Fig. 8, are very similar to those obtained with the initial (ρ, V_P, V_S) parametrization (shown in the first column of Fig. 8). In particular, the similarity of the V_P/V_S models highlights the consistency of the V_P and V_S models. These simple tests thus confirm that the parametrization choice is not so important provided that the model covariance matrix is transformed accordingly in order to make the inversion consistent.

3.4.3 (ρ, V_P, V_S) inversion with a non-diagonal model covariance matrix

We now consider the (ρ, V_P, V_S) parametrization and a non-diagonal model covariance matrix in the form (20). We perform a hierarchical inversion in which we impose $r = 0.97$ for all parameter pairs for lower cut-off periods ranging from 10 to 4 s, and $r = 0.80$ in the final inversion at 2.5 s. The resulting models at each stage of this hierarchical inversion are shown in Fig. 9. At 10 s, the ρ, V_P, and V_S structures of the subducted slab and basin are generally well recovered, with a notable blurring effect that comes from the limited spatial resolution of this low-frequency inversion. The final ρ, V_P, and V_S models are all nicely recovered with a sharp definition. The remaining small-scale artefacts could perhaps be reduced

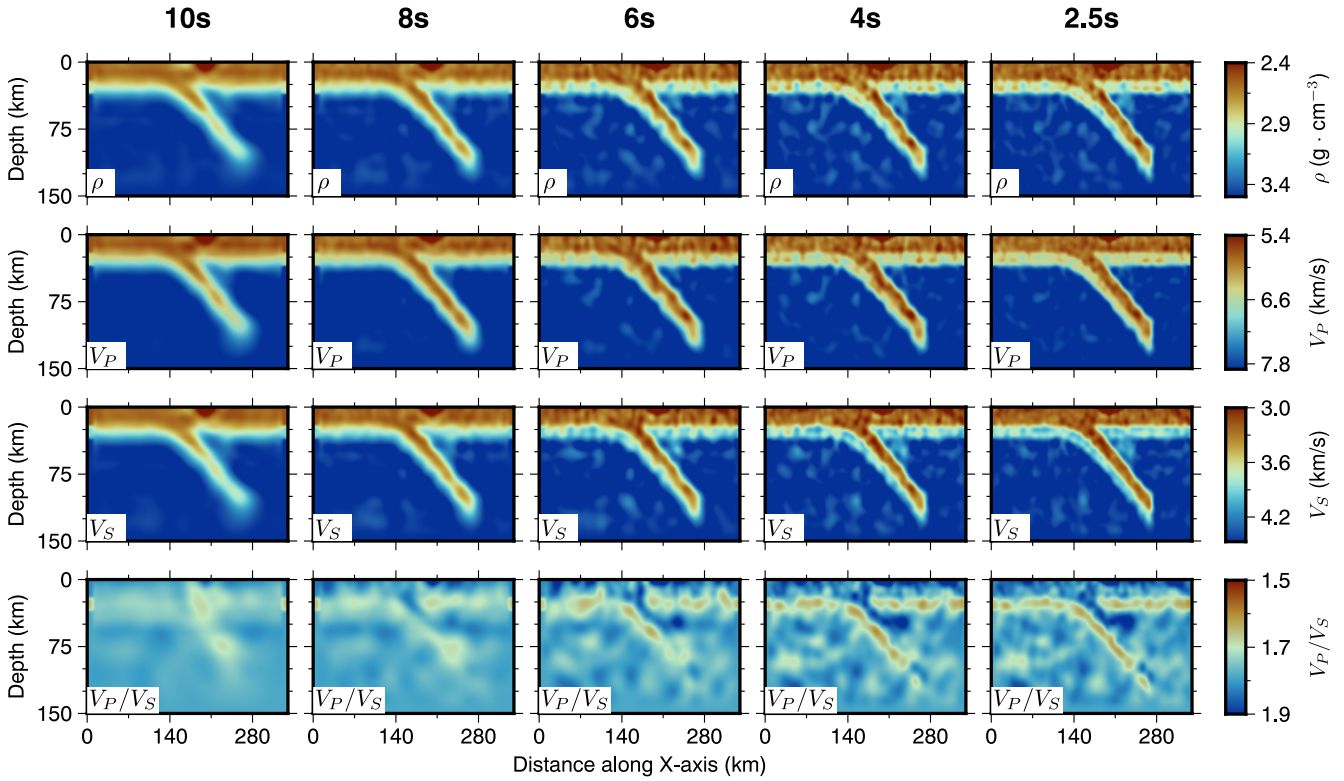

Figure 9. Results of five stages of multiscale FWI with (ρ, V_P, V_S) parametrization, assuming a non-diagonal model covariance. The first inversion starts from long period filtered in 25 to 10 s, from which we take the model as the input of the next run and decrease the cut-off period progressively to 8, 6, 4 and finally 2.5 s.

by additional smoothing, but this would probably also degrade the spatial resolution. Comparing these models to those previously obtained with a diagonal model covariance matrix (Fig. 7 right), we can therefore conclude that introducing strong correlations between model parameters significantly improves the results of FWI.

Fig. 10 shows the 2-D scatter plots of model parameters corresponding to the first and last stage of the hierarchical inversions with (Fig. 9) and without correlation (Fig. 6). After the long period inversions (10 s), in all cases, neither the basin (brown circle) nor the mantle wedge (green circle) are recovered. The spatial resolution is simply insufficient to resolve these small-scale structures correctly. When the model parameters are assumed to be uncorrelated, they can deviate significantly from the input model values (Fig. 10b). When high correlation coefficients are imposed ($r = 0.97$), we observe a very different behaviour. The model parameters follow a quasi-linear distribution between the values in the different homogeneous tectonic units, with very small deviations from these average trends (Fig. 10c). The algorithm has imposed a local correlation, on a support corresponding to the spatial extent of the exponential kernels. In summary, the inversion successfully exploited the averaged scaling relationships that exist between the model parameters, for a better reconstruction of the input model. Fig. 11 presents the errors on the density, V_P, V_S and V_P/V_S computed in the final models with or without correlation imposed between the model parameters. The standard deviations obtained are 0.15 g cm^{-3}, 0.23 km s^{-1}, 0.14 km s^{-1} and 0.05 for density, V_P, V_S and V_P/V_S respectively for the diagonal model covariance matrix, and 0.08 g cm^{-3}, 0.16 km s^{-1}, 0.11 km s^{-1}, 0.03 respectively for the non-diagonal model covariance matrix. Therefore, using a non-diagonal model covariance matrix leads to variance reductions of 71 per cent for the density, 51 per cent for V_P, 38 per cent for V_S, and 64 per cent for the V_P/V_S ratio. The improvement of the recovered model is thus very significant.

Fig. 12 compares the evolution of the residual variance with the number of iterations in long period inversions with diagonal and non-diagonal model covariance matrices. When the model parameters are assumed to be uncorrelated, convergence is slower in the first 10 iterations, and a much larger number of iterations is necessary to reach convergence. Intuitively, when strong correlations are imposed between model parameters, this restricts the model space explored by the iterative inversion algorithm, which both accelerates the convergence and restricts the occurrence of artefacts in the reconstructed models.

4 APPLICATION TO REAL DATA: THE SOUTHERN PERU TRANSECT

We now illustrate the impact of using a complete non-diagonal model covariance matrix on a real data set, the PE profile of the Peru Subduction Experiment (Clayton 2013; Fig. 13). The motivation of this experiment was to image the subduction of the Nazca plate beneath South America. Under this profile, the subduction of the Nazca plate is normal, with a dip of the Nazca plate around 30°. Receiver functions have shown that the crustal thickness is increasing from 25 km close to the coast to 75 km beneath the Altiplano (Phillips *et al.* 2012).

