Dimensionality Reduction in Full-Waveform Inversion Uncertainty Analysis

Mulder, W. A.; Kuvshinov, B. N.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ORIGINAL ARTICLE OPEN ACCESS

Dimensionality Reduction in Full-Waveform Inversion Uncertainty Analysis

W. A. Mulder[1,2] | B. N. Kuvshinov[1]

[1]Shell Global Solutions International B.V., Den Haag, The Netherlands | [2]Department of Geoscience & Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

Correspondence: W. A. Mulder (w.a.mulder@tudelft.nl)

ABSTRACT

The uncertainty of model parameters obtained by full-waveform inversion can be determined from the Hessian of the least-squares error functional. A description of uncertainty characterisation is presented that takes the null space of the Hessian into account and does not rely on the Bayesian formulation. Because the Hessian is generally too costly to compute and too large to be stored, a segmented representation of perturbations of the reconstructed subsurface model in the form of geological units is proposed. This enables the computation of the Hessian and the related covariance matrix on a larger length scale. Synthetic two-dimensional isotropic elastic examples illustrate how conditional and marginal uncertainties can be estimated for the properties per geological unit by themselves and in relation to other units.

1 | Introduction

Subsurface model reconstruction from seismic data with full-waveform inversion (FWI) has become a routine approach. To characterise the uncertainty of the model parameters, the classic approach involves a series expansion around the, hopefully, global minimum of the least-squares data misfit functional or one of its many variants (Backus and Gilbert 1970; Tarantola 2005). The locally quadratic cost function around the minimum is described by the Hessian, assuming the cost or loss function is sufficiently smooth. The uncertainty can be quantified as the region in model parameter space where this function stays below a certain threshold value. The latter is determined by the inversion accuracy and noise level of the data. The region is bounded by the ellipsoid where the paraboloid of the quadratic approximation cuts the threshold value. The principal axes of the ellipsoid are the eigenvectors of the Hessian, and the reciprocals of the lengths of the semi-axes are the square roots of the Hessian's eigenvalues.

A typical three-dimensional model for finite-difference modelling requires Gigabytes of storage and the associated Hessian Exabytes. Its direct computation is out of reach, except for smaller problems (Pratt et al. 1998) or small subsets of points (Hak and Mulder 2010; Mulder and Kuvshinov 2025; Plessix and Mulder 2004, for instance). There are many methods that find some approximation to the Hessian, for instance, by using the Lanczos algorithm (Minkoff 1996; Vasco et al. 2003), low-rank approximations (Bui-Thanh et al. 2012; Eckart and Young 1936; Liu and Peter 2019a; 2019b; Riffaud et al. 2024; H. Zhu et al. 2016) and Kalman filtering (Eikrem et al. 2019; Hoffmann et al. 2024; Huang et al. 2020; Thurin et al. 2019). Crude estimates can be obtained from methods that estimate 'true-amplitude' weights (Chen and Xie 2015; Rickett 2003; Riyanti et al. 2008,

for instance) or checkerboard tests (Inoue et al. 1990; Lévêque et al. 1993). Examples of approaches that estimate the uncertainty without using the Hessian are null-space shuttles (Deal and Nolet 1996; Fichtner and Zunino 2019; Keating and Innanen 2021) and their generalisations (Meju 2009; Vasco 2007), the Markov-chain Monte Carlo method (Barbosa et al. 2020; Ely et al. 2018; Fichtner and van Leeuwen 2015; Guo et al. 2020; Martin et al. 2012; Piana Agostinetti et al. 2015; Ray et al. 2017, among others), the Hamiltonian Monte Carlo method (Betancourt 2018; Duane et al. 1987; Fichtner and Zunino 2019; Revelo Obando 2018; Sen and Biswas 2017; Zhao and Sen 2021), variational inference (Biswas et al. 2023; Izzatullah et al. 2023; Liu and Wang 2016; Wang et al. 2023; Zhang and Curtis 2020), and machine learning (Qu et al. 2024; Rizzuti et al. 2020; Siahkoohi et al. 2023; Sun et al. 2021), to mention just a few. Reviews on the subject can be found in the paper by Rawlinson et al. (2014), geared towards tomography and in the references mentioned above.

The grid size in FWI modelling is always smaller than the characteristic wavelength, typically 4–5 grid cells per wavelength. If a finite-difference scheme on a uniform grid is used, this value is attained in parts of the model where the (shear) velocity is smallest and much larger values are reached elsewhere. As a result, FWI is inherently uncertain at the grid-spacing scale and small-scale perturbations of model parameters typically fall within the null space of the full Hessian, having a negligibly small influence on the cost function. For this reason, the use of the full Hessian for perturbations of each model parameter in each point of a finite-difference grid is impractical, even if such a Hessian can be calculated.

To quantify uncertainty, we must assess it at larger scales. If we do not eliminate small scales, the uncertainty becomes formally infinite. However, filtering out small-scale variations makes the analysis subjective, because the final result depends on which scales we choose to eliminate. Thus, the question 'what is FWI uncertainty?' can only be meaningfully answered in a relative sense and specifically as FWI uncertainty with respect to a chosen class of perturbations. This issue is evident in the description of FWI uncertainty in terms of the associated Hessian matrix. The Hessian is singular or nearly singular, with eigenvalues spanning many orders of magnitude. The uncertainty is governed by the smallest eigenvalues, which correspond to poorly resolved features. When moving to larger scales, we effectively eliminate small eigenvalues, reducing the uncertainty. However, the final result depends on where we set the threshold. This reinforces the fundamental point: FWI uncertainty is not a single well-defined quantity but must always be considered relative to the scale and nature of perturbations under investigation.

One approach to specify larger subsurface blocks is to segment the model, resulting from FWI, into geological units that define parts of the subsurface of a similar rock type. The model subspace consists of model parameters that are defined as perturbations of the original model and are, in the simplest case, piecewise constant within each unit. This restriction to a lower-dimensional space results in a compression of the Hessian with a smaller size than the original one. The size of a unit influences the uncertainty estimates, and we examine its effect by a number of examples.

From a mathematical point of view, Poincaré's separation theorem (Gradshteyn and Ryzhik 2000, for instance) describes the relation between the eigenvalues of the Hessian before and after projection. Here, the projection is similar to the compression but mapped back to the original space. In this way, the projection operator still acts on the same subspace but is defined as a map from the original space to itself. In the numerical examples, the projection replaces perturbations by their average inside a unit and acts as a spatial high-cut filter removing shorter wavelengths from the model. Our approach is similar to a common operation in the multigrid method, originally designed as an optimal numerical algorithm for solving elliptic partial differential equations by using a sequence of discretisations on grids with different scales (Hackbusch 1985; Mulder 2021). It also bears some similarity to spectral coarse graining of graphs (Gfeller and Rios 2007). Our method involves dimensionality reduction and, in that sense, is similar to other approximation methods, referenced earlier, that do not construct the full Hessian.

Section 2 reviews the basics of the Hessian computation and uncertainty estimation. We treat the minimisation problem as a projection of the observed data on the hypersurface formed by the model range in data space. With a proper choice of the weight in the cost functional, the covariance matrix characterising uncertainties induced by noisy data is proportional to the pseudo-inverse of Hessian. The covariance matrix characterising uncertainties that appear due to imperfect minimisation is proportional to the *square* of the pseudo-inverse of Hessian. Both types of uncertainties can be analysed in the same way, and without loss of generality, we consider uncertainties of the former type only. The Hessian is calculated in the Gauss–Newton approximation. Appendix A explains how to find the Hessian using the adjoint-state method. Appendix B explains how to combine Hessian in the case where one has several independent datasets.

Our approach is different from the standard one, described by, for instance, Tarantola (2005), in two respects. First, we do not rely on the Bayesian formalism nor do we specify a particular shape of the data noise distribution, such as Gaussian. All calculations are made in terms of covariance matrices. This approach is motivated by the fact that, in practice, we want to know the region of uncertainty inside which the model parameters lie within a specified level of confidence, instead of the exact probability for model parameters to have certain prescribed values – which cannot be reliably evaluated anyway. The geometry of the confidence ellipsoid, which represents the uncertainty range, is primarily influenced by the covariance matrix derived from the noise distribution rather than the noise distribution itself. Secondly, our derivations are more general, compared to what can be found in the geophysical literature. The matrices we are dealing with are intrinsically singular, and we therefore use the Moore–Penrose pseudo-inverse rather than the inverse.

The intersection of the confidence ellipsoid with a hyperplane in the model-parameter space provides a lower-dimensional ellipsoid. This smaller ellipsoid is described by the compressed Hessian, which acts only on those components of model vectors that lie in the hyperplane. The compressed Hessian can be constructed by partitioning the original Hessian into two parts, as described in Appendix C. The construction of the compressed

Hessian in the general case requires an operator that maps the solution from a fine to a coarse grid. As already mentioned, such an operator is called restriction, and the reverse operator is called prolongation. Section 3 explains how to choose the restriction operator, depending on the desired grouping of the model parameters. In particular, restriction operators can perform averaging over a set of model parameters, corresponding to high-cut filtering in the wavenumber domain.

In Section 4, we consider a series of two-dimensional models, starting with a 2D homogeneous acoustic model for which the Hessian can be found analytically, a 2D isotropic elastic model with a horizontally layered model with a numerically computed Hessian, and another 2D isotropic elastic model. In the examples, we consider a constant relative perturbation of each model parameter prescribed per geological unit. However, the background model obtained by FWI does not have to be constant inside the unit. Also, in the case of finite-difference modelling, grid points belonging to the same unit might be disconnected, even when the unit itself is a connected set. Its interior grid points may still exhibit gaps near, for instance, dipped sharp pinch-outs. We examine the effect of the projection on the null-space components and overall uncertainty estimates, both for the conditional and marginal cases.

2 | Framework to Quantify Full-Waveform Inversion Uncertainties

We will analyse the full-waveform inversion (FWI) uncertainty based on the reconstructed model, assuming that FWI has converged to the global minimum, although in the numerical examples, we will instead use synthetic models created for the occasion. The uncertainty of a model parameter is evaluated from the condition that the norm of the perturbations in the modelled data due to model perturbations is the same as the norm of the expected noise in the observed data. This norm of the perturbed modelled data, which by themselves are supposed to be free of noise apart from unavoidable numerical noise, has a quadratic dependence on the model perturbations, and it is characterised by the Hessian. We review the role of the Hessian in uncertainty analysis and explain its relation with the covariance matrix. We then introduce the confidence ellipsoid and explain the geometrical meaning of the conditional and marginal uncertainties. Conditional uncertainties follow from subsets of the Hessian and marginal uncertainties from subsets of its pseudo-inverse, the covariance matrix.

2.1 | Sources of Full-Waveform Inversion Inaccuracies

FWI reconstructs a subsurface model parameterised by a vector $\mathbf{m}$ by minimising a cost functional

$$\mathcal{X}(\mathbf{m}) = \frac{1}{2} \|\mathbf{u}(\mathbf{m}) - \mathbf{d}_{obs}\|_{\mathbf{W}}^2, \tag{1}$$

measuring the difference between modelled data $\mathbf{u}(\mathbf{m})$ and observed data $\mathbf{d}_{obs}$. Here, $\mathbf{S}_r$ is an operator that samples the wavefield $\mathbf{u}$ at the receiver locations for each shot. The modelled
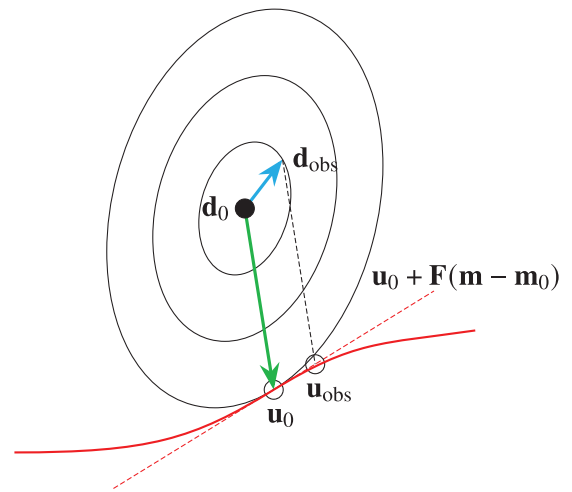
Figure 1 | Ellipses show points in the data space that are equidistant from noiseless data $\mathbf{d}_0$ with respect to some metric $\mathbf{W}$. The cost functional (1) with noiseless observed data $\mathbf{d}_0$ is minimised by projecting $\mathbf{d}_0$ on the range of the forward operator $\mathbf{u}$, which gives the least-squares solution $\mathbf{u}_0$ of the inversion problem for model parameters $\mathbf{m}_0$. The ellipse passing through $\mathbf{u}_0$ is tangent to the model hyperplane $\mathbf{u}$. If the observed data $\mathbf{d}_{obs}$ contain noise, the solution of the minimisation problem shifts to the point $\mathbf{u}_{obs}$. That point can be estimated by projecting $\mathbf{d}_{obs}$ on the linearised model range, which is shown by the red dashed line and is given by $\mathbf{u} = \mathbf{u}_0 + \mathbf{F}(\mathbf{m} - \mathbf{m}_0)$, with $\mathbf{F} = \nabla_{\mathbf{m}} \mathbf{u}(\mathbf{m}_0)$ the Fréchet derivative of the data modelling operator.

data $\mathbf{u}$ are found by solving a partial differential equation

$$\mathcal{L}(\mathbf{v}, \mathbf{m}) = \mathbf{f} \tag{2}$$

for $\mathbf{v}(\mathbf{m})$, for a given model $\mathbf{m}$ and source $\mathbf{f}$, and comparing $\mathbf{u} = \mathbf{S}_r \mathbf{v}$ with $\mathbf{d}_{obs}$ at the times and positions where the data were acquired, using the sampling operator $\mathbf{S}_r$. The L_2-norm in Equation (1) is defined by the inner product, $\|\mathbf{u}(\mathbf{m}) - \mathbf{d}_{obs}\|_{\mathbf{W}}^2 = [\mathbf{u}(\mathbf{m}) - \mathbf{d}_{obs}]^\top \mathbf{W}[\mathbf{u}(\mathbf{m}) - \mathbf{d}_{obs}]$, where $\mathbf{W}$ is a weight matrix, accounting for weighting in time, frequency, offset, depth, etc., and the superscript $(\cdot)^\top$ denotes the transpose. The weight matrix $\mathbf{W}$ is positive definite and plays the role of a metric tensor in the data space.

Let $\mathbf{m}_0$ be the parameters that represent the, hopefully, global minimum of the cost functional $\mathcal{X}$ in the absence of noise. The corresponding noiseless observed data are denoted by $\mathbf{d}_0$. Levels of constant $\frac{1}{2} \|\mathbf{d}_{obs} - \mathbf{d}_0\|_{\mathbf{W}}^2$ form a family of nested ellipsoids in the data space. The model range, that is, all possible values of $\mathbf{u}(\mathbf{m}) = \mathbf{S}_r \mathbf{v}(\mathbf{m})$ with $\mathbf{v}$ satisfying Equation (2), forms a hyperplane in the full data space. This hyperplane is tangent to one of the ellipsoids at the point $\mathbf{u}_0 = \mathbf{u}(\mathbf{m}_0)$ that is obtained by projecting $\mathbf{d}_0$ on the model range $\mathbf{u}(\mathbf{m})$. The remainder $\mathcal{X}(\mathbf{m}_0) = \frac{1}{2} \|\mathbf{u}(\mathbf{m}_0) - \mathbf{d}_0\|_{\mathbf{W}}^2$ involves data that the modelling operator $\mathbf{u}(\mathbf{m})$ cannot explain. The inverted model parameters might be different from $\mathbf{m}_0$ due to reconstruction errors. Another reason is that actual observed data $\mathbf{d}_{obs}$ contain ambient and instrumental noise, and its projection $\mathbf{u}_{obs}$ on the model range is not the same as $\mathbf{u}_0$ as Figure 1 illustrates.

2.2 | Hessian and Covariance

In what follows, we consider the inversion problem in the vicinity of $\mathbf{u}_0$, where the model can be linearised as $\mathbf{u} = \mathbf{u}_0 + \mathbf{F}(\mathbf{m} - \mathbf{m}_0)$. Here, $\mathbf{F} = \nabla_{\mathbf{m}}\mathbf{u}(\mathbf{m}_0)$ is the Fréchet derivative of the modelling operator, which can be computed by perturbing model parameters individually and recording the corresponding data perturbations for all shots and receivers. The cost functional and its derivative expand in the Taylor series around $\mathbf{m}_0$ as

$$\mathcal{X} = \mathcal{X}_0 + \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^\top \mathbf{H}(\mathbf{m} - \mathbf{m}_0), \tag{3}$$

$$\nabla_{\mathbf{m}}\mathcal{X} = \mathbf{H}(\mathbf{m} - \mathbf{m}_0), \tag{4}$$

where $\mathbf{H}$ is the Hessian, describing the second derivatives of the cost functional with respect to the model parameters. In the Gauss–Newton approximation, the Hessian follows from weighted dot products of the full datasets for each perturbation, summing over all receivers for all sources, and it is equal to $\mathbf{H} = \mathbf{F}^\top \mathbf{W} \mathbf{F}$.

Appendix A describes the adjoint state method for calculating the Hessian in the more general case. For a given value of $\nabla_{\mathbf{m}}\mathcal{X}$, the minimum-norm solution of Equation (4) is $\delta\mathbf{m} = \mathbf{H}^\dagger \nabla_{\mathbf{m}}\mathcal{X}$, where $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ and the superscript $\dagger$ denotes the Moore–Penrose pseudo-inverse. If the misfit function only depends on the data, possibly by ignoring penalty terms, the Fréchet derivative of the misfit function with respect to the model can be factored into $\nabla_{\mathbf{u}}\mathcal{X}$ and the $\mathbf{F} = \nabla_{\mathbf{m}}\mathbf{u}$ used here. The first will be required as input for the reverse-time part of an adjoint-state gradient computation in FWI, and should therefore be available. With these building blocks, the implementation of our method should be straight-forward for more general cost functions that involve time adaptivity (Bharadwaj et al. 2016; Bozdağ et al. 2011; Jiao et al. 2015; van Leeuwen and Mulder 2008; Warner and Guasch 2016) or optimal transport (Engquist and Yang 2022; Métivier et al. 2018).

If random vectors $\mathbf{Y}$ and $\mathbf{X}$ are related linearly as $\mathbf{Y} = \mathbf{L}\mathbf{X}$, their covariance matrices $\mathbf{C}_\mathbf{X}$ and $\mathbf{C}_\mathbf{Y}$ satisfy the equation $\mathbf{C}_\mathbf{Y} = \mathbf{L}\mathbf{C}_\mathbf{X}\mathbf{L}^\top$. If the gradient of the cost functional does not vanish exactly for the inverted model parameters but has a random distribution with covariance matrix $\mathbf{C}_\mathcal{X}$, the deviation $\delta\mathbf{m}$ of the inverted model parameters from $\mathbf{m}_0$ has a random distribution with the covariance matrix

$$\mathbf{C}_{\mathrm{m},\mathcal{X}} = (\mathbf{H}^\dagger)^\top \mathbf{C}_\mathcal{X} \mathbf{H}^\dagger = \sigma_\mathcal{X}^2 (\mathbf{H}^\dagger)^2 = \sigma_\mathcal{X}^2 (\mathbf{H}^2)^\dagger. \tag{5}$$

Here, we have taken into account that the Hessian is symmetric and assumed that $\mathbf{C}_\mathcal{X}$ is a diagonal matrix with all elements equal to $\sigma_\mathcal{X}^2$. We represent the weight matrix $\mathbf{W}$ in the form $\mathbf{W} = \mathbf{V}^\top \mathbf{V}$ and use the normalisations $\hat{\mathbf{u}} = \mathbf{V}\mathbf{u}$, $\hat{\mathbf{d}} = \mathbf{V}\mathbf{d}$. The normalised data space has the Euclidean metric, where the L_2 norm is defined as $\|\mathbf{u} - \mathbf{d}\|^2 = (\hat{\mathbf{u}} - \hat{\mathbf{d}})^\top(\hat{\mathbf{u}} - \hat{\mathbf{d}})$. The normalised difference between the inverted modelled data with and without noise $\delta\hat{\mathbf{u}}_{\mathrm{obs}} = \mathbf{V}\delta\mathbf{d}_{\mathrm{obs}}$, where $\delta\mathbf{d}_{\mathrm{obs}} = \mathbf{u}_{\mathrm{obs}} - \mathbf{u}_0$, is equal to the normal projection of $\delta\hat{\mathbf{d}}_{\mathrm{obs}} = \mathbf{V}\delta\mathbf{d}_{\mathrm{obs}}$ on the range of matrix $\hat{\mathbf{F}} = \mathbf{V}\mathbf{F}$. Taking into account that the normal projection operator on the range of $\hat{\mathbf{F}}$ is equal to $\hat{\mathbf{F}}\hat{\mathbf{F}}^\dagger$, we obtain $\delta\hat{\mathbf{u}}_{\mathrm{obs}} = \hat{\mathbf{F}}\hat{\mathbf{F}}^\dagger \delta\hat{\mathbf{d}}_{\mathrm{obs}}$ or $\hat{\mathbf{F}}\delta\mathbf{m} =$

$\hat{\mathbf{F}}\hat{\mathbf{F}}^\dagger \delta\hat{\mathbf{d}}_{\mathrm{obs}}$. The minimum norm solution of this equation is $\delta\mathbf{m} = \hat{\mathbf{F}}^\dagger \hat{\mathbf{F}} \hat{\mathbf{F}}^\dagger \delta\hat{\mathbf{d}}_{\mathrm{obs}} = \hat{\mathbf{F}}^\dagger \delta\hat{\mathbf{d}}_{\mathrm{obs}} = (\mathbf{V}\mathbf{F})^\dagger \mathbf{V}\delta\mathbf{d}_{\mathrm{obs}}$. Similarly to Equation (5), we conclude that in the case where $\delta\mathbf{d}_{\mathrm{obs}}$ is distributed with the covariance matrix $\mathbf{C}_{\mathrm{d}}$, the value $\delta\mathbf{m}$ is distributed with a covariance matrix

$$\mathbf{C}_{\mathrm{m,d}} = (\mathbf{V}\mathbf{F})^\dagger \mathbf{V}\mathbf{C}_{\mathrm{d}}\mathbf{V}^\top (\mathbf{V}\mathbf{F})^{\dagger^\top}. \tag{6}$$

If the reconstruction inaccuracies due to variations of $\nabla_{\mathbf{m}}\mathcal{X}$ and noise of the observed data $\mathbf{d}_{\mathrm{obs}}$ do not correlate with each other, the value $\delta\mathbf{m}$ is distributed with a covariance matrix equal to the sum of covariances $\mathbf{C}_{\mathrm{m},\mathcal{X}}$ and $\mathbf{C}_{\mathrm{m,d}}$.

According to Aitken's (1935) generalised least-squares method, a best unbiased estimator is obtained when the weight matrix $\mathbf{W}$ in the cost function is chosen in such a way that $\mathbf{V}\mathbf{C}_{\mathrm{d}}\mathbf{V}^\top = \sigma_{\mathrm{d}}^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and σ_{d} is the proportionality coefficient. This condition is satisfied if $\mathbf{C}_{\mathrm{d}}$ is invertible and $\mathbf{V} = \sigma_{\mathrm{d}}\mathbf{C}_{\mathrm{d}}^{-1/2}$. With this choice, we obtain $\mathbf{C}_{\mathrm{m,d}} = \sigma_{\mathrm{d}}^2 (\mathbf{V}\mathbf{F})^\dagger [(\mathbf{V}\mathbf{F})^\dagger]^\top = \sigma_{\mathrm{d}}^2 [(\mathbf{V}\mathbf{F})^\top(\mathbf{V}\mathbf{F})]^\dagger = \sigma_{\mathrm{d}}^2 (\mathbf{F}^\top \mathbf{W}\mathbf{F})^\dagger = \sigma_{\mathrm{d}}^2 \mathbf{H}^\dagger$.

In what follows, we assume that the main contribution to FWI uncertainties comes from the noise in the observed data so that the covariance matrix is proportional to $\mathbf{H}^\dagger$. The case where the uncertainties are mostly due to inaccurate inversion can be analysed in the same way, using $\mathbf{H}^2$ instead of $\mathbf{H}$.

2.3 | Confidence Ellipsoid

Let $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{U}^\top$ be the singular value decomposition of the Hessian, with singular values defined by the vector $\mathbf{s}$ and $\mathbf{S} = \mathrm{diag}(\mathbf{s})$, so that $\delta\mathbf{m}$ is distributed with the covariance matrix $\mathbf{C}_{\mathrm{m}} = \sigma_{\mathrm{d}}^2 \mathbf{U}\mathbf{S}^\dagger\mathbf{U}^\top$. If $\mathbf{H}$ has M non-zero singular values, the vectors $\mathbf{s}$ and

$$\delta\hat{\mathbf{m}} = \sigma_{\mathrm{d}}^{-1}\mathbf{S}^{1/2}\mathbf{U}^\top \delta\mathbf{m} \tag{7}$$

have M non-zero components (or components with amplitude exceeding the numerical accuracy), which are distributed with a unit covariance matrix. Consider the scalar $\zeta = \delta\mathbf{m}^\top \mathbf{H}\,\delta\mathbf{m}/\sigma_{\mathrm{d}}^2 = \delta\hat{\mathbf{m}}^\top \delta\hat{\mathbf{m}} = \delta\hat{m}_1^2 + \cdots + \delta\hat{m}_M^2$. By construction, ζ is the sum of squares of M independent random variables with zero average values and unit standard deviations $\sigma_j^2 = \mathrm{Var}(\delta\hat{m}_j) = \langle\delta\hat{m}_j^2\rangle - \langle\delta\hat{m}_j\rangle^2 = 1$, where the angular brackets denote averaging. Individual squares are distributed with the variances $\mathrm{Var}(\delta\hat{m}_j^2) = \langle\delta\hat{m}_j^4\rangle - \langle\delta\hat{m}_j^2\rangle^2 = \kappa\sigma_j^4$. The proportionality coefficient κ in the above equation depends on the actual distribution of the $\delta\hat{m}_j$. Given the large number of model parameters, the central limit theorem implies that ζ is distributed normally with mean value M and standard deviation $(\kappa M)^{1/2}$. The probability that ζ does not exceed the value ζ_c equals