We only provide here a brief outline of the data selection and preparation. The results of this tomographic study will be presented and discussed in full detail elsewhere. We retrieved all the data corresponding to teleseismic events with magnitude larger than 6 from IRIS-DMC. For this study, we selected the records of 16

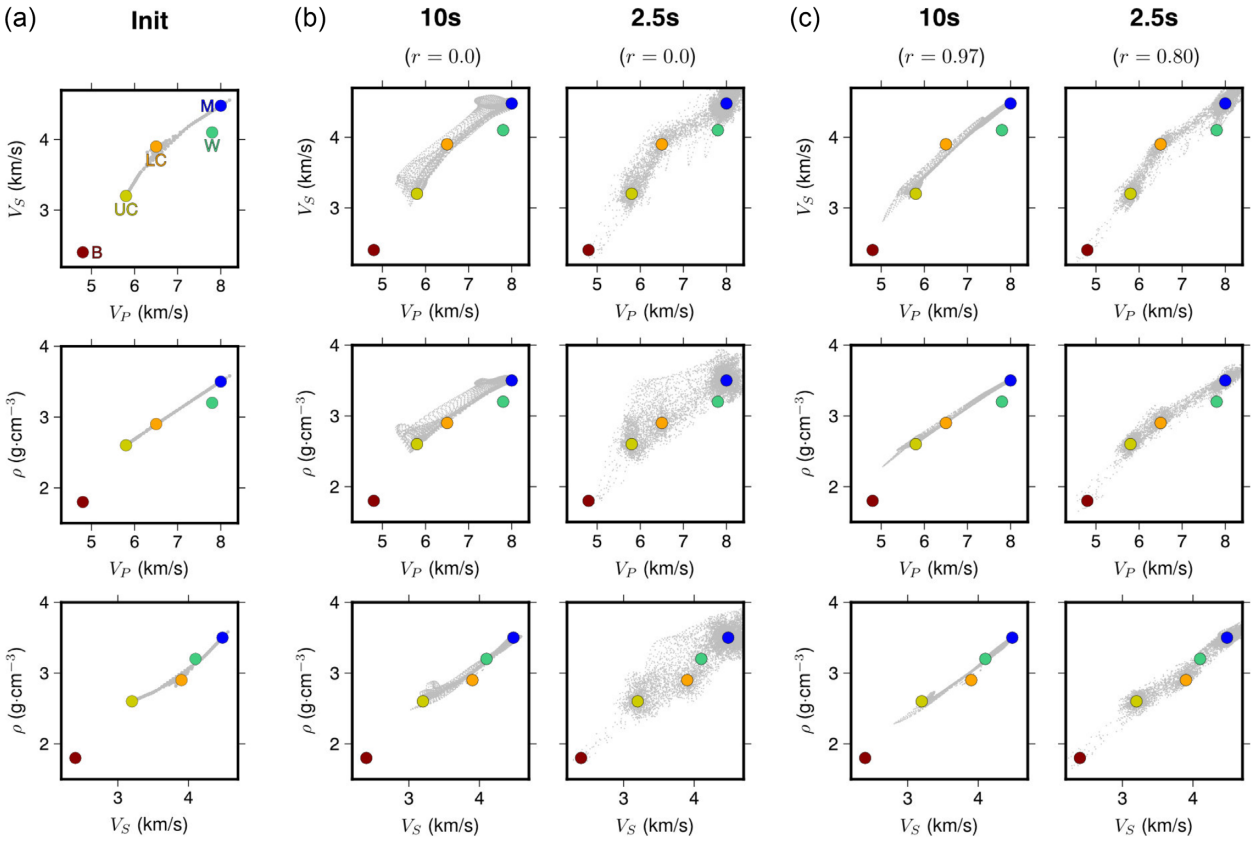

Figure 10. Scatter plots for inversion with and without correlations between model parameters ρ, V_P and V_S. (a) Scatter plots for initial smooth 1-D model. (b) Scatter plots for models with zero correlations between model parameters for the 10 s and 2.5 s inversions. (c) Scatter plots for models with strong correlations (0.97 and 0.80) between model parameters corresponding to the 10 and 2.5 s inversions. The grey dots represent the value in the corresponding model. The colour dots are the same as detailed in Fig. 2(b).

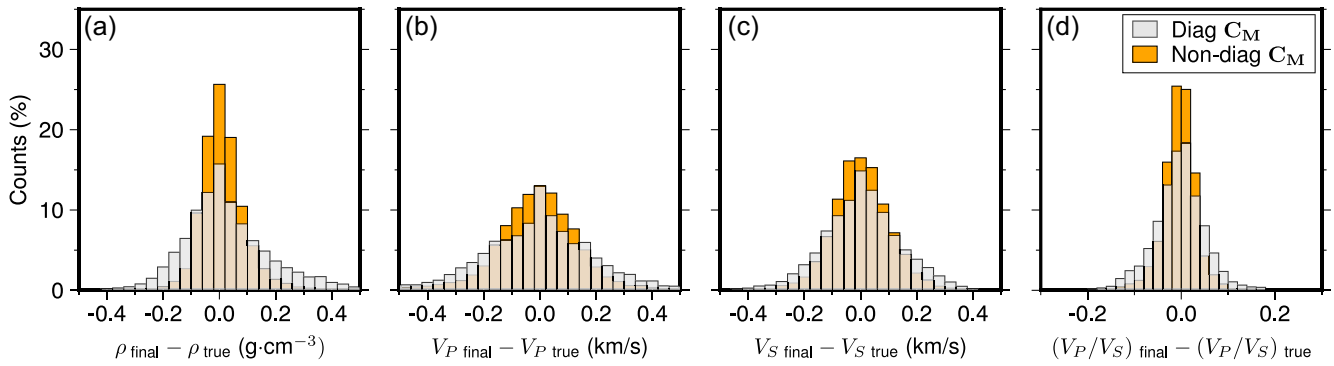

Figure 11. Histogram of the differences of density (a), V_P (b), V_S (c) and V_P/V_S (d) between the values in the true model and the final model obtained with a diagonal covariance matrix (grey) and a non-diagonal covariance matrix (orange). The standard deviations are 0.15 g cm^{-3}, 0.23 km s^{-1}, 0.14 km s^{-1} and 0.05 for density, V_P, V_S and V_P/V_S respectively for the diagonal model covariance matrix, and 0.08 g cm^{-3}, 0.16 km s^{-1}, 0.11 km s^{-1} and 0.03 respectively for the non-diagonal model covariance matrix.

events with large S/N ratios. We invert both the radial and vertical component records of the *P* waves, over time windows that start 5 s before the *P* arrival and end from 50 to 70 s after, depending on the event. The data are band-pass filtered between 0.04 and 0.08 Hz. We first compute synthetic impulse responses using a Gaussian source wavelet with a dominant period of 4 s with the hybrid AxiSEM/SEM method (Monteiller *et al.* 2021). For each event, we estimate the source wavelet by deconvolving the vertical component records from their corresponding synthetic impulse responses, aligning the resulting traces, and finally stacking them. The complete synthetic seismograms are obtained by convolving the impulse responses with the corresponding source wavelet. The inversions are started from a smooth 1-D reference model derived from the ak135 reference Earth model (Kennett *et al.* 1995). We perform two types of inversions. In the first inversion, we assume that model parameters are independent and thus consider a diagonal model covariance matrix. In the second inversion, we impose a correlation coefficient of 0.9 between all the pairs of parameters. The smoothing length in both inversions is set to 30 km.

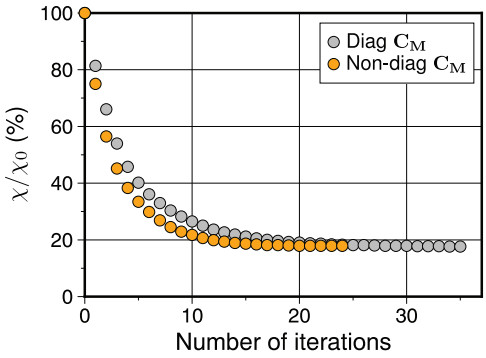

Figure 12. Evolution of the normalized cost reduction in the long period inversions (data filtered between 0.04 and 0.10 Hz) with (orange dots) and without (grey) correlation imposed between model parameters. The inversion with the non-diagonal covariance matrix converges faster.

The convergence is achieved after 19 and 17 iterations in inversions 1 and 2 respectively, with a variance reduction of about 80 per cent in both cases. The final models obtained are presented in Fig. 14(a), which shows the absolute values of the density and seismic velocities, while Fig. 14(b) shows the deviations of density and seismic velocities with respect to the depth-averaged 1-D models. The V_P and V_S models are relatively similar in both inversions, with a pronounced low-velocity anomaly down to about 80 km depth that corresponds to the thickened crust beneath the Altiplano. The V_P model is slightly better resolved at depth, as in the synthetic experiments shown previously. In particular, the fast velocity anomaly related to the subducting Nazca plate is imaged in the V_P model, but with a much finer resolution in the inversion that accounts for correlations between model parameters. However, the benefits of accounting for the correlations between model parameters are more pronounced on the density and V_S models. From this very simple inversion experiment on real data, we can therefore already conclude that taking into account the correlation between density and seismic velocities seems very promising to improve the FWI results.