$$p(\zeta \leq \zeta_c) = \frac{1}{2}[1 + \mathrm{erf}(\alpha\{(\zeta_c/M) - 1\})], \quad \alpha = \sqrt{M/(2\kappa)}, \tag{8}$$

where $\mathrm{erf}(\cdot)$ is the error function. For a normal distribution, where $\langle\delta\hat{m}_j^4\rangle = 3\sigma_j^4$ and $\kappa = 2$, Equation (8) follows from the chi-squared test in the limit $M \gg 1$.
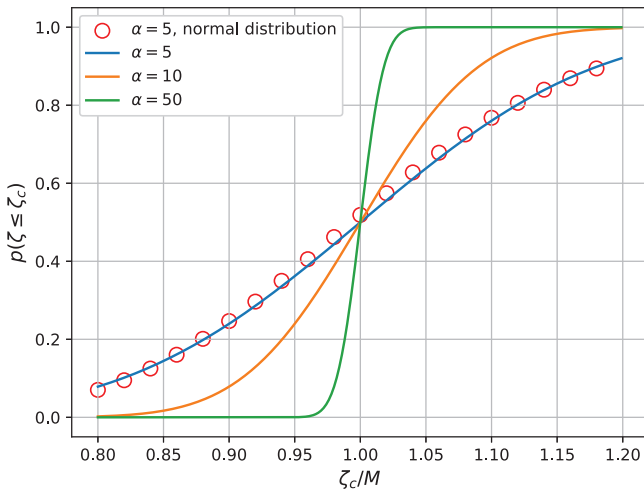
Figure 2 | Solid lines show how the confidence interval $p(\zeta \leq \zeta_c)$ depends on ζ_c/M for $\alpha = 5, 10, 50$. For normal distributions, $p(\zeta \leq \zeta_c) = \int_0^{\xi_c} \chi_M^2(\xi) d\xi$, where $\chi_M^2(\zeta) = \zeta^{M/2-1} e^{-\zeta/2} / [2^{M/2}\Gamma(M/2)]$ is the chi-squared distribution. The cumulative distribution function for χ_M^2 with $M = 100$ is shown by open circles, and it is well approximated by Equation (8) with $\alpha = 5$.

Figure 2 illustrates the behaviour of $p(\zeta \leq \zeta_c)$ and shows that Equation (8) provides a good approximation for exact distributions if $\alpha \geq 5$. The function $p(\zeta \leq \zeta_c)$ exhibits a jump when crossing the point $\zeta_c = M$, which becomes sharper with increasing M. For typical subsurface modelling, the solution of equation $p(\zeta \leq \zeta_c) = p_c$ can be approximated by $\zeta_c \simeq M$ for any p_c that is not too close to 0 or 1. This happens because of the high dimensionality of the problem. If points are randomly distributed inside an M-dimensional ellipsoid with $M \gg 1$, most of them are located near the ellipsoid boundary.

Let D be the number of data samples used for FWI. The energy of the noise present in the data is $\frac{1}{2}\delta\mathbf{d}^\top \delta\mathbf{d} = \frac{1}{2}\sigma_d^2 D$. Using this relation, we write the condition $\zeta_c \leq M$ as

$$\frac{1}{2}\delta\mathbf{m}^\top \mathbf{H}\, \delta\mathbf{m} \leq \varepsilon_0, \quad \varepsilon_0 = \epsilon(M/D)\mathcal{E}, \qquad (9)$$

where $\mathcal{E} = \mathbf{d}^\top \mathbf{d}/2$ is the energy of the measured signal and $\epsilon = \delta\mathbf{d}^\top \delta\mathbf{d}/\mathbf{d}^\top \mathbf{d}$ is the ratio of the noise energy to the signal energy. Inequality (9) can be interpreted as follows. If the noise is distributed uniformly over the data space, its energy per degree of freedom is equal to $\epsilon\mathcal{E}/D$. Since the modelled data form a subspace of dimension M in the data space, the energy of noise that is projected on the range of the linearised modelling operator (dashed line in Figure 1) is equal to $\varepsilon_0 = \epsilon(M/D)\mathcal{E}$. This part of the noise introduces uncertainty in reconstructed model parameters, which is described by condition (9). The remaining noise lies in the space complementary to the model range, and it changes the minimal residual of the cost function $\mathcal{X} - \mathcal{X}_0$, but not values of the model parameters where the residual is minimised.

Inequality (9) defines a confidence ellipsoid in the model parameter space, which contains viable solutions to the minimisation problem. The confidence ellipsoid provides a complete characterisation of uncertainty in the linear approximation. However, even if the Hessian is known, analysis of the corresponding
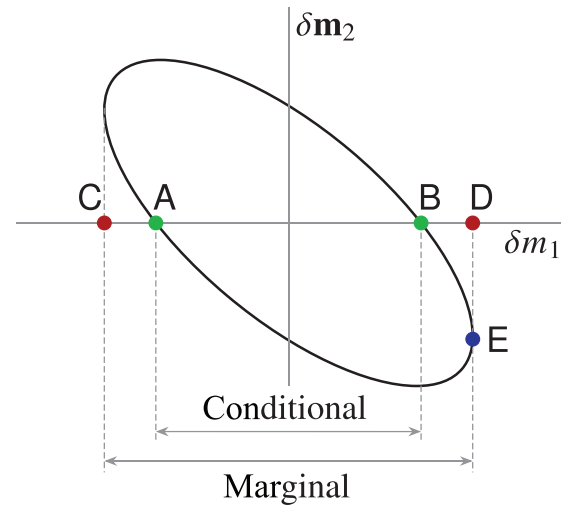
Figure 3 | The conditional uncertainty range of the parameter m_1 is the interval between points A and B, where the δm_1-axis crosses the confidence ellipsoid. The marginal uncertainty range of parameter m_1 is the interval between points C and D, which is the projection on the δm_1-axis of the confidence ellipsoid. The conditional and marginal uncertainty ranges of δm_1 are proportional to $\sqrt{1/h_{1,1}}$ and to $\sqrt{h_{1,1}^\dagger}$, respectively.

ellipsoid is difficult because of the high dimensionality of the model space. This dimensionality can be reduced by projecting the ellipsoid on specific hyperplanes in the parameter space and by considering its intersections with these hyperplanes. Figure 3 illustrates the procedure. We introduce axes $\delta m_1, \delta m_2, \ldots$ in the parameter space. The components of a vector along these axes are equal to perturbations of the corresponding model parameter. The set of axes $\delta m_2, \delta m_3, \ldots$ that does not include the axis δm_1 is denoted by $\delta\boldsymbol{m}_2$. Points lying on axis δm_1 satisfy the condition $\delta\boldsymbol{m}_2 = \mathbf{0}$. The range between points A and B, where the axis δm_1 intersects the ellipsoid, represents the conditional uncertainty of the parameter δm_1, that is, the uncertainty range of δm_1 under the condition that all the other model parameters are fixed.

The conditional uncertainty of δm_1 is described by the inequality $\delta m_1 H_{11} \delta m_1/2 \leq \varepsilon_0$. It provides the smallest uncertainty bound for this parameter: $|\delta m_1| \leq (2\varepsilon_0/H_{11})^{1/}$.

The largest uncertainty bound – the marginal uncertainty – describes the case where changes in the cost function associated with a given model parameter are maximally compensated by varying other model parameters. The marginal uncertainty is obtained by projecting the ellipsoid on the parameter axis considered. In the example illustrated by Figure 3, the marginal uncertainty of parameter m_1 is the range between points C and D, where the lines (actually, hyperplanes) $\delta m_1 = \text{const.}$ are tangent to the ellipse. At point E, δm_1 reaches its largest value.

In general, we can consider the problem of finding $\mu_i = \max(\delta m_i)$ subject to $\psi = \frac{1}{2}\delta\mathbf{m}^\top \mathbf{H}\, \delta\mathbf{m} = \varepsilon_0$. The Lagrangian $\mathcal{L}(\delta\mathbf{m}, \lambda) = \delta m_i - \lambda(\psi - \varepsilon_0)$ has a derivate $\partial\mathcal{L}/\partial\delta m_j = \delta_{ij} - \lambda(\mathbf{H}\delta\mathbf{m})_j$ that should vanish. Defining $\mathbf{v} = \mathbf{H}\delta\mathbf{m}$, this leads to $v_j = \delta_{ij}/\lambda$ and $\delta m_i = (\mathbf{C}\mathbf{v})_i = c_{ii}/\lambda$, for the pseudo-inverse $\mathbf{C} = \mathbf{H}^\dagger$. Then, $\psi = \frac{1}{2}\delta\mathbf{m}^\top \mathbf{v} = \frac{1}{2}\delta m_i/\lambda = \frac{1}{2}c_{ii}/\lambda^2 = \varepsilon_0$. The positive solution for λ yields $\mu_i = (2\varepsilon_0 c_{ii})^{1/}$ for the maximum of δm_i, and the minus sign

produces the minimum. The resulting the marginal uncertainty range is $|\delta m_i| \leq [2\varepsilon_0 (\mathbf{H}^\dagger)_{ii}]^{1/}$.

Instead of extracting a single model parameter m_1, one can split the vector $\mathbf{m}$ into two complementary parts $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2)^\top$. The Hessian is partitioned accordingly as

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} \tag{10}$$

and acts on $(\delta\mathbf{m}_1, \delta\mathbf{m}_2)^\top$. As is explained in Appendix C, the action of the Hessian on the vectors can then be represented by a sum of two terms, given by Equation (C.6). The Hessian block $\mathbf{H}_{11}$ describes the conditional uncertainties of the parameters $\mathbf{m}_1$, assuming the parameters $\mathbf{m}_2$ are fixed. The combination of Hessian blocks $\mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^\dagger\mathbf{H}_{12}$ describes the marginal uncertainty of the parameters $\mathbf{m}_2$, where one sets $\delta\mathbf{m}_1 = -\mathbf{H}_{11}^\dagger\mathbf{H}_{12}\delta\mathbf{m}_2$ to compensate for changes in the cost function due to variations of parameters $\mathbf{m}_2$.

3 | Restriction and Compression

3.1 | General Formalism

Instead of a model with n parameters, we can consider its restriction to a subspace with dimension $r < n$. An arbitrary restriction operator can be represented by an $r \times n$ matrix $\mathbf{Q}$, whose rows $\mathbf{q}_j$ with $j = 1, 2, \ldots, r$ are linearly independent vectors in the n-dimensional parameters space. The action of the Hessian on those components of $\mathbf{m}$ that do not lie in the space spanned by vectors $\mathbf{q}_j$, that is, the range $\mathcal{R}(\mathbf{Q}^\top)$ of $\mathbf{Q}^\top$, is ignored. This is equivalent to the condition that the model parameter vectors $\mathbf{m}$ lying in the null space $\mathcal{N}(\mathbf{Q})$ of $\mathbf{Q}$ are fixed.

Theorem 1. *The projected Hessian* $\mathbf{H}_p = \mathbf{PHP}$ *and the compressed Hessian* $\mathbf{H}_c = \mathbf{QHQ}^\dagger$ *have the same eigenvalues. The corresponding eigenvectors can be mapped to each other by* $\mathbf{Q}^\dagger$ *and* $\mathbf{Q}$, *respectively, apart from null-space components.*

The $n \times n$ orthogonal projection operator $\mathbf{P}$, defined on the subspace with dimension $r < n$ spanned by the row-vectors of $\mathbf{Q}$, is equal to $\mathbf{P} = \mathbf{Q}^\dagger\mathbf{Q}$.

The $n \times n$ projected Hessian $\mathbf{H}_p$ is defined by the condition that its action on parameter vectors $\mathbf{m}$ is the same as the action of the original Hessian $\mathbf{H}$ on projected parameter vectors $\mathbf{Pm}$, that is, $\mathbf{m}^\top\mathbf{H}_p\mathbf{m} = (\mathbf{Pm})^\top\mathbf{HPm}$. This requirement implies $\mathbf{H}_p = \mathbf{Q}^\dagger\mathbf{QHQ}^\dagger\mathbf{Q} = \mathbf{PHP}$.

We also introduce the $r \times r$ compressed Hessian $\mathbf{H}_c$ such that $\mathbf{m}_c^\top\mathbf{H}_c\mathbf{m}_c = \mathbf{m}^\top\mathbf{H}_p\mathbf{m} = \mathbf{m}_c^\top\mathbf{QHQ}^\dagger\mathbf{m}_c$, where $\mathbf{m}_c$ is an r-dimensional vector defined by the equation $\mathbf{m}_c = \mathbf{Qm}$. The minimum-norm solution of the above equation is $\mathbf{m} = \mathbf{Q}^\dagger\mathbf{m}_c$, which shows that $\mathbf{Q}^\dagger$ is the prolongation operator.

The compressed Hessian $\mathbf{H}_c$ is expressed via $\mathbf{H}$ and $\mathbf{H}_p$ as $\mathbf{H}_c = \mathbf{QHQ}^\dagger = \mathbf{QH}_p\mathbf{Q}^\dagger$, where we have used the following properties of pseudo-inverses: $\mathbf{QQ}^\dagger\mathbf{Q} = \mathbf{Q}$ and $\mathbf{Q}^\dagger\mathbf{QQ}^\dagger = \mathbf{Q}^\dagger$.

PROOF. An eigenvector $\mathbf{v}'$ with eigenvalue λ' of the compressed $\mathbf{H}_c$ obeys $\mathbf{QHQ}^\dagger\mathbf{v}' = \lambda'\mathbf{v}'$. Since $\mathbf{PQ}^\dagger = \mathbf{Q}^\dagger$, we have $\mathbf{Q}^\dagger\mathbf{QHQ}^\dagger\mathbf{v}' = \mathbf{H}_p(\mathbf{Q}^\dagger\mathbf{v}') = \lambda'(\mathbf{Q}^\dagger\mathbf{v}')$. The other way around, consider $\mathbf{H}_p\mathbf{v} = \lambda\mathbf{v}$ and note that $\mathbf{v}$ lies in the range of $\mathbf{P}$ and, therefore, of $\mathbf{Q}^\dagger$. Then, $\mathbf{QH}_p\mathbf{v} = (\mathbf{QHQ}^\dagger)(\mathbf{Qv}) = \mathbf{H}_c(\mathbf{Qv}) = \lambda(\mathbf{Qv})$. $\quad\square$

3.2 | Compression With Semi-Orthogonal Matrices

As mentioned above, the operator $\mathbf{P}$ projects vectors on the space spanned by rows of the restriction matrix $\mathbf{Q}$, which is the same as the range $\mathcal{R}(\mathbf{Q}^\top)$ of the transposed matrix. The matrix $\mathbf{Q}$ can be constructed such that its rows $\mathbf{q}_j$ form an orthonormal basis in $\mathcal{R}(\mathbf{Q}^\top)$. Then, $\mathbf{Q}^\dagger = \mathbf{Q}^\top$ and $\mathbf{QQ}^\top = \mathbf{I}_r$, where $\mathbf{I}_r$ is a $r \times r$ unit matrix. The corresponding $n \times n$ operator $\mathbf{P} = \mathbf{Q}^\top\mathbf{Q}$ projects vectors on the same subspace of the model space as the original operator $\mathbf{P}$.

Theorem 2. *Given a real symmetric matrix* $\mathbf{A}$ *of size* $n \times n$ *with eigenvalues* λ_i, $i = 1, 2, \ldots, n$, *sorted in descending order. The semi-orthogonal* $r \times n$ *matrix* $\mathbf{Q}$, *with* $r \leq n$, *defines its restriction to an* r-*dimensional linear subspace and obeys* $\mathbf{QQ}^\top = \mathbf{I}_m$, *where* $\mathbf{I}_m$ *is the* $r \times r$ *identity matrix. Then, the eigenvalues* μ_i, $i = 1, 2, \ldots, r$, *in descending order, of* $\mathbf{B} = \mathbf{QAQ}^\top$ *obey*

$$\lambda_i \geq \mu_i \geq \lambda_{n-r+i}, \quad i = 1, 2, \ldots, r. \tag{11}$$

The compression $\mathbf{m}_c = \mathbf{Qm}$ with semi-orthogonal matrix $\mathbf{Q}$ can be viewed as a map to a lower-dimensional space. The projected Hessian $\mathbf{H}_p$ and the corresponding covariance matrix $\mathbf{C}_p = \mathbf{H}_p^\dagger$ are transformed as $\mathbf{H}_c = \mathbf{QH}_p\mathbf{Q}^\top$ and $\mathbf{C}_c = \mathbf{QC}_p\mathbf{Q}^\top$. By the definition of the Moore–Penrose pseudo-inverse, $\mathbf{C}_p\mathbf{H}_p\mathbf{C}_p = \mathbf{C}_p$, $\mathbf{H}_p\mathbf{C}_p\mathbf{H}_p = \mathbf{H}_p$, $(\mathbf{H}_p\mathbf{C}_p)^\top = \mathbf{H}_p\mathbf{C}_p$, $(\mathbf{C}_p\mathbf{H}_p)^\top = \mathbf{C}_p\mathbf{H}_p$. Also, $\mathbf{C}_p\mathbf{H}_p = \mathbf{P}$. It can be verified that the compressed matrices $\mathbf{H}_c$ and $\mathbf{C}_c$ also satisfy the above four properties. Hence, if $\mathbf{Q}$ is semi-orthogonal, the pseudo-inverse of the compressed Hessian is the same as the compression of the projected covariance matrix: $\mathbf{H}_c^\dagger = \mathbf{QC}_p\mathbf{Q}^\top$.

Semi-orthogonal matrices allow for the application of Poincaré's separation theorem (Gradshteyn and Ryzhik 2000, for instance), which relates the eigenvalues of a given real symmetric matrix of size $n \times n$ to those of its compression to a subspace with dimension $m < n$.

The theorem implies that the eigenvalues of $\mathbf{H}_p$ and $\mathbf{H}_c$ are not larger than the eigenvalues of the original Hessian $\mathbf{H}$.

Poincaré's separation theorem is not applicable to the covariance matrix $\mathbf{C}_p$ in relation to $\mathbf{H}$ because the pseudo-inverse of the projected Hessian $\mathbf{H}_p$ is not the same as the projection of the pseudo-inverse of the Hessian $\mathbf{H}$, that is, $\mathbf{C}_p = \mathbf{H}_p^\dagger \neq \mathbf{PH}^\dagger\mathbf{P}$. This can also be understood from a heuristic point of view because restricting the types of possible perturbations also limits the opportunity to find a combination where such perturbations maximally compensate each other.

An alternative way to determine the marginal uncertainty of a parameter m_1 that represents a geological unit is the following. Choose a $\delta m_1 = \Delta m_1$ and minimise the cost func-

tional on the original model over all parameters while keeping $\delta m_1 = \Delta m_1$ fixed. The minimum is reached at $\delta \mathbf{m}_{2,\min} = (\delta m_{2,\min}, \delta m_{3\min}, \ldots)^\top$, and the value of the cost function at the minimum is $\mathcal{X}_{\min} = \mathcal{X}(\Delta m_i, \delta \mathbf{m}_{2,\min})$. If $\mathcal{X}_0 = 0$ in Equation (3), the quadratic behaviour of $\mathcal{X}(\delta \mathbf{m}) = \frac{1}{2}\delta \mathbf{m}^\top \mathbf{H}\, \delta \mathbf{m} = \varepsilon_0$ in δm_1 can be used to rescale $\mathcal{X}_{\min}$ to $\mathcal{X}(\delta \mathbf{m})$, resulting in a marginal uncertainty $|\delta m_i| \leq (\varepsilon_0/\mathcal{X}_{\min})^{1/}|\Delta m_i|$. In Figure 3, the ellipse described by $\mathcal{X}_{\min} = \mathcal{X}(\Delta m_i, \delta \mathbf{m}_{2,\min})$ would be an enlarged or shrunk version of the original ellipse given by $\mathcal{X}(\delta \mathbf{m}) = \varepsilon_0$ and the rescaling would make them the same.

3.3 | Construction of Semi-Orthogonal Restriction Matrices

The simplest way to construct a semi-orthogonal restriction matrix $\mathbf{Q}$ is to specify its rows as $\mathbf{q}_1 = (1, 0, \ldots, 0)$, $\mathbf{q}_2 = (0, 1, \ldots, 0)$, so that j-th vector $\mathbf{q}_j$ has a single non-zero j-th component equal to 1. The corresponding compressed Hessian $\mathbf{H}_c$ is the $r \times r$ upper-left block of the original Hessian $\mathbf{H}$, which characterises conditional uncertainties of first r model parameters. One can also choose $\mathbf{q}_j$ with r arbitrary indices j, with $1 \leq j \leq n$. In that case, the compressed Hessian consists of the elements of the original Hessian formed at the intersections of rows and columns for the selected indices. More generally, a semi-orthogonal restriction matrix $\mathbf{Q}$ can be constructed from arbitrarily chosen rows from an orthogonal matrix.

Orthogonal matrices with unit determinants act as rotations. A rotation in n dimensions can be described by the pivot axis $\boldsymbol{\xi}$, and a unit $(n-1)$-dimensional 'pull' vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_{n-1})^\top$ in the plane perpendicular to $\boldsymbol{\xi}$ (Hanson 1995, Equation 7). The coordinate axes are transformed in such a way that the pivot axis $\boldsymbol{\xi}$ rotates towards $\boldsymbol{\eta}$ by an angle θ in the plane formed by $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. In the case where the pivot axis is the last coordinate axis, the matrix $\mathbf{R}^\top$ that performs the above rotation has the form (Hanson 1995)

$$\mathbf{R}^\top = \begin{pmatrix} 1 - r_{1,1} & -r_{1,2} & \ldots & -r_{1,n-1} & -s\eta_1 \\ -r_{2,1} & 1 - r_{2,2} & \ldots & -r_{2,n-1} & -s\eta_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -r_{n-1,1} & -r_{n-1,2} & \ldots & 1 - r_{n-1,n-1} & -s\eta_{N-1} \\ s\eta_1 & s\eta_2 & \ldots & s\eta_{N-1} & c \end{pmatrix}. \quad (12)$$

Here, $r_{ij} = (1-c)\eta_i\eta_j$, $c = \cos\theta$ and $s = \sin\theta$. Rows of $\mathbf{R}^\top$ are coordinates of the new coordinate axes in the original coordinate systems. The transpose of matrix (12) is commonly called the rotation matrix $\mathbf{R}$ (fixed coordinate system rotation), and it relates point coordinates $\mathbf{m}'$ in the rotated coordinate system to the original coordinates $\mathbf{m}$ by $\mathbf{m}' = \mathbf{R}\mathbf{m}$.

As an example, consider a 2D coordinate system where the pivot axis $\boldsymbol{\xi} = \mathbf{e}_y$ rotates outwards the 'pull' vector $\boldsymbol{\eta} = \mathbf{e}_x$ over an angle $\pi/4$. Setting $\theta = -\pi/4$ in Equation (12), we have

$$\mathbf{R}^\top = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}. \quad (13)$$

If the matrix $\mathbf{Q}$ is chosen as the first row of $\mathbf{R}^\top$, the compressed Hessian $\mathbf{H}_c$ acts along the first axis of the rotated coordinate system where $\delta m_1 = \delta m_2$. Figure 4a provides an illustration of the cost function near the minimum if two parameters m_1 and m_2 are involved. If the noise level is set to $\mathcal{X}_{\text{noise}} = 1$, the cross section $\mathcal{X} = \mathcal{X}_{\text{noise}}$ through the paraboloid defines an ellipse. Figure 4a displays the values $\mathcal{X} \leq \mathcal{X}_{\text{noise}}$. If one parameter is fixed, in this case, the value of m_2 at the minimum of the cost function, the conditional probability distribution measures the width of m_1 inside the ellipse, whereas the marginal distribution describes the outer bounds of m_1 on the ellipse, as shown in Figure 4c. The line segment between the two small open circles in Figure 4c, which falls inside the original confidence ellipse, represents the compressed confidence range. This segment has a length between the shortest and longest axis, illustrating Poincaré's theorem.

Another approach to construct $\mathbf{Q}$, used in Section 4 is the following. The computational domain is partitioned into several disjoint subsets of model parameters. The number of model parameters in the j-th set is denoted as M_j, and the set of their indices is denoted as S_j. The j-throw of $\mathbf{Q}$ is defined as $q_{j,k} = 1/\sqrt{M_j}$ if $k \in S_j$ and $q_{j,k} = 0$ if $k \notin S_j$. For example, the selection of two sets of model parameters $\{m_1, m_2, m_3\}$ and $\{m_4, m_5\}$ out of a larger set results in the restriction matrix
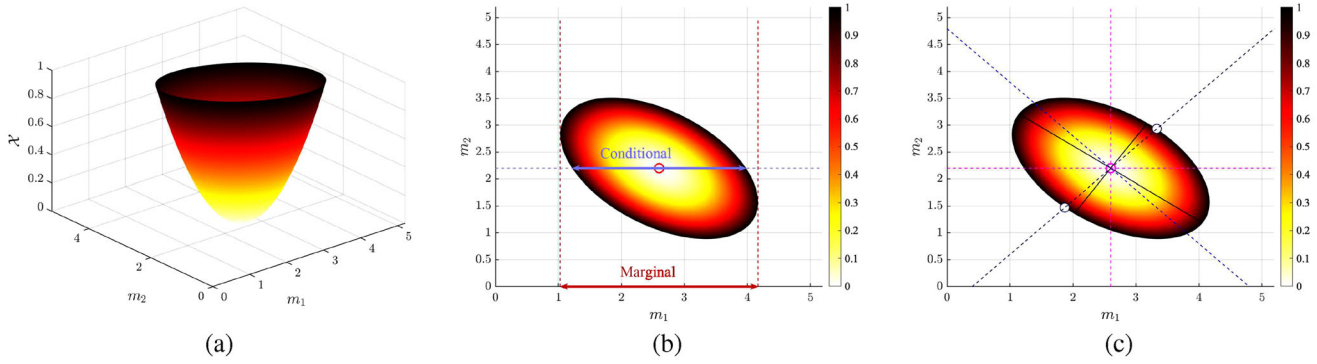
$$\mathbf{Q} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \ldots \end{pmatrix}. \quad (14)$$

In this case, the compressed Hessian represents perturbations for which the model parameters within the same set are changed by the same value.