5 DISCUSSION

5.1 Tuning FWI hyperparameters: model parametrization, standard deviations, correlation coefficients and smoothing length

If we assume that the model parameters are not correlated, then the parametrization (ρ, V_P, V_P/V_S) should give better results than the parametrization (ρ, V_P, V_S). Indeed, Fig. 2 suggests that to first order V_P/V_S and V_P are weakly correlated ($r = -0.37$), while the assumption that V_P and V_S are independent is clearly wrong, as these two parameters are strongly correlated ($r = 0.96$). The fact that the (V_P, V_P/V_S) parametrization leads to a better reconstruction of the V_P/V_S ratio than the (V_P, V_S) parametrization has been known for a long time (e.g. Thurber & Eberhart-Phillips 1999), and this property has been widely exploited in local earthquake tomography, where one seeks to jointly image P and S velocity variations form P and S traveltimes (e.g. Thurber *et al.* 1995). The scatter plots shown in Fig. 2 show that the parameters are strongly correlated with the parametrizations (ρ, I_P, I_S) and (ρ, V_P, V_S). This suggests that choosing one of these two parametrizations, and also assuming high ($r > 0.95$) correlation coefficients, is probably the best and simplest strategy for obtaining images of the Earth's interior by multiparametric inversion of complete waveforms.

We now explore the influence of correlations between model parameters by performing inversions at low frequency (0.04–0.1 Hz) with r equal to 0.00, 0.50 and 0.97, with a constant smoothing length of 5 km. The results of these three inversions are presented in Fig. 15. The most notable differences are observed in the reconstructed ρ models, for which the correlations imposed between V_P and V_S compensate for the limited sensitivity of P waves to the density structure. The best density model is obtained for very large correlation coefficients ($r = 0.97$) between density and seismic velocities. In this case, the sedimentary basin is more sharply defined, the slab is better imaged, especially in its deep part, and the small-scale spurious artefacts that are clearly visible in the inversion with uncorrelated parameters have almost completely disappeared. A similar though less dramatic improvement is observed on the reconstructed V_S models. As with the density reconstruction, the correlation imposed between V_P and V_S compensates for the lack of sensitivity to deep, large-scale V_S heterogeneities. Not surprisingly, the correlation coefficient imposed between V_P and V_S has also a large impact on the V_P/V_S ratio. For small values of r, the V_P/V_S model is contaminated by strong artefacts resulting from the marked difference in resolution between the V_P and V_S models, the latter being characterized by a very low resolution at depth. This long period inversion test suggests that complete, non-diagonal model covariance matrices are key ingredients to image lateral variations of density.

Different tests (not shown) suggest that the models are best reconstructed with moderate smoothing lengths (5–10 km) and that significant artefacts appear when smoothing lengths greater than about 20 km are used. The reason for this can be understood by looking at Fig. 16, which shows the normalized gradients of V_S obtained after the first long period inversion (10–25 s period) with different smoothing lengths. When no smoothing is applied (Fig. 16a), strong and localized negative singularities are observed in the very shallow part of the grid, just below the stations. These anomalies are produced by the sources of the adjoint wavefield, located at the positions of the sensors on the free surface. With a 5 km smoothing length (Fig. 16b), these singularities are smoothed out while at the same time all the details of the gradient are preserved. With a larger smoothing length of 20 km (Fig. 16c), most of the structural details of the gradient are lost. This comparison suggests that a 5 km smoothing length represents a good compromise between cleaning up numerical artefacts and preserving structural details in the gradients.

Intuitively, we can understand this by the fact that in the hierarchical inversion algorithm, we first invert long period signal ($T > 10$ s) to obtain smooth long wavelength models. At this point, the gradients are smooth and do not require additional smoothing. As the frequency cut-off is progressively increased, the inversions start to resolve smaller structural details that could potentially be erased by imposing too much smoothing. It is therefore advisable to impose a minimum amount of smoothing in inversions.

The last hyperparameters that need to be defined are the standard deviations for each class of model parameters. In all the inversions we have presented so far, we have used $\sigma_\rho = 0.27$ g cm^{-3}, $\sigma_{V_P} = 0.65$ km s^{-1} and $\sigma_{V_S} = 0.37$ km s^{-1}. These values were simply deduced from inspection of the scatter plots shown in Fig. 2. Fig. 17 shows the results of the long period inversion obtained for different values of σ_{V_P}. When the value of σ_{V_P} is reduced, the anomalies in the V_P models are smaller in amplitude and smoother. In contrast, the density model shows sharper variations, especially

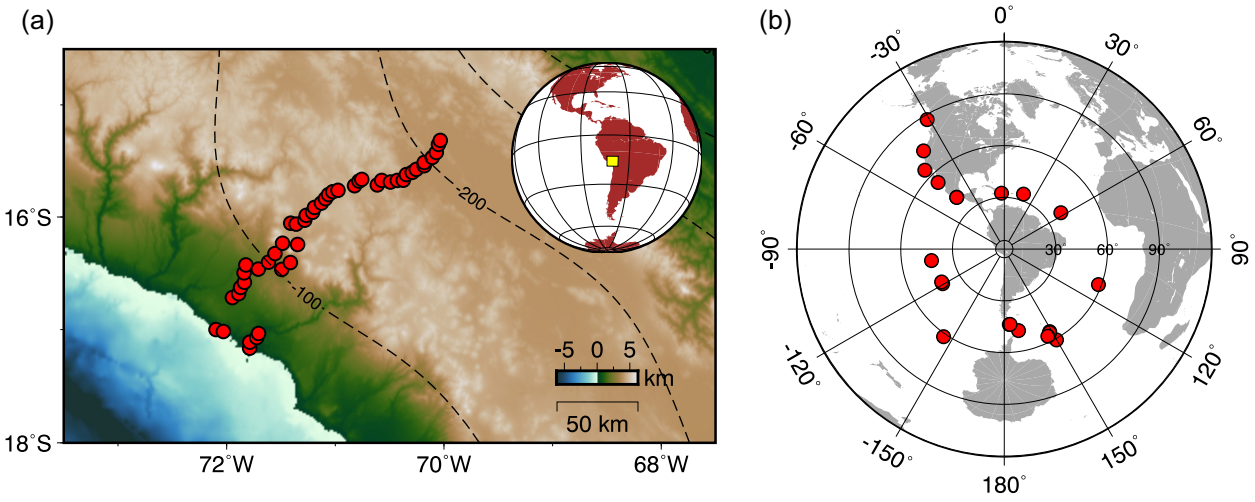

Figure 13. Geometry of the PERUSE deployment (a) and teleseismic sources (b) used in the real data inversion. The dashed lines show the depth contours of the Nazca plate taken from the Slab2 model (Hayes 2018).

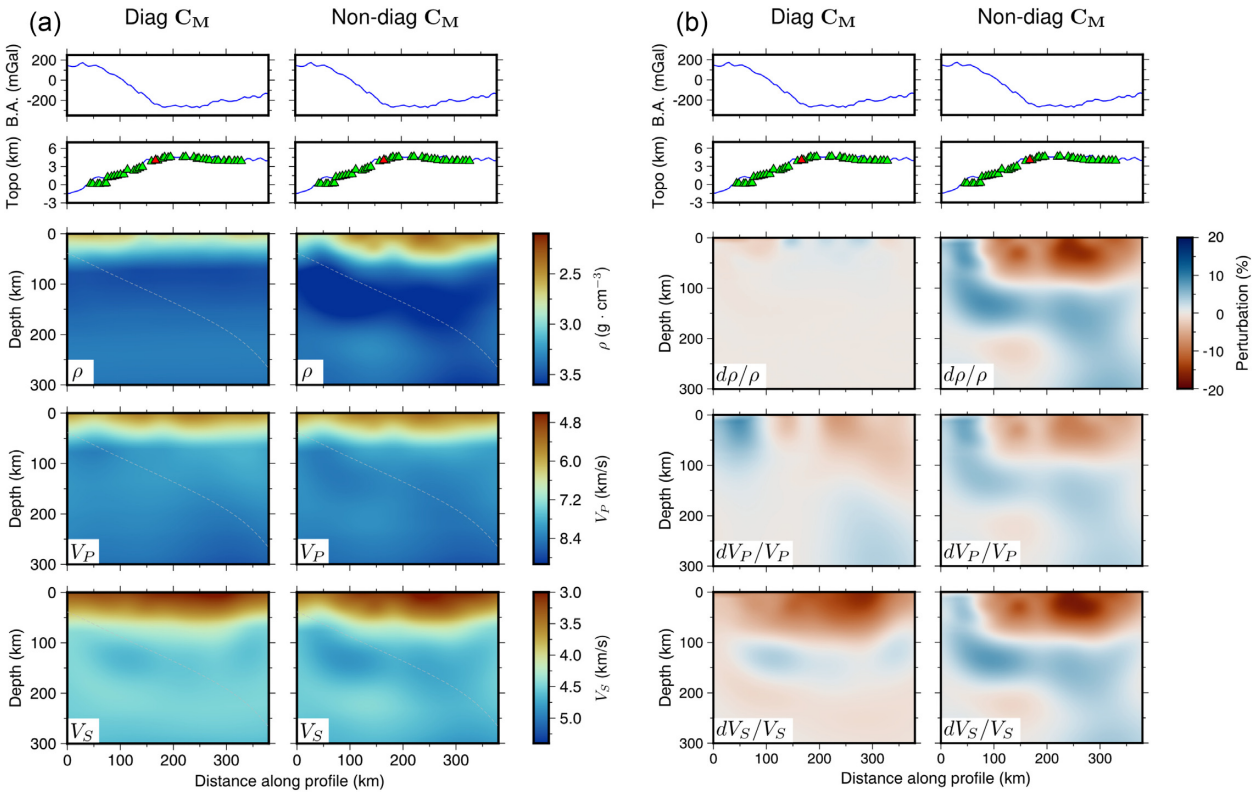

Figure 14. Absolute ρ, V_P and V_S models (a) obtained by FWI of 16 teleseismic events recorded by the profile PE of PeruSE array, with data filtered between 0.04 and 0.08 Hz. (b) Deviations of ρ, V_P and V_S from their average 1-D models. The green triangles represent the seismic stations whereas the red triangle indicates the location of the active volcanic arc. B.A. represents the Bouguer anomaly.

in the shallow part of the slab. The V_S is almost not affected. When σ_{V_P} is increased, we observe opposite effects on ρ and V_P. The V_P model shows stronger and sharper contrasts while the ρ model is smoother with anomalies of reduced amplitude. The use of strongly unbalanced standard deviations of the model parameters can thus lead to trade-offs between the model parameters, in particular between ρ and V_P.

So far, we have focused on our synthetic subduction model. It remains to be seen whether the hyperparameters we found suitable in synthetic inversion experiments can also be used in real waveform inversions, although we expect that this will indeed be the case. In any case, the use of a full non-diagonal matrix with spatial covariance kernels provides additional flexibility to regularize the inversion. With this approach, it is possible to freely adjust standard deviations, correlation lengths, and correlation coefficients between model parameters. Our approach also allows the implementation of model covariance matrices that can vary with the position in the model, and in particular with depth. This could provide additional flexibility to regularize FWI on large tomographic grids, extending to the transition zone or even the lower mantle.

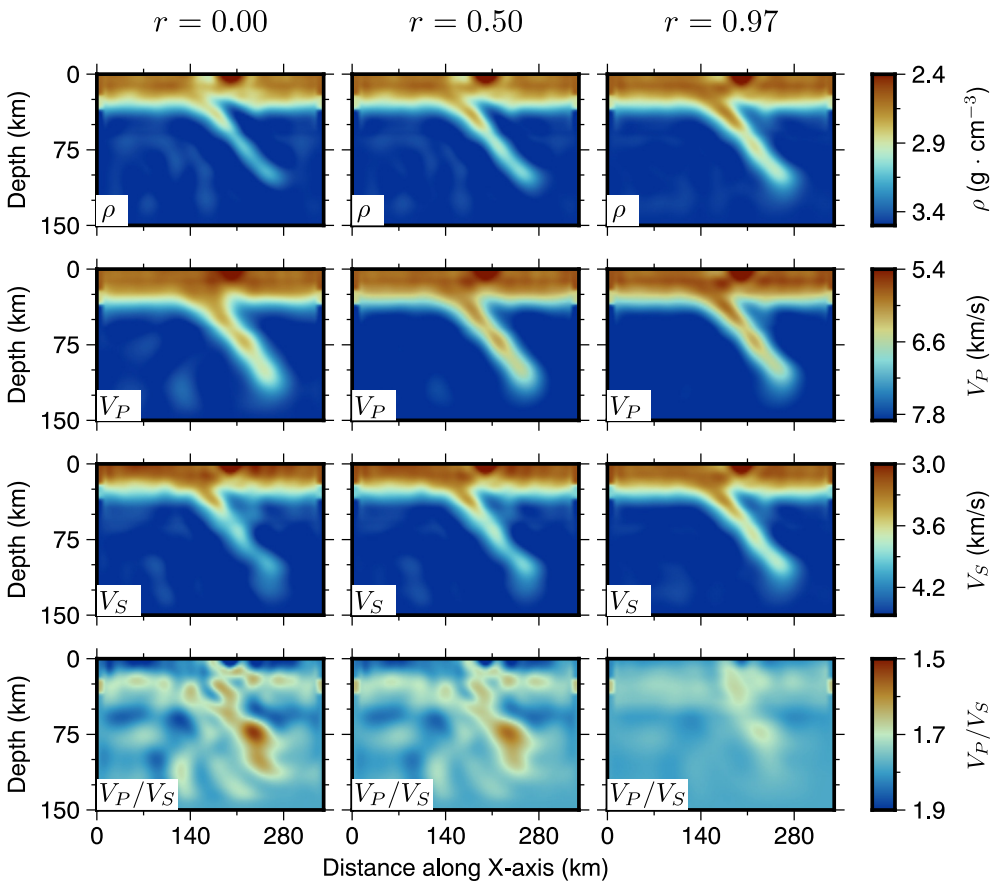

Figure 15. FWI models obtained with data filtered between 0.04 and 0.1 Hz, using correlation coefficients between model parameters varying from 0 (no correlation) to 0.97 (strong correlation).

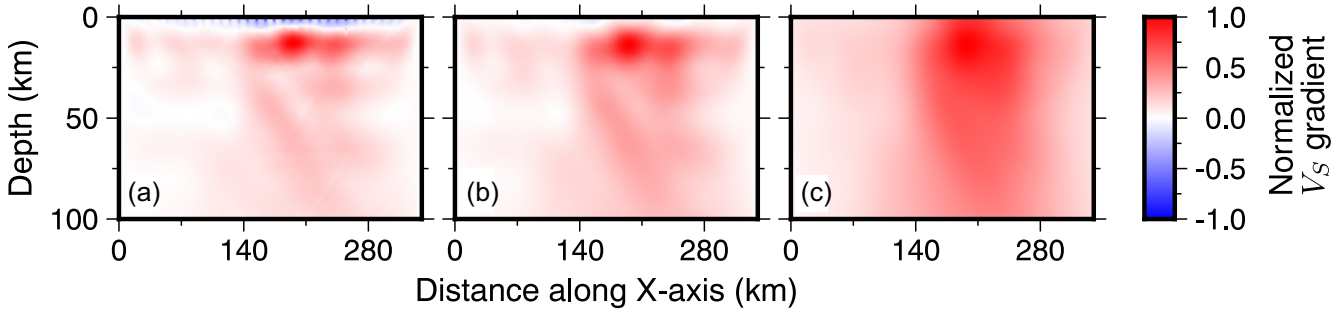

Figure 16. Normalized V_S gradient computed in the initial smooth model, with different smoothing lengths. (a) No smoothing. (b) Smoothing length of 5 km. (c) Smoothing length of 20 km. Note that the strong singularities beneath the stations when no spatial smoothing is applied in (a).