We can also construct a compressed Hessian that describes perturbations with certain spectral properties. In particular, filtering out small-scale perturbations bears some similarity to the homogenisation method (Cao et al. 2024; Capdeville and Métivier 2018; Cupillard and Capdeville 2018; Gibson et al. 2014; Owhadi and Zhang 2008). Another approach is to compose $\mathbf{Q}$ from rows of discrete cosine or sine Fourier transforms (Fichtner and Trampert 2011b). If n is a power of 2, the Walsh–Hadamard matrices with entries equal to $\pm 1/\sqrt{n}$ can be used to define $\mathbf{Q}$ (Fino and Algazi 1976; Thompson 2017). Hessian compression with Hadamard matrices is conceptually the same as filtering in the wavelet-domain representation. If we denote the ν-thcolumn of the 2^q-dimensional Walsh–Hadamard matrix by $\mathbf{v}_\nu^{(q)}$, the column vectors $\mathbf{v}_\nu^{(q)}$ can be constructed as follows (Ben-Artzi et al. 2007). Starting from a one-dimensional vector $\mathbf{v}_1^{(1)}$ whose single component is equal to 1, 2D Walsh–Hadamard vectors are obtained by joining the components of $\mathbf{v}_1^{(1)}$ with plus and minus signs: $\mathbf{v}_1^{(2)} = (1, 1)^\top$ and $\mathbf{v}_2^{(2)} = (1, -1)^\top$. From the 2^j-dimensional vector $\mathbf{v}_\mu^{(j)}$, two 2^{j+1}-dimensional vectors are constructed by $\mathbf{v}_{2\mu-1}^{(j+1)\top} = [(\mathbf{v}_\mu^{(j)})^\top, (-1)^{\mu-}(\mathbf{v}_\mu^{(j)})^\top]$, and $\mathbf{v}_{2\mu}^{(j+1)\top} = [(\mathbf{v}_\mu^{(j)})^\top, (-1)^\mu(\mathbf{v}_\mu^{(j)})^\top]$. We then can introduce the Walsh–Hadamard restriction matrices $\mathbf{Q}_{p,\nu}^{(q)}$ with dimensions $2^{p-q} \times 2^p$ as

$$\mathbf{Q}_{p,\nu}^{(q)} = \frac{1}{2^{q/2}} \begin{pmatrix} \mathbf{v}_\nu^{(q)\top} & 0 & \cdots & 0 \\ 0 & \mathbf{v}_\nu^{(q)\top} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{v}_\nu^{(q)\top} \end{pmatrix}, \quad (15)$$

Figure 4 | (a) The cost functional $\mathcal{X}(m_1, m_2)$ for two parameters has the shape of a paraboloid near the minimum at $(m_{0,1}, m_{0,2})$. Its map view (b) shows the conditional and marginal uncertainty for given $m_2 = m_{0,2}$ for a threshold value $\epsilon = 1$. (c) The eigenvalues of the Hessian describing the paraboloid correspond to the axes of the ellipsoid. The dotted lines indicate the coordinate axes after a 45° rotation. The projection on the subspace defined by $m_1 - m_{0,1} = m_2 - m_{0,2}$ is the line segment between the two white dots, where the line intersects the ellipse. It has a length in between that of the shortest and longest axis.

where $\mathbf{v}_\nu^{(q)}$ is the Walsh–Hadamard vectors constructed above. Restricting the $2^p \times 2^p$ matrix $\mathbf{H}$ with $\mathbf{Q}_{p,\nu}^{(q)}$ is the same as (spatial) frequency filtering. The Fourier domain of $\mathbf{H}$ is split into 2^{p-q} blocks of equal length, and the action of $\mathbf{H}$ is restricted by vectors lying in the ν-th Fourier block.

4 | Examples

4.1 | Computation of Perturbation Data

The Gauss–Newton approximation of the Hessian provides the same result as the Hessian for a modelling operator based on its Born approximation. The computation of Born scattering data for each model perturbation with, for instance, a finite-difference code, involves the simultaneous solution of two (systems of) equations. The partial differential equation (2) can be split into $\mathcal{L}(\mathbf{m}_0, \mathbf{u}_0) = \mathbf{f}$ for the background wavefield $\mathbf{u}_0 = \mathbf{u}(\mathbf{m}_0)$ and $\mathcal{L}(\mathbf{m}_0, \delta\mathbf{u}) \simeq -[\partial\mathcal{L}(\mathbf{m}_0, \mathbf{u}_0)/\partial\mathbf{m}]\delta\mathbf{m}$ for the scattered field $\delta\mathbf{u}$, thereby doubling the compute cost. The Born approximation is usually applied to models that are split into a smooth background model that does not provide scattering in the frequency band of interest, and rough components that define the reflectors (Tarantola 1984; Østmo et al. 2002). In our case, the background model is assumed to be the full-waveform inversion (FWI) result and the perturbation data may contain free-surface and interbed multiples.

For the two-dimensional isotropic elastic examples shown later, we use a Taylor series approach, which requires the computation of $\mathcal{L}(\mathbf{m}_1, \mathbf{u}_1) = \mathbf{f}$ with $\mathbf{m}_1 = \mathbf{m}_0 + \varepsilon\,\delta\mathbf{m}$ and provides $\delta\mathbf{u} \simeq (\mathbf{u}_1 - \mathbf{u}_0)/\varepsilon$. In that case, the receiver data for the background wavefield $\mathbf{S}_r\mathbf{u}_0$, with sampling operator $\mathbf{S}_r$, only have to be computed once for all model perturbations. While this avoids the doubling of the cost, the choice of ε is critical. If too small, the data will be severely affected by numerical noise and round-off errors. If too large, non-linear effects will appear. Nevertheless, we have used this method for the 2D isotropic elastic examples that are shown later. The elements of the Hessian follow from dot products between data for different perturbations (Mulder and Kuvshinov 2025).

According to Equation (13) of Huang (2023), the Born approximation in the scalar constant-density acoustic case produces scattering data of the form $G_j(\mathbf{x}, \mathbf{x}') - G^{(0)}(\mathbf{x}, \mathbf{x}') = \int_{\Omega_j} d\mathbf{x}''\, G^{(0)}(\mathbf{x}, \mathbf{x}'') V_j(\mathbf{x}'') G^{(0)}(\mathbf{x}'', \mathbf{x}')$. In our setting, the full domain $\Omega = \bigcup_{j=1}^m \Omega_j$ is partitioned into m disjoint subsets Ω_j and the perturbation $V_j(\mathbf{x})$ has a unit amplitude inside Ω_j and is zero elsewhere. The Taylor series approach yields an approximation somewhere in between the Born approximation $G^{(0)}G^{(0)}$ and the $G^{(0)}G$ of Equation (19) in Huang (2023), known as the Dyson equation in quantum mechanics or as the primary–secondary formulation in controlled-source electromagnetics. The differences between these three should be small for small perturbations, in the order of percent, implying the implicit assumption that the uncertainties are also small, of the same order of magnitude.

In practice, it is more convenient to work with relative perturbations of the form

$$m_i = m_{0,i}\left(1 + \frac{(\delta m)_i}{m_{0,i}}\right) = m_{0,i}(1 + \delta\log m_i), \tag{16}$$

for all model parameters enumerated by i. This will rescale the Hessian by $\mathrm{diag}(\mathbf{m}_0)\mathbf{H}\,\mathrm{diag}(\mathbf{m}_0)$, with $\mathrm{diag}(\mathbf{m}_0)$ the diagonal matrix with values $m_{0,i}$. Here and in what follows, we simplify the notation and will use m_i to denote the relative perturbation $\delta\log m_i$. The Hessian is assumed to be scaled accordingly.

4.2 | Two-Dimensional Homogeneous Acoustic Problem

We start with a Hessian for the 2D constant-density acoustic wave equation, computed analytically in the frequency domain with the exact Green functions. The model has a density of $\rho = 2\,\mathrm{g/cm}^3$ and a P-wave velocity of $v_p = 1.5$ km/s. We choose a 15-Hz Ricker wavelet and only consider frequencies from 4 to 30 Hz at a 0.5-Hz interval. The Hessian is computed on a regular grid in a subdomain defined by $x \in [-250, 250]$ m and $z \in [750, 1250]$ m with a 5-m spacing. Sources and receivers are located at zero depth with lateral positions $x_s \in [-887.5, 887.5]$ m for the
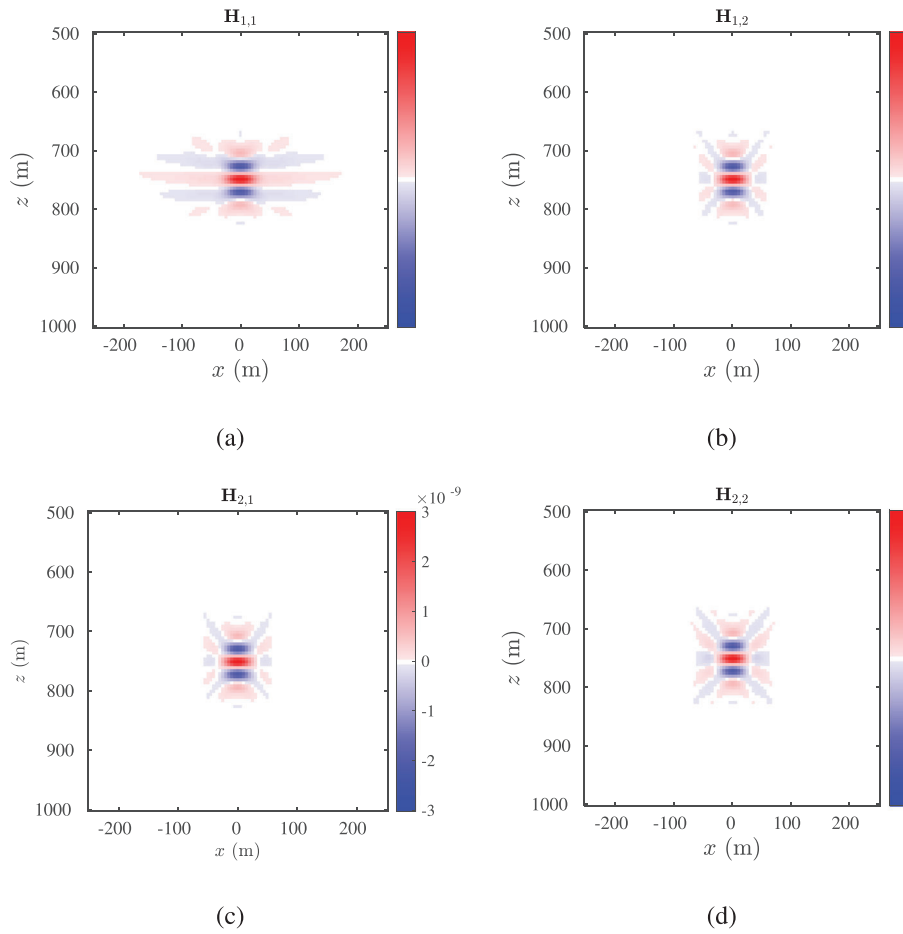
Figure 5 | Action of the Hessian on a scatterer at $(x, z) = (0, 1000)$ m, for a perturbation $m_1 = \delta \log[1/(\rho v_p^2)]$, shown in (a) and (b), or a perturbation $m_2 = \delta \log(1/\rho)$, shown in (c) and (d).

shots and $x_r \in [-900, 900]$ m for the receivers, both with a 25-m spacing.

Figure 5 displays two 'lines' of the Hessian, the response of a scatterer at the centre of the domain, at $x = 0$ m and $z = 1000$ m, for either a perturbation $m_1 = \delta \log[1/(\rho v_p^2)]$, in Figure 5a, b, or $m_2 = \delta \log(1/\rho)$, in Figure 5c,d. The imprint of the Ricker wavelet is visible in the vertical direction, whereas longer wavelengths appear in the horizontal direction. As is clear from these images, the finer scales at the level of the 5-m grid spacing are not resolved, and we expect a large null space.

Figure 6a displays the eigenvalues of a subset of the Hessian as a black line, for 100×100 points instead of 101×101, dropping the results for positions at the highest value of x and of z. The reason for taking a subset is that is easier to build its projections by combining 2×2 or ☐ points inside small squares of the grid. The resulting eigenvalues are shown as red and blue lines, respectively. All curves are scaled by the maximum eigenvalue for the original grid. On the latter, less than 3500 of the 20,000 eigenvalues are not zero, taking 10^{-16} as a rather small threshold. When projected to groups of 2×2 points, about 3000 of the 5000 are not zero. For groups of 4×4, about 1200 of the 1250 eigenvalues are not zero.

This shows that the projection helps to remove the null-space components, in particular, sub-resolution model features that cannot be reconstructed from the data and have an infinite uncertainty. We also observe that the projections do not increase the eigenvalues, in agreement with Poincaré's theorem. Alternatively, by stretching the horizontal axis for the compressed case, we obtain results that are closer to the original black curve, in particular for the 2×2 compression represented by the red curve in Figure 6b. This behaviour is expected as long as the group of grid points lies in a Fresnel zone and their responses add coherently.

4.3 | Two-Dimensional Ocean Bottom Node Data, One-Dimensional Isotropic Elastic Model

Figure 7a displays a deep-water 1D isotropic elastic model, in terms of density ρ, P-wave velocity v_p and S-wave velocity v_s. For the computation of the Hessian, the water layer, down to a depth of 1400 m, is described by a constant density and water velocity. In reality, they will vary with temperature and salinity and column pressure and abrupt depth changes such as a thermocline layer may even produce reflections in the seismic bandwidth. For the deepest layer, beyond 4700 m, the three elastic parameters are also assumed to be constant with depth. The grid spacing for 2D finite-difference modelling was set at 10 m. The parameters
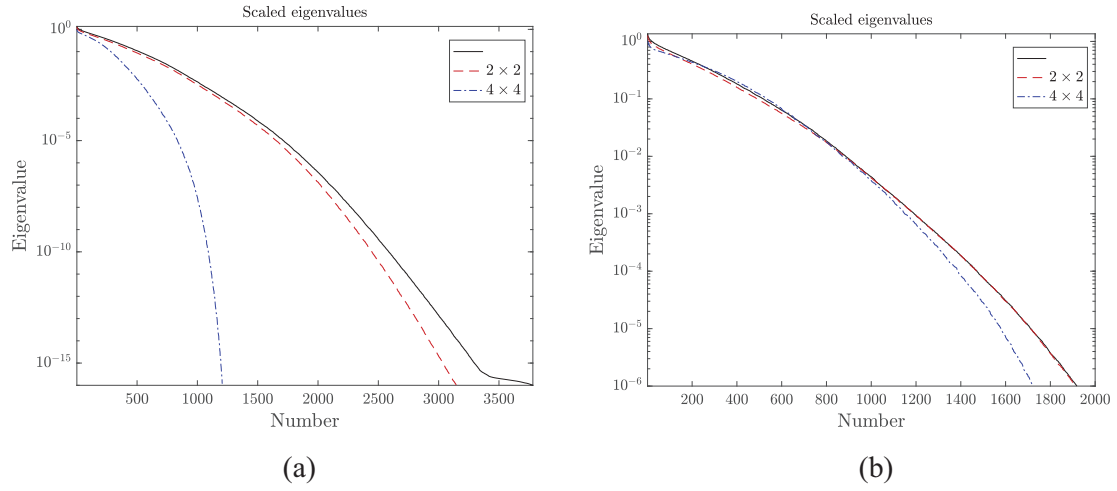
Figure 6 | (a) Eigenvalues of the Hessian (black), of its projection on 2×2 (red dashed) and 4×4 points (blue dash-dotted). (b) A subset with the horizontal scale multiplied by 1, 1.035 and 1.88, respectively.
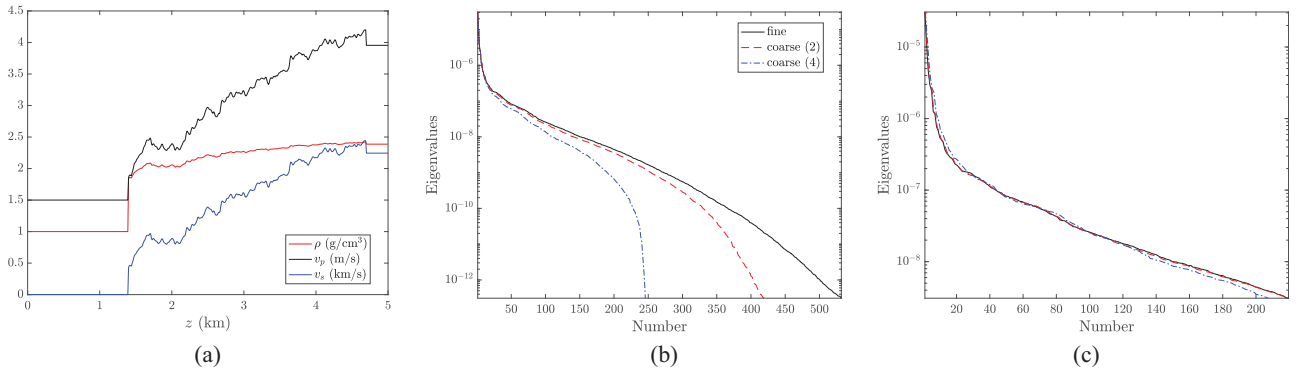
Figure 7 | (a) 1D isotropic elastic model. (b) Eigenvalues of the Hessian on the modelling grid (black) and when coarsened per 2 (red) and per 4 (blue). (c) As (b) but with the larger eigenvalues and the horizontal axis scaled by 1, 1.06 and 1.3, respectively.

for the water layer and the deepest were assumed to be known, leaving $3 \times 330 = 990$ model parameters on the 10-m modelling grid. A total of 161 shots were fired at a depth of 10 m and horizontal offsets from 0 to 8000 m at a 50-m interval, using an 8-Hz Ricker wavelet. For the receiver at the sea bottom, only the P-data were used with a recording time of 10 s and a 4-ms sampling. Reciprocity was applied for modelling. A free-surface boundary condition was imposed.

Figure 7b shows the eigenvalues of the Hessian as a black line. As already mentioned, the water and deepest layer were ignored. The result was not scaled by the number of points in the x-coordinate. When the 330 points were coarsened by combining adjacent depth pairs, the compressed Hessian has $3 \times 330/2 = 495$ eigenvalues, drawn in red. When 4 points in depth are combined, the last layer has only 2 points combined, and there are 249 eigenvalues, drawn in blue. In the last case, the null-space components have effectively been removed. Because the uniform finite-difference grid is much finer than the resolvable scales at a larger depth, the null space is expected to have a substantial size. Figure 7c shows a subset of the same eigenvalues but with the horizontal axis stretched.

Figure 8 displays three types of uncertainty estimates. The first in Figure 8a is the conditional one, obtained by fixing all parameters and selecting one value on the main diagonal of the projected Hessian $\mathbf{H}_p$ and finding σ_k from $\frac{1}{2}\sigma_k h_{p,k,k}\sigma_k = \epsilon'\mathcal{X}_0$ with $\epsilon' = 10^{-4}$ and $\mathcal{X}_0$ the data energy in the reference or background model.

We have plotted the results for the projected Hessian $\mathbf{H}_p$ rather than the compressed $\mathbf{H}_c$, because the former is defined on the original model space. Appendix D offers a pictorial description of three ways to map uncertainties obtained from the compressed Hessian to the modelling grid, using $\mathbf{Q}^\top$, simple copies, or estimates from $\mathbf{H}_p$. The latter is more suitable when the geological units are small, and null-space components dominate, causing ellipsoids to be elongated in the direction perpendicular to that of the compression, as illustrated in Figure A1c.

With the chosen restriction operator, the projection with $\mathbf{H}_p$ replaces the original model perturbations by their average in each segment. This can be easily seen by an example. Consider the compression operator from Equation (14) and a diagonal Hessian $\mathbf{H}$ with $\mathrm{diag}(\mathbf{H}) = (a, a, a, b)^\top$ and zeros elsewhere. Then, $\mathbf{H}_c$ is diagonal with $\mathrm{diag}(\mathbf{H}_c) = (a, b)^\top$ and
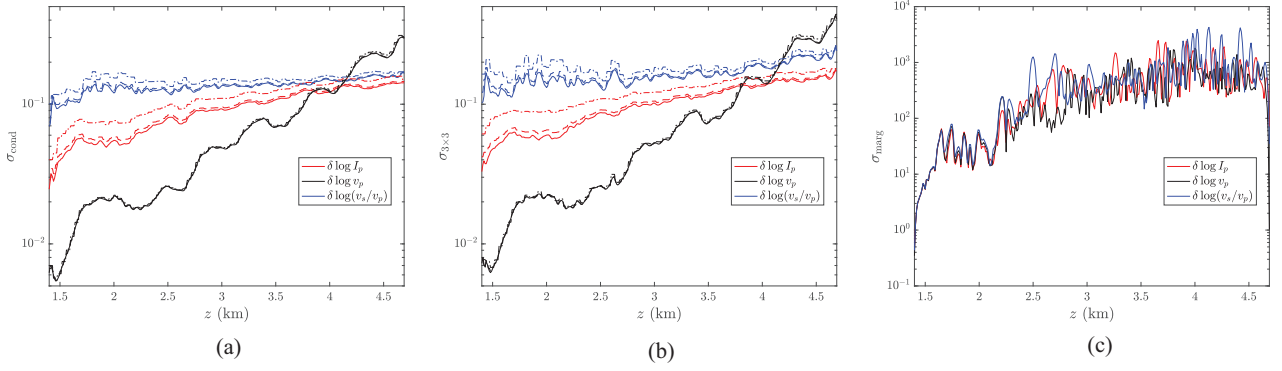
Figure 8 | (a) Conditional uncertainty based on $\mathbf{H}_p$ for single parameters and (b) for 3×3 blocks. (c) Marginally uncertainty. The drawn lines correspond to the 1D model grid, the dashed to combinations of 2 points into a single relative model perturbation and the dash-dotted lines to 4 points combined. For the marginals in (c), only the finest grid is shown.

$$\mathbf{H}_p = \begin{pmatrix} a/3 & a/3 & a/3 & 0 & 0 \\ a/3 & a/3 & a/3 & 0 & 0 \\ a/3 & a/3 & a/3 & 0 & 0 \\ 0 & 0 & 0 & b/2 & b/2 \\ 0 & 0 & 0 & b/2 & b/2 \end{pmatrix}.$$

The Hessians $\mathbf{H}$, $\mathbf{H}_c$ and $\mathbf{H}_p$ share the same eigenvalues a and b, but for $\mathbf{H}$ they are repeated 3 and 2 times, respectively, whereas $\mathbf{H}_p$ has three zeros because of the repeated entries in the matrix.

The second type of uncertainty estimate in Figure 8b is partially conditional, fixing parameters everywhere except at one depth, and plotting the diagonal of the local covariance matrix. This amounts to selecting one 3×3 block of the Hessian for each point and extracting the diagonal of its inverse. Figure 8c shows the marginal uncertainty, based on the diagonal of the covariance matrix for the full problem.

The uncertainty increases with depth, as expected, and also after projection to the lower-dimensional subspaces. The marginal uncertainty is very large but decreases after projection when null-space and near null-space components, in particular those related to unresolved features, are removed. We also observe a decrease towards the bottom boundary in Figure 8c, which is presently not understood.

A potential disadvantage of using $\mathbf{H}_p$ instead of $\mathbf{H}_c$ is its size. However, that storage can be saved by working with the compressed Hessian $\mathbf{H}_c$ instead of $\mathbf{H}_p$, since the latter contains duplicate entries, as explained in Appendix E.

4.4 | Two-Dimensional Marine Example

Figure 9 shows a 2D marine model, used earlier (Mulder and Kuvshinov 2023; 2025). Figure 10a displays the index map, where each index value denotes a geological unit. The four negative values refer to four reservoirs. In the model, we have assumed that elastic properties are constant inside each fine-grid unit, although that is not required for the method, as only the relative perturbations are assumed to be constant. Figure 10b depicts a coarser version, obtained by combining pairs of adjacent layers, excluding the seawater, down to a depth of 800 m and the four

reservoirs. Both index maps define projections, a finer and a coarser one, relative to the modelling grid that has a 10-m grid spacing.

For the acquisition, 199 shot positions range from $x_s = -2900$ to 7000 m at a 50-m interval and a depth of 10 m. The source wavelet is a 15-Hz Ricker integrated twice in time, that is, a Gaussian with a standard deviation $\sigma_w = (\pi \sqrt{2} f_{\text{peak}})^{-1}$ and $f_{\text{peak}} = 15$, Hz. Receivers at an 8-m depth have offsets at a 25-m interval from 100 to 6000 m or less when the rightmost boundary of the domain is reached, and 7 s of data were recorded and sampled at 4 ms. A free-surface boundary condition is imposed, suppressing low frequencies in the data.

The finest grid has points in the set $V^{(0)}$. When compressed with a projection operator $\mathbf{Q}_{(0)}^{(1)}$, the larger scale geological units are elements of the set $V^{(1)}$. A further compression produces a set $V^{(2)}$. Then, $\mathbf{Q}_{(1)}^{(2)} = \mathbf{Q}_{(0)}^{(2)} \left(\mathbf{Q}_{(0)}^{(1)}\right)^{\top}$. The resulting operator involves the following steps: undo the $1/\sqrt{n_{\text{f},j}}$ scaling for the finer $\mathbf{H}_c^{(1)}$, where $n_{\text{f},j}$ is the number of points inside each geological unit j, add contributions from finer to coarser, apply the $1/\sqrt{n_{\text{c},k}}$ scaling after summation, where $n_{\text{c},k}$ is the number of points inside each larger geological unit after projection, assuming each coarser unit is obtained by combining one or more finer ones. This describes the relation between the Hessian used for Figures 11 and 12 and the one used for Figures 13 and 14. These figures are based on the compressed Hessians $\mathbf{H}_c$ and their pseudo-inverses, for the finer and coarser segmentation, and the uncertainty estimates are just copied to the modelling grid for display purposes (option 2 in Appendix D).

Figure 15a shows a subset of the covariance matrix, restricted to the reservoir with index -1 and describing the marginal distribution. The result for the coarser projection is shown in Figure 15b and has somewhat smaller values. Figure 15c displays the result for the conditional case, assuming all parameters outside the reservoir are known. Since this part of the model is the same in the coarser projection, the corresponding matrix is also the same.