5.2 Joint inversions of *P* and *SH* events

As we have seen in the previous sections, accounting for the correlation between V_P and V_S in the inversions allowed us to improve the resolution in the deep parts of the V_S models. Fig. 18 compares the reconstructed models obtained by adding 4 *SH* events to the data set, with or without taking into account the correlations between model parameters. Fig. 19 presents the synthetic seismograms computed in the final models (black lines), and the seismograms computed in the initial (grey lines) and final (red lines for *P* waves and blue lines for *SH* waves) models. This comparison reveals that the *SH* waveforms are already quite well fitted by the model derived from the inversion of the *P* waveforms only, provided that a strong correlation between V_P and V_S is assumed. When *SH* waves are included in the inversion, the fits of *SH* waveforms are significantly improved, in particular for the multiples that arrive ∼15 s after the direct *SH* waves. We can first note that in the joint inversion of *P* and *S* waveforms, the reconstructed V_P and V_S models assuming a diagonal or non-diagonal model covariance matrix are very similar. Furthermore, these models are also remarkably similar to the models previously obtained from the inversion of *P* waveforms alone but using the full covariance matrix. These results therefore suggest that taking into account the correlation between V_P and V_S indeed allows us to reconstruct the V_S model even if we only invert *P* waveforms. However, the V_P/V_S ratio model is best recovered when both *P* and *S* waveforms are inverted and with a non-diagonal model covariance matrix.

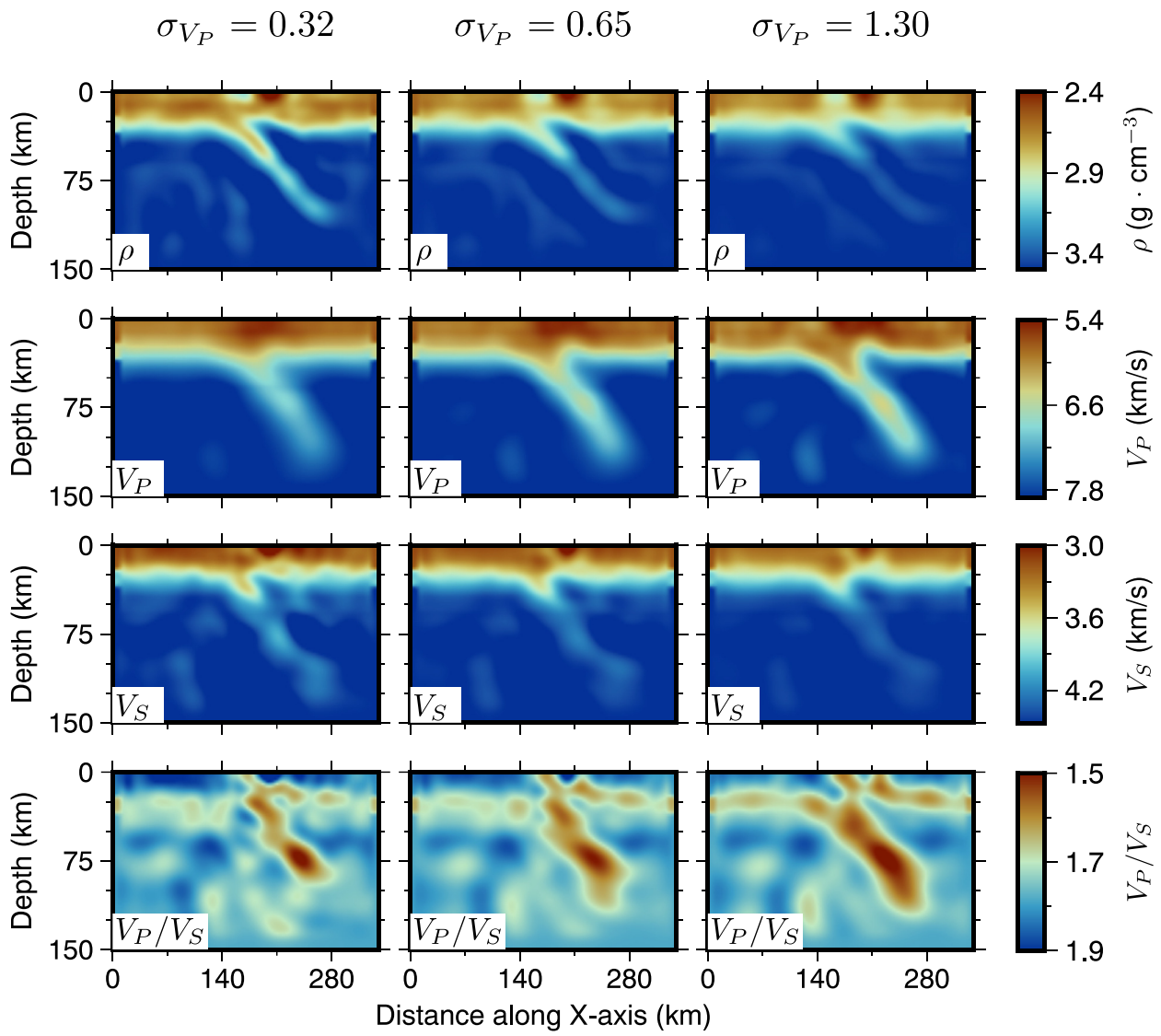

Figure 17. Comparison of long period FWI results for different σ_{V_P} (km s^{-1}). All the other hyperparameters are the same as in the reference inversion shown in Fig. 6.

Although these results seem to suggest that it may not be necessary to include shear waves to image V_S heterogeneities, the contribution of shear waves in waveform inversion is in fact far from negligible. First, shear waves provide an independent source of data that can be used to assess the quality of the model obtained by the inversion of teleseismic *P* waveforms. Second, the addition of shear waves will allow us to obtain more robust constraints on the V_P/V_S ratio, which would otherwise suffer from the contrast in spatial resolution of the V_P and V_S models. In addition to improving the spatial coverage, especially of deep and long wavelengths V_S heterogeneities, the combination of *S* waveforms with *P* waveforms is also essential to properly constrain seismic anisotropy, as demonstrated by Beller & Chevrot (2020).

6 CONCLUSIONS

We have presented a new implementation of Bayesian FWI that introduces *a priori* information through a full 3-D model covariance matrix. The spatial correlation function that describes the statistical spatial correlations of model parameters is an exponential kernel. Its inverse admits a simple semi-analytical expression and it can thus be easily applied to any vector model within the spectral-element framework that is used to simulate the wave propagation. With this formalism, it is also possible to account for the physical correlation between density and compressional and shear wave velocities by introducing off-diagonal terms in the model covariance matrix.

Synthetic inversion experiments demonstrate that with this new inversion method the parameter correlations compensate for the lower sensitivity of *P* waveforms to shear velocities and the final shear velocity model is on par with the one that would have been obtained by jointly inverting *P* and *S* waveforms. The improvement on the reconstruction of the density is even more spectacular. The reconstruction of the V_P/V_S ratio, a key parameter to constrain the composition and thermal state of the lithosphere, is also greatly improved. Whereas the models obtained by assuming the model parameters are uncorrelated, that is assuming a diagonal model covariance matrix, are contaminated by significant artefacts, most of these artefacts disappear when the correlations between model parameters are accounted for during the inversion. In addition, complete non-diagonal model covariance matrices reduce the number of