Figure 16 displays the conditional distribution with all parameters fixed with the exception of the P-impedances for the two units
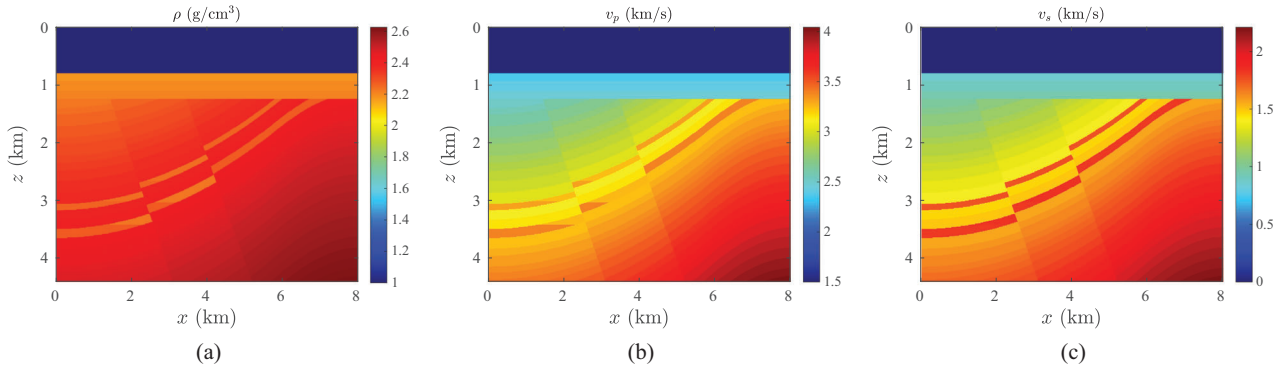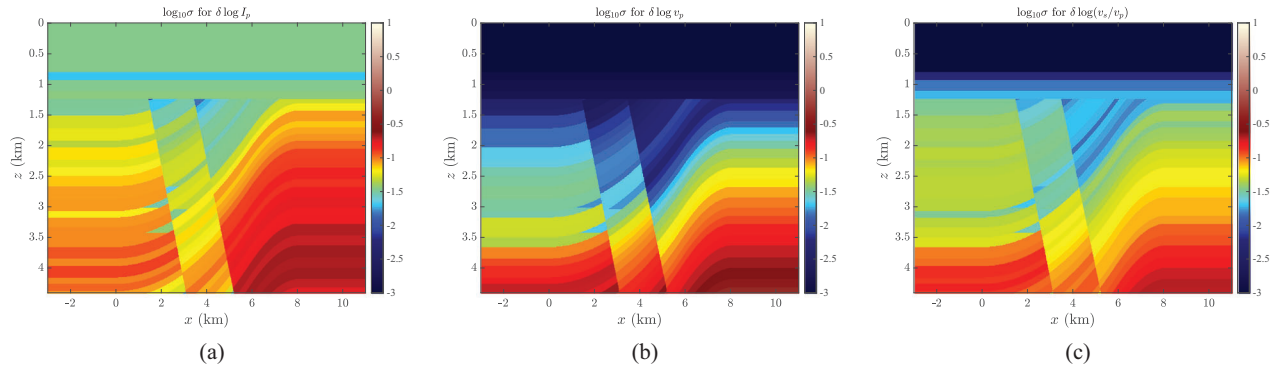
Figure 9 | Isotropic elastic model with (a) density, (b) P- and (c) S-wave velocity.
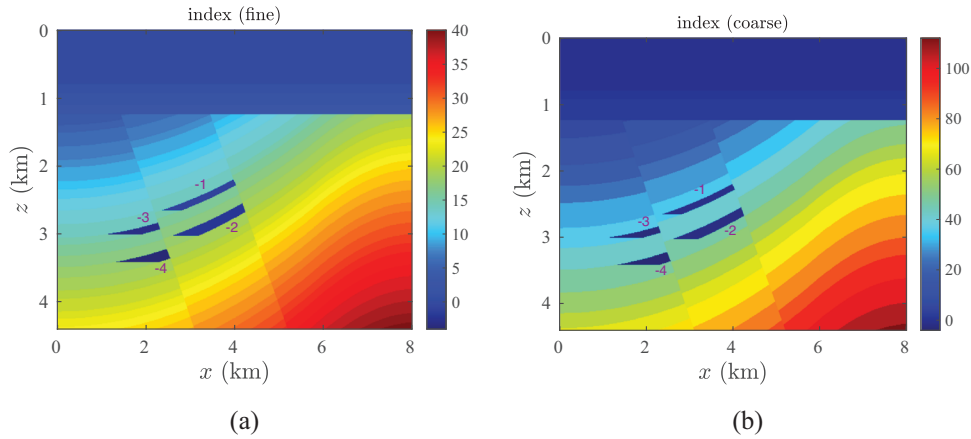
Figure 10 | The index map (a) defines the piecewise constant values per layer for the model parameters in Figure 9. The negative indices correspond to 4 reservoirs. A coarser version (b) is obtained by pairwise combinations of layers, excluding seawater, top layer, and reservoirs.


Figure 11 | Standard deviations on a logarithmic scale for the marginal distributions for components (a) $\delta \log I_p$, (b) $\delta \log v_p$, and (c) $\delta \log(v_s/v_p)$. Values may be clipped at the extrema of the colour scale.

with index 41 and 44, deeper below it, corresponding to the central part of the model in between the two faults and the reservoirs with index −1 and −2. The image depicts $(\mathcal{X}_{unc}/\mathcal{X}_{noise})^{1/}$ as a function of the two model parameters, ignoring other model parameters. The white ellipse is the boundary of the uncertainty region for a noise energy $\mathcal{X}_{noise}$ taken as 10^{-8} of the data energy $\mathcal{X}_0$. It follows from the singular value decomposition $\mathbf{H} = \mathbf{USU}^\top$, with singular values $\mathbf{s} = \mathrm{diag}(\mathbf{S})$ in the diagonal matrix $\mathbf{S}$, and setting $\mathcal{X}_{unc} = \frac{1}{2}\mathbf{y}_H^\top\mathbf{y}_H$ with $\mathbf{y}_H = \mathbf{S}^{1/}\mathbf{U}^\top\delta\mathbf{m}$

and $\delta\mathbf{m} = \mathbf{U}(\mathbf{S}^\dagger)^{1/}\mathbf{y}_H$, similar to Equation (7). The ellipsoid is parameterised by $\mathbf{y}_H$ on a high-dimensional sphere with a radius $(2\mathcal{X}_{noise})^{1/}$.

Similarly, the covariance matrix $\mathbf{C} = \mathbf{US}^\dagger\mathbf{U}^\top$, with $\mathbf{S}^\dagger$ the pseudo-inverse of $\mathbf{S}$, determines ellipses defined by constant values of $\mathbf{m}^\top\mathbf{Cm} = \mathbf{y}_C^\top\mathbf{y}_C$, with $\mathbf{y}_C = (\mathbf{S}^\dagger)^{1/}\mathbf{U}^\top\mathbf{m}$ on a hypersphere and $\mathbf{m} = \mathbf{US}^{1/}\mathbf{y}_C$. The magenta ellipse in Figure 16 corresponds to a 2×2 subset of $\mathbf{C}$ and represents the marginal distribution of the two
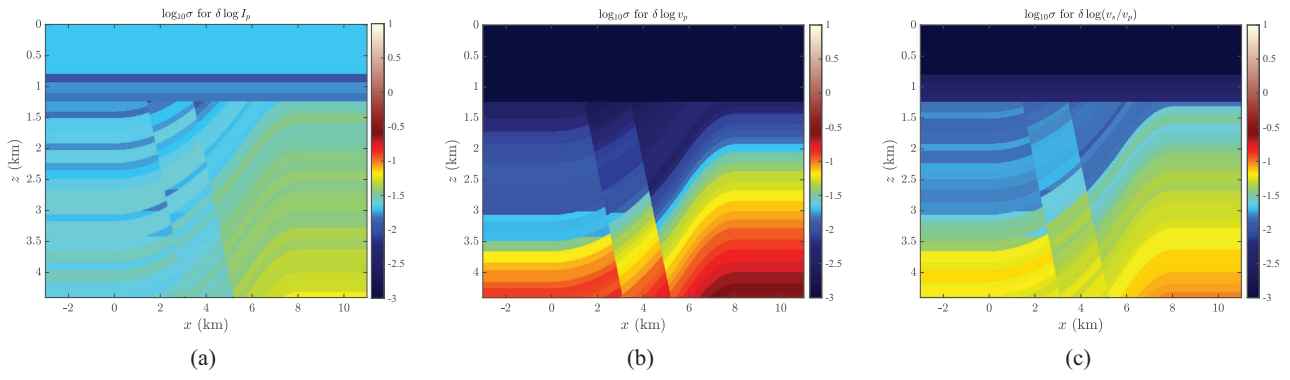
Figure 12 | As Figure 11, but for the conditional distributions of (a) $\delta \log I_p$, (b) $\delta \log v_p$, and (c) $\delta \log(v_s/v_p)$.
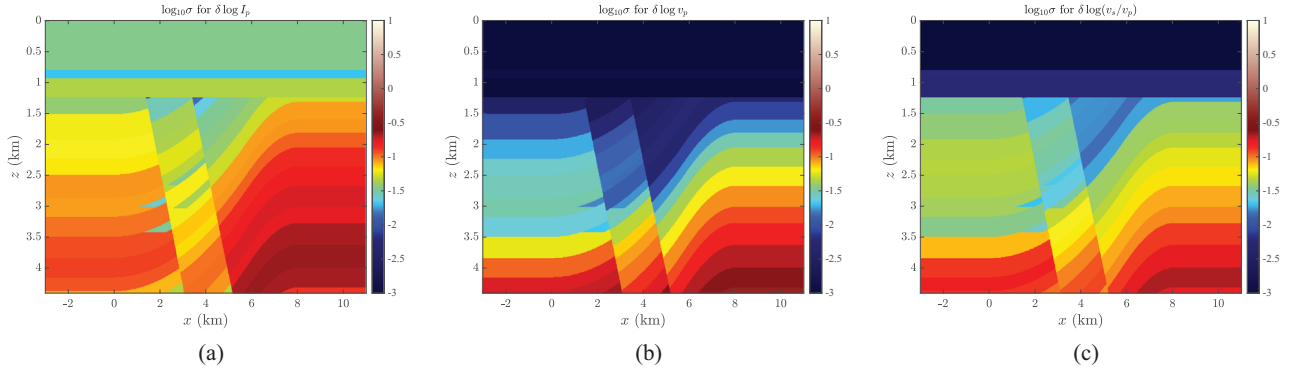
Figure 13 | As the marginal distributions of Figure 11, but for coarser units.
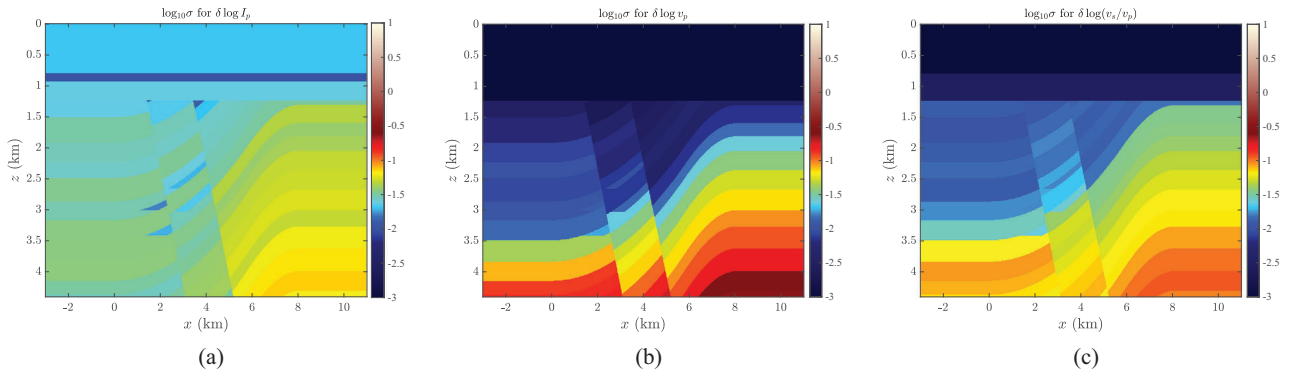
Figure 14 | As the conditional distributions of Figure 12, but for coarser units.
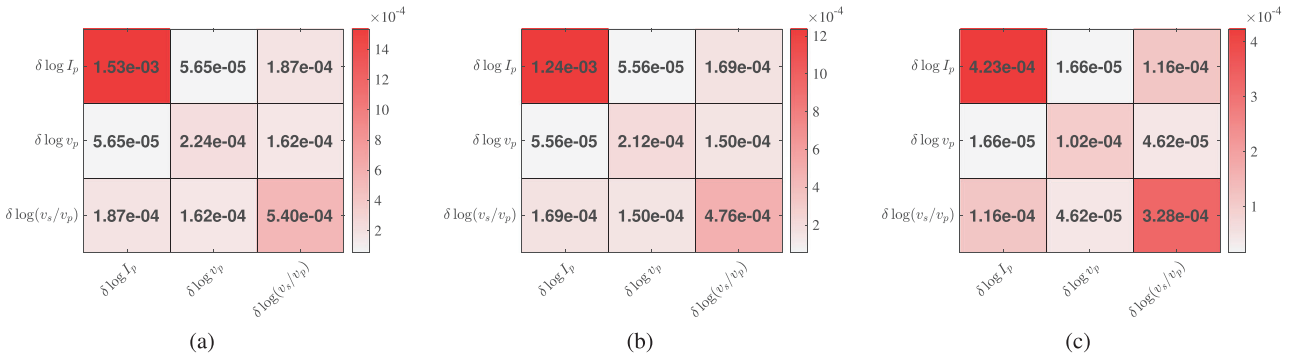
Figure 15 | Marginal covariance matrices for the reservoir with index −1 in the fine (a) and coarse (b) case. For the conditional case (c), fine and coarse results are the same, by definition.