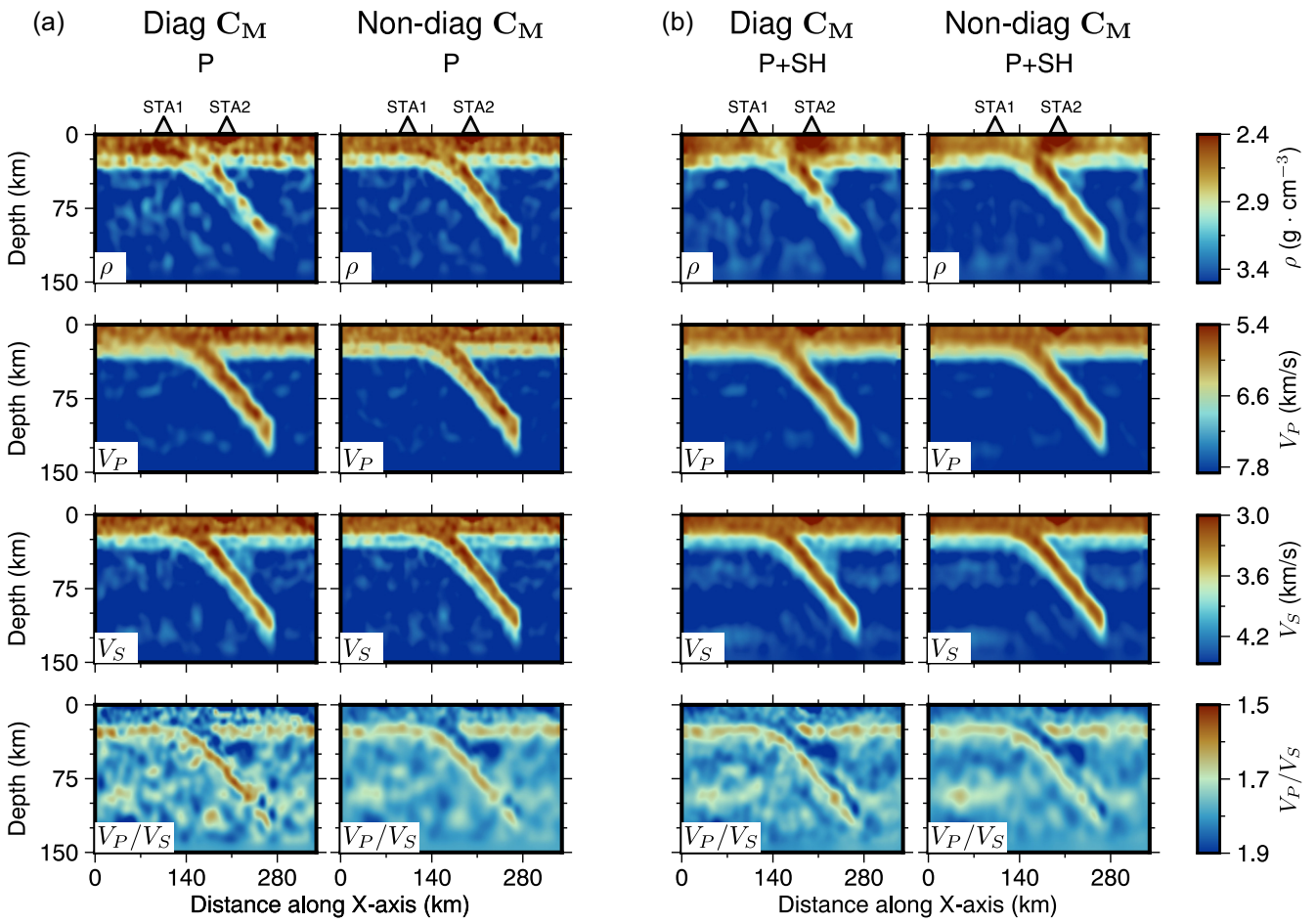

Figure 18. Comparison of the results obtained after multiscale (0.04–0.40 Hz frequency bandwidth) FWI considering *P* waveforms only (a) and both *P* and *SH* waveforms (b), with a diagonal model covariance matrix or with a non-diagonal model covariance matrix. The multiscale inversion strategies are the same as in Figs 6 and 9. The corresponding waveforms for STA1 and STA2 are shown in Fig. 19.

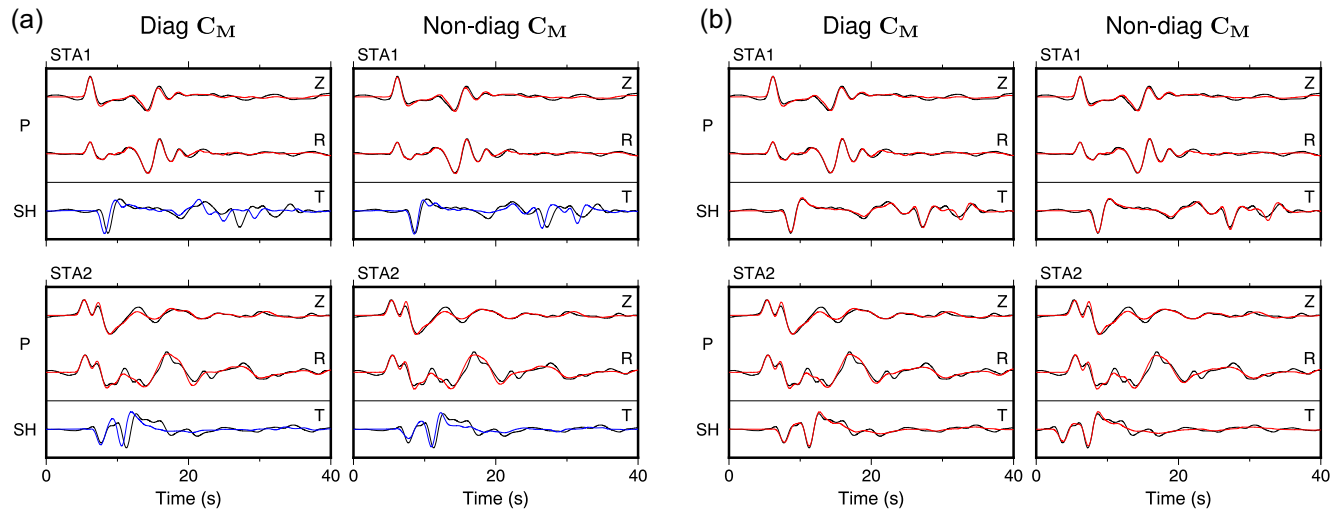

Figure 19. Waveform fits for the event with a backazimuth of 70°. (a) Waveform fits obtained by inverting *P* waveforms considering a diagonal and a non-diagonal model covariance matrix. (b) Waveform fits obtained by inverting both *P* and *SH* waveforms considering a diagonal and a non-diagonal model covariance matrix. All the traces are bandpass filtered between 0.04 and 0.40 Hz. Black lines show the noisy data. Red traces represent the corresponding waveforms computed in the final model shown in Fig. 18. Blue lines represent the predicted *SH* waveforms when only *P* waves are included in the inversion. The position of stations STA1 and STA2 are marked in Fig. 18.

degrees of freedom of the inverse problem and thus the size of the model space to explore. The convergence of the inversion algorithm is consequently faster.

Future work will be devoted to applying this new approach to real teleseismic waveforms recorded by dense regional arrays. The extension of the Bayesian FWI approach described here to general anisotropic media will be presented in a future contribution.

ACKNOWLEDGMENTS

We thank the two anonymous reviewers for their constructive and thoughtful comments that helped us to improve the manuscript. VM was supported by the European Union's Horizon 2020 - Research and Innovation Framework Programme under the ChEESE project, grant agreement No. 823844. This work was granted access to the HPC resources of CALMIP supercomputing center under the allocation P20012 and the HPC resources of IDRIS under the allocation 2021-101085 made by GENCI. We also acknowledge PRACE for awarding us access to Marconi100 at CINECA, Italy.

DATA AVAILABILITY

The FWI code is available upon request to the corresponding author, and it will be released to the public in the near future. The waveform data of station MLS from the French RLBP network and of the profile PE in PERUSE are available from IRIS-DMC.

REFERENCES

Araujo, S., Valette, B., Potin, B. & Ruiz, M., 2021. A preliminary seismic travel time tomography beneath ecuador from data of the national network, *J. South Am. Earth Sci.,* **111,** 103486, doi:10.1016/j.jsames.2021.103486.

Babuška, V. & Cara, M., 1991. *Seismic Anisotropy of the Earth,* Kluwer Academic.

Becker, T.W. & O'Connell, R.J., 2001. Predicting plate velocities with mantle circulation models, *Geochem. Geophys. Geosyst.,* **2**(12), 2001GC000171, doi:10.1029/2001gc000171.

Beller, S. & Chevrot, S., 2020. Probing depth and lateral variations of upper-mantle seismic anisotropy from full-waveform inversion of teleseismic body-waves, *Geophys. J. Int.,* **222,** 352–387.

Beller, S., Monteiller, V., Operto, S., Nolet, G., Paul, A. & Zhao, L., 2017. Lithospheric architecture of the south-western Alps revealed by multi-parameter teleseismic full-waveform inversion, *Geophys. J. Int.,* **212**(2), 1369–1388.

Birch, F., 1961. The velocity of compressional waves in rocks to 10 kilobars, part 2, *J. geophys. Res.,* **66**(7), 2199–2224.

Bolton, H. & Masters, G., 2001. Travel times of P and S from the global digital seismic networks: implications for the relative variation of P and S velocity in the mantle, *J. geophys. Res.,* **106**(B7), 13527–13540.

Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. seism. Soc. Am.,* **95**(6), 2081–2092.

Cammarano, F., Goes, S., Vacher, P. & Giardini, D., 2003. Inferring upper-mantle temperatures from seismic velocities, *Phys. Earth planet. Inter.,* **138**(3-4), 197–222.

Chevrot, S. & van der Hilst, R., 2000. The Poisson ratio of the Australian crust: geological and geophysical implications, *Earth planet. Sci. Lett.,* **183,** 121–132.

Christensen, N.I., 1996. Poisson's ratio and crustal seismology, *J. geophys. Res.,* **101,** 3139–3156.

Clayton, R., 2013. Peru subduction experiment, Caltech. Dataset. doi:10.7909/C3H41PBZ.

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

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A. & Virieux, J., 2013. Which parameterization is suitable for acoustic vertical transverse isotropic full waveform inversion? Part 1: sensitivity and trade-off analysis, *Geophysics,* **78,** R81–R105.

Hayes, G., Moore, G.L.,Portner, D.E.,Hearne, M., Flamme, H., Furtney, M. & Smoczyk, G.M., 2018. Slab2 - a comprehensive subduction zone geometry model, *Science,* **362,** 58–61.

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

Lee, C. T.A., 2003. Compositional variation of density and seismic velocities in natural peridotites at STP conditions: implications for seismic imaging of compositional heterogeneities in the upper mantle, *J. geophys. Res.,* **108**(B9), Art No. 2441, doi:10.1029/2003JB002413.

Lees, J.M. & VanDecar, J.C., 1991. Seismic tomography constrained by Bouguer gravity anomalies: applications in western Washington, *Pure appl. Geophys.,* **135**(1), 31–52.

Lines, L.R., Schultz, A.K. & Treitel, S., 1988. Cooperative inversion of geophysical data, *Geophysics,* **53**(1), 8–20.

Martin, R., Giraud, J., Ogarko, V., Chevrot, S., Beller, S., Gégout, P. & Jessell, M., 2021. Three-dimensional gravity anomaly data inversion in the Pyrenees using compressional seismic velocity model as structural similarity constraints, *Geophys. J. Int.,* **225**(2), 1063–1085.

Monteiller, V., Beller, S., Plazolles, B. & Chevrot, S., 2021. On the validity of the planar wave approximation to compute synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.,* **224**(3), 2060–2076.

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

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

Nafe, J.E. & Drake, C.L., 1957. Variation with depth in shallow and deep water marine sediments of porosity, density and the velocities of compressional and shear waves, *Geophysics,* **22**(3), 523–552.

Nataf, H.-C., Nakanishi, I. & Anderson, D.L., 1986. Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy: 3. Inversion, *J. geophys. Res.,* **91**(B7), 7261–7307.

Nocedal, J. & Wright, S., 2006. *Numerical Optimization,* Springer.

Oliver, D.S., 1998. Calculation of the inverse of the covariance, *Math. Geol.,* **30**(7), 911–933.

Operto, S., Gholami, S., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L. & Virieux, J., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice, *Leading Edge,* **32,** 1040–1054.

Phillips, K. *et al.*, 2012. Structure of the subduction system in southern Peru from seismic array data, *J. geophys. Res.,* **117**(B11), doi:10.1029/2012jb009540.

Potin, B., 2016. Les Alpes occidentales: tomographie, localisation de séismes et topographie du Moho, *PhD thesis*, Université Grenoble Alpes.

RESIF, 1995. RESIF-RLBP French Broad-band network, RESIF-RAP strong motion network and other seismic stations in metropolitan France. RESIF – Réseau Sismologique et géodésique Français. , doi:10.15778/resif.fr.

Robertson, G.S. & Woodhouse, J.H., 1997. Comparison of *P* and *S* station corrections and their relationship to upper mantle structure, *J. geophys. Res.,* **102,** 27 355–27 366.

Savard, G., Bostock, M.G. & Christensen, N.I., 2018. Seismicity, metamorphism, and fluid evolution across the northern Cascadia fore arc, *Geochem. Geophys. Geosyst.,* **19**(6), 1881–1897.

Simmons, N., Forte, A. & Grand, S., 2006. Constraining mantle flow with seismic and geodynamic data: a joint approach, *Earth planet. Sci. Lett.,* **246**(1–2), 109–124.

Tarantola, A., 2005. *Inverse Problem Theory,* SIAM.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.,* **20**(2), 219–232.

Thurber, C. & Eberhart-Phillips, D., 1999. Local earthquake tomography with flexible gridding, *Comput. Geosci.,* **25**(7), 809–818.

Thurber, C.H., Atre, S.R. & Eberhart-Phillips, D., 1995. Three-dimensional vp and vp/vs structure at Loma Prieta, California, from local earthquake tomography, *Geophys. Res. Lett.,* **22**(22), 3079–3082.

Trinh, P., Brossier, R., Métivier, L., Virieux, J. & Wellington, P., 2017. Bessel smoothing filter for spectral-element mesh, *Geophys. J. Int.,* **209**(3), 1489–1512.

Tromp, J., Tape, C. & Liu, Q.Y., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

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

Watanabe, T., 1993. Effects of water and melt on seismic velocities and their application to characterization of seismic reflectors, *Geophys. Res. Lett.,* **20,** 2933–2936.

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

APPENDIX A: DECOMPOSITION OF THE MODEL COVARIANCE MATRIX

We consider a model covariance matrix of the form

$$\mathbf{C_M} = \begin{bmatrix} \sigma_\rho^2\mathbf{C} & r\sigma_\rho\sigma_{V_\mathrm{P}}\mathbf{C} & r\sigma_\rho\sigma_{V_\mathrm{S}}\mathbf{C} \\ r\sigma_\rho\sigma_{V_\mathrm{P}}\mathbf{C} & \sigma_{V_\mathrm{P}}^2\mathbf{C} & r\sigma_{V_\mathrm{P}}\sigma_{V_\mathrm{S}}\mathbf{C} \\ r\sigma_\rho\sigma_{V_\mathrm{S}}\mathbf{C} & r\sigma_{V_\mathrm{P}}\sigma_{V_\mathrm{S}}\mathbf{C} & \sigma_{V_\mathrm{S}}^2\mathbf{C} \end{bmatrix} \tag{A1}$$

where r is the correlation coefficient, $\mathbf{C}$ is the spatial covariance kernel, σ_ρ is the standard deviation of density, σ_{V_P} is the standard deviation of V_P and σ_{V_S} is the standard deviation of V_S. Note that we assume here for simplicity that the correlation coefficients between each pair of parameters are equal.

The model covariance matrix can be rewritten as

$$\mathbf{C_M} = \Sigma\mathbf{S}\mathbf{R}\mathbf{S}\Sigma, \tag{A2}$$

where

$$\Sigma = \begin{bmatrix} \sigma_\rho & 0 & 0 \\ 0 & \sigma_{V_\mathrm{P}} & 0 \\ 0 & 0 & \sigma_{V_\mathrm{S}} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{C}^{\frac{1}{2}} & 0 & 0 \\ 0 & \mathbf{C}^{\frac{1}{2}} & 0 \\ 0 & 0 & \mathbf{C}^{\frac{1}{2}} \end{bmatrix}, \tag{A3}$$

and

$$\mathbf{R} = \begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}. \tag{A4}$$

The eigenvalues λ of $\mathbf{R}$ are solutions of

$$|\Theta - \lambda\mathbf{I}| = (1-\lambda)^3 - 3r^2(1-\lambda) + 2r^3 = 0. \tag{A5}$$

The third-order polynomial in $x = 1 - \lambda$ can be factorized as

$$x^3 - 3r^2x + 2r^3 = (x+2r)(x-r)^2. \tag{A6}$$

The first eigenvalue of $\mathbf{R}$ is thus $\lambda_1 = 1 + 2r$ and the first eigenvector

$$\mathbf{v_1} = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}. \tag{A7}$$

The second and third eigenvalues are $\lambda_2 = \lambda_3 = 1 - r$ and

$$\mathbf{v_2} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \quad \mathbf{v_3} = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{bmatrix}. \tag{A8}$$

The singular value decomposition of $\mathbf{R}$ is

$$\mathbf{R} = \mathbf{V}\Lambda\mathbf{V}^t, \tag{A9}$$

with

$$\mathbf{V} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \text{and } \Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \tag{A10}$$

The model covariance matrix can therefore be decomposed into the product

$$\mathbf{C_M} = \Sigma\mathbf{S}\mathbf{V}\Lambda\mathbf{V}^t\mathbf{S}\Sigma. \tag{A11}$$

This decomposition provides some insights into the anatomy of the physical constraints that are imposed on the model parameters. The matrix Σ defines a metric that normalizes the perturbations of model parameters. It makes the inversion insensitive to the units chosen and equalizes the contribution of each class of parameter to the gradient of the misfit function. The matrix $\mathbf{V}$ can be seen as a transformation matrix that rotates the covariance matrix along its principal axes. The first axis, oriented along $\mathbf{v_1}$, describes the principal linear relationship between the different parameters. The second and third eigenvectors describe the deviations of model perturbations with respect to this linear relationship. When $r \to 1$, $\lambda_1 \to 3$, and $\lambda_{2,3} \to 0$. The three parameters are perfectly correlated and described by their position along the line oriented along $\mathbf{v_1}$. When $r \to 0$, $\lambda_{1,2,3} \to 1$, variations of the parameters are completely random and the model covariance matrix is diagonal. This is the implicit assumption that has been made so far in most multiparameter tomographic inversions. The matrix $\mathbf{S}$ describes how model parameters vary as a function of the distance between two points and can thus be seen as a smoothing operator.

APPENDIX B: CHANGE OF MODEL PARAMETRIZATION AND TRANSFORMATION OF THE MODEL COVARIANCE MATRIX

We illustrate the transformation of the model covariance matrix after a change of model parametrization. We first consider a diagonal covariance matrix for model parametrization (ρ, V_P, V_S) with mean $\overline{V}_\mathrm{P} = 7\ \mathrm{km\,s^{-1}}$ and $\overline{V}_\mathrm{S} = 4\ \mathrm{km\,s^{-1}}$, $\sigma_{V_\mathrm{P}} = 0.65\ \mathrm{km\,s^{-1}}$ and $\sigma_{V_\mathrm{S}} = 0.37\ \mathrm{km\,s^{-1}}$. We draw 2000 samples from the bivariate probability density function

$$f(V_\mathrm{P}, V_\mathrm{S}) = \frac{1}{2\pi\sigma_{V_\mathrm{P}}\sigma_{V_\mathrm{S}}\sqrt{1-r^2}} \cdot \exp\left(-\frac{1}{2(1-r^2)}\left[\left(\frac{V_\mathrm{P}-\overline{V}_\mathrm{P}}{\sigma_{V_\mathrm{P}}}\right)^2 + \left(\frac{V_\mathrm{S}-\overline{V}_\mathrm{S}}{\sigma_{V_\mathrm{S}}}\right)^2 - 2r\left(\frac{V_\mathrm{P}-\overline{V}_\mathrm{P}}{\sigma_{V_\mathrm{P}}}\right)\left(\frac{V_\mathrm{S}-\overline{V}_\mathrm{S}}{\sigma_{V_\mathrm{S}}}\right)\right]\right), \tag{B1}$$

where r is the correlation coefficient between V_P and V_S. The scatter plot showing the distribution of these samples is displayed in Fig. B1(a). Since ρ is independent of either V_P, V_S or $V_\mathrm{P}/V_\mathrm{S}$, the scatter plots with ρ are not shown. The same distribution after conversion to the (V_P, $V_\mathrm{P}/V_\mathrm{S}$) parametrization is plotted in Fig. B1(b). Using eq. (19), we can derive the transformed covariance matrix for the (V_P, $V_\mathrm{P}/V_\mathrm{S}$) parametrization. We draw a new set of 2000 samples that follow the distribution described by this transformed covariance matrix and plot them in Fig. B1(c). As expected, the distributions shown in Figs B1(b) and (c) are very similar.

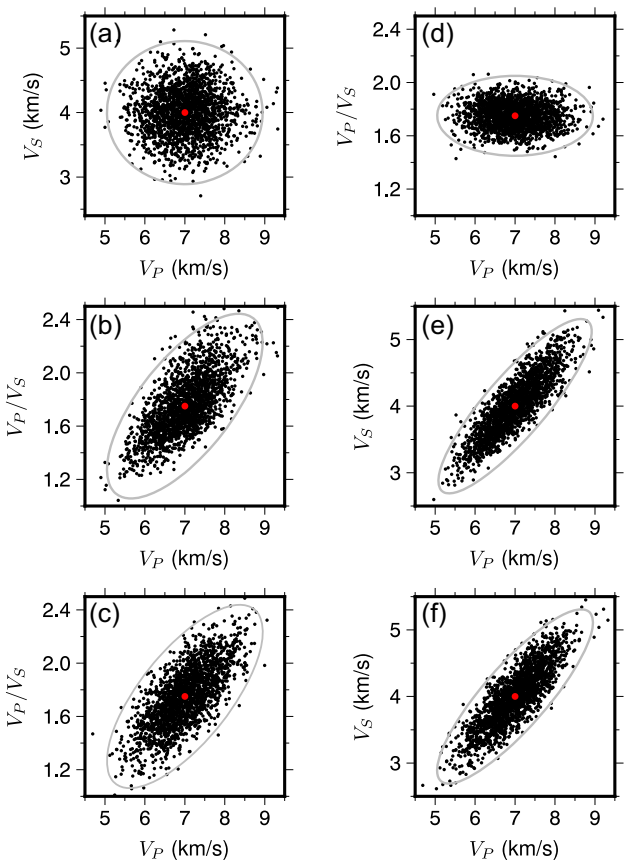

Figure B1. The consistency of bivariate normal distributions over different parametrizations, with the covariance matrix properly transformed. (a) Bivariate normal distribution of V_P and V_S, generated with a diagonal covariance matrix. (b) Same distribution of (a) but plotted in V_P and V_P/V_S. (c) Bivariate normal distribution of V_P and V_P/V_S, generated with the transformed covariance matrix. (d) Bivariate normal distribution of V_P and V_P/V_S, generated with a diagonal covariance matrix. (e) Same distribution of (d) but plotted in V_P and V_S. (f) Bivariate normal distribution of V_P and V_S, generated with the transformed covariance matrix. The grey lines correspond to three standard deviations.

We now consider the diagonal covariance matrix

$$\mathbf{C}_{(\rho,V_\mathrm{P},\frac{V_\mathrm{P}}{V_\mathrm{S}})} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_\mathrm{P}}^2 & 0 \\ 0 & 0 & \sigma_{\frac{V_\mathrm{P}}{V_\mathrm{S}}}^2 \end{bmatrix} \tag{B2}$$

for the (ρ, V_P, $V_\mathrm{P}/V_\mathrm{S}$) parametrization, which transforms to

$$\mathbf{C}_{(\rho,V_\mathrm{P},V_\mathrm{S})} = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_{V_\mathrm{P}}^2 & \frac{V_\mathrm{S}}{V_\mathrm{P}}\sigma_{V_\mathrm{P}}^2 \\ 0 & \frac{V_\mathrm{S}}{V_\mathrm{P}}\sigma_{V_\mathrm{P}}^2 & \frac{V_\mathrm{S}^2}{V_\mathrm{P}^2}\sigma_{V_\mathrm{P}}^2 + \frac{V_\mathrm{S}^4}{V_\mathrm{P}^2}\sigma_{\frac{V_\mathrm{P}}{V_\mathrm{S}}}^2 \end{bmatrix} \tag{B3}$$

for the (ρ, V_P, V_S) parametrization.

We draw 2000 samples from the bivariate normal distribution (B2) using $\overline{V}_\mathrm{P} = 7$ km s^{-1}, $\overline{\frac{V_\mathrm{P}}{V_\mathrm{S}}} = 1.75$, $\sigma_{V_\mathrm{P}} = 0.65$ km s^{-1} and $\sigma_{\frac{V_\mathrm{P}}{V_\mathrm{S}}} = 0.1$. The resulting scatter plot is shown in Fig. B1(d). The same distribution after conversion to the (V_P, V_S) parametrization is plotted in Fig. B1(e). Finally, the distribution obtained by sampling the bivariate distribution (B3) is shown in Fig. B1(f). Again, the two distributions in Figs B1(e) and (f) are similar. These two examples emphasize the importance of the non-diagonal terms of the model covariance matrix when the model parametrization is changed.