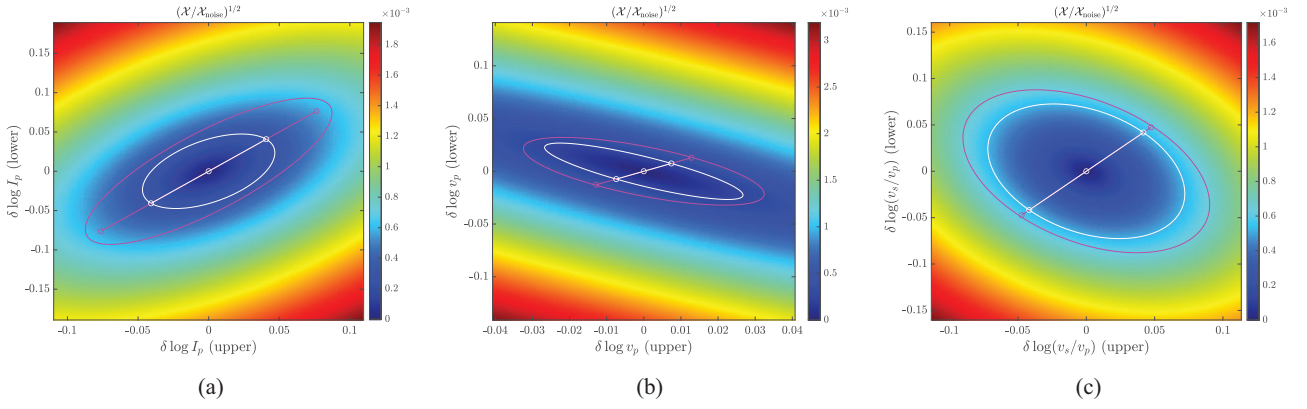**Figure 16** | Uncertainty for a pair of parameters in two adjacent layers, for (a) $\delta \log I_p$, (b) $\delta \log v_p$ and (c) $\delta \log(v_s/v_p)$, scaled by $\mathcal{X}_{\text{noise}} = \epsilon' \mathcal{X}_0$. The white ellipses correspond to the value of σ for the conditional uncertainty and the magenta ellipses to the marginal one. The projection to single parameters is shown as a line segment.

parameters. The latter has a slightly different orientation of the axes.

In the example of Figure 16, units 41 and 44 below it can be taken together, leaving a single value for each of the three elastic model components. The projection operator in this case is a subset of $\mathbf{Q}_{(1)}^{(2)}$ that combines units 41 (upper) and 44 (lower). It can be expressed as the first row of the transposed rotation matrix

$$\mathbf{R}^{\mathsf{T}} = \begin{pmatrix} 0.7037 & 0.7105 \\ -0.7105 & 0.7037 \end{pmatrix},$$

related to an angle of 45.28°. The white line segments in Figures 16a–c describe the range of uncertainty for the conditional distribution. The line segment and ellipse provide another illustration of Poincaré's separation theorem. The line segment is a 1D subset of the area inside the ellipse, with a length between the short and long axis of the ellipse. In the conditional case, the endpoints coincide with points on the ellipse. The projection for the marginal distribution does not necessarily coincide with the ellipse, as the inverse of the projection is not the same as the projection of the inverse, although the endpoints of the line segment are nevertheless very close to points on the magenta ellipse in Figures 16b and 16c but not in Figure 16a.

5 | Discussion and Conclusions

We have considered full-waveform inversion (FWI) uncertainty as the range of model parameters within which the sensitivity of the modelled data to parameter variations does not exceed the noise present in the observed data. This sensitivity is estimated with noise-free forward modelling and is characterised by the Hessian of a cost or loss function. Full characterisation of noise in the observed data requires specifying its covariance matrix. We show that for our purpose the overall noise energy level, which is assumed to be known and which does not influence the sensitivity estimation, is a sufficient proxy. Computation of the full Hessian is usually not feasible in practice. Moreover, due to a key feature of FWI – the use of grid spacings significantly smaller than the wavelengths of the modelled waves – the Hessian is inevitably singular. This leads to formally infinite uncertainty in directions

corresponding to parameter perturbations that lie in the null space, that is, combinations that do not affect the modelled data. The null-space may occupy 80% or more of the full model space. To obtain meaningful and finite uncertainty estimates, it is necessary to project the Hessian onto a lower-dimensional subspace. As a consequence, FWI uncertainty estimates are not absolute but relative, depending on the choice of dimensionality reduction or projection approach.

We developed a formalism to find lower-dimensional projections of the Hessian, taking into account that its pseudo-inverse is generally required. A number of examples show that the projection removes a part or all of the null-space components. When a finite-difference method with a fixed grid spacing is used for modelling and inversion, the grid is typically too dense in the deeper parts of a model where the velocities are higher. Suppression of the null-space components related to sub-resolution structures is therefore necessary. If the scale length after projection is relatively small, the eigenvalues of the compressed Hessian are distributed following the same pattern as the eigenvalues of the original Hessian. However, if the length scale after projection becomes too large, the spectrum of the Hessian and the related uncertainty estimates will be distorted. Apparently, this happens at scales larger than the size of the Fresnel zone where the variations of parameters at the grid points that belong to the same geological unit influence the perturbation of the measured signal incoherently.

The examples show that the proposed approach provides reasonable estimates of the conditional uncertainty. When confined to a subset of the domain, partially conditional estimates can be useful to quantify relative uncertainties for multi-parameter inversion. The estimates of the marginal uncertainty, which are affected by the global effect of the model parameters, are less reliable because the inverse of the projected Hessian differs from the projection of the inverse Hessian, required for computing the marginal uncertainty.

We have only considered the simplest projection operator based on averaging and representing the relative model perturbations as piecewise constant per geological unit. The relation to the Walsh–Hadamard transform and its generalisations would facilitate the

selection of additional components, other than long wavelength structures. Also, smoother representations replacing the current blocky choice could be useful (Loris et al. 2007; Simons et al. 2011, for instance).

We envisage a workflow that starts, possibly, with an initial velocity model. FWI on successively finer length scales and higher frequencies provides a subsurface model (Bunks et al. 1995). Seismic interpretation leads to a segmentation in geological units. Automatic horizon tracking can provide a space-filling partitioning into subsurface volumes and pattern recognition can assist in combining those into units of similar rock type. Nevertheless, segmentation of highly heterogeneous three-dimensional models obtained by FWI can be a challenge. Once accomplished, the Hessian follows from constant relative perturbations of each model parameter per unit, even if the parameters vary per unit, at the cost of a forward simulation of the seismic dataset. Since inversion requires $O(100)$ iterations for $O(1)$ model parameters per point or per set of points, this should be feasible for $O(100)$ geological units, either of rather large size for the full model, or much smaller in target-oriented applications, or a combination of both.

An alternative, elegant approach combines segmentation and inversion (Bodin et al. 2009; Burdick and Lekić 2017; Guo et al. 2020; Hawkins and Sambridge 2015; Malinverno and Leaney 2005; D. Zhu and Gibson 2018). An application for FWI (Ray et al. 2017) uses a quad tree to represent the 2D subsurface model by the Haar wavelet basis and a reversible-jump Markov-chain Monte Carlo method to sample the posterior model. Its disadvantages are compute cost and a blocky model representation. Other ways to reduce compute cost are operator upscaling (Stuart et al. 2019) and homogenisation (Cao et al. 2024; Capdeville and Métivier 2018; Cupillard and Capdeville 2018; Gibson et al. 2014; Owhadi and Zhang 2008).

The analysis based on the Hessian is limited to small perturbations around the global minimum. Uncertainty quantification for cases where the reference model selected for the analysis is not close to the global minimum requires other tools, several of which are mentioned in the Introduction.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analysed in this study. However, if the computational results shown in the figures are considered as new data, then the authors elect not to share those.

References

Aitken, A. C. 1935. "On Least Squares and Linear Combinations of Observations." *Proceedings of the Royal Society of Edinburgh* 55: 42–48. https://doi.org/10.1017/S0370164600014346

Albert, A. 1972. *Regression and the Moore-Penrose Pseudoinverse*. Academic Press.

Backus, G., and F. Gilbert. 1970. "Uniqueness in the Inversion of Inaccurate Gross Earth Data." *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 266, no. 1173: 123–192. https://doi.org/10.1098/rsta.1970.0005

Barbosa, C. H., L. N. Kunstmann, R. M. Silva, C. D. Alves, B. S. Silva, D. M. Filho, M. Mattoso, F. A. Rochinha, and A. L. Coutinho. 2020. "A Workflow for Seismic Imaging With Quantified Uncertainty" *Computers & Geosciences* 145: 104615. https://doi.org/10.1016/j.cageo.2020.104615

Ben-Artzi, G., H. Hel-Or, and Y. Hel-Or. 2007. "The Gray-Code Filter Kernels" *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, no. 3: 382–393. https://dx.doi.org/10.1109/TPAMI.2007.62

Betancourt, M. 2018. "A Conceptual Introduction to Hamiltonian Monte Carlo." Preprint, arXiv, July 16. https://doi.org/10.48550/arXiv.1701.02434

Bharadwaj, P., W. Mulder, and G. Drijkoningen. 2016. "Full Waveform Inversion With An Auxiliary Bump Functional" *Geophysical Journal International* 206, no. 2: 1076–1092. https://doi.org/10.1093/gji/ggw129

Biswas, R., M. Walker, J. Zhang, P. Paramo, K. Wolf, S. Gerth, J. Winterbourne, A. Roy, P. Morris, C. Decalf, Y. Zheng, and R. Warnick. 2023. "Bayesian AVA Elastic Seismic Inversion Using Stein Variational Gradient Descent (SVGD)." In *Conference Proceedings, 84th EAGE Annual Conference & Exhibition*, vol. 2023, 1–5. European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.202310835

Bodin, T., M. Sambridge, and K. Gallagher. 2009. "A Self-Parametrizing Partition Model Approach to Tomographic Inverse Problems" *Inverse Problems* 25, no. 5: 55009. https://doi.org/10.1088/0266-5611/25/5/055009

Bozdağ, E., J. Trampert, and J. Tromp. 2011. "Misfit Functions for Full Waveform Inversion Based on Instantaneous Phase and Envelope Measurements" *Geophysical Journal International* 185, no. 2: 845–870. https://doi.org/10.1111/j.1365-246X.2011.04970.x

Bui-Thanh, T., C. Burstedde, O. Ghattas, J. Martin, G. Stadler, and L. C. Wilcox. 2012. "Extreme-Scale UQ or Bayesian Inverse Problems Governed by PDEs." In *2012 International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–11. IEEE. https://doi.org/10.1109/SC.2012.56

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent. 1995. "Multiscale Seismic Waveform Inversion" *Geophysics* 60, no. 5: 1457–1473. https://doi.org/10.1190/1.1443880

Burdick, S., and V. Lekić. 2017. "Velocity Variations and Uncertainty From Transdimensional P-Wave Tomography of North America" *Geophysical Journal International* 209, no. 2: 1337–1351. https://doi.org/10.1093/gji/ggx091

Cao, J., R. Brossier, Y. Capdeville, L. Métivier, and S. Sambolian. 2024. "A Fully Scalable Homogenization Method to Upscale 3-D Elastic Media" *Geophysical Journal International* 238, no. 1: 72–90. https://doi.org/10.1093/gji/ggae132

Capdeville, Y., and L. Métivier. 2018. "Elastic Full Waveform Inversion Based on the Homogenization Method: Theoretical Framework and 2-D Numerical Illustrations" *Geophysical Journal International* 213, no. 2: 1093–1112. https://doi.org/10.1093/gji/ggy039

Chen, B., and X.-B. Xie. 2015. "An Efficient Method for Broadband Seismic Illumination and Resolution Analyses." In *SEG Technical Program Expanded Abstracts 2015*, 4227–4231. Society of Exploration Geophysicists. https://doi.org/10.1190/segam2015-5926976.1

Cupillard, P., and Y. Capdeville. 2018. "Non-Periodic Homogenization of 3-D Elastic Media for the Seismic Wave Equation" *Geophysical Journal International* 213, no. 2: 983–1001. https://doi.org/10.1093/gji/ggy032

Deal, M. M., and G. Nolet. 1996. "Null-Space Shuttles" *Geophysical Journal International* 124, no. 2: 372–380. https://doi.org/10.1111/j.1365-246X.1996.tb07027.x

Duane, S., A. Kennedy, B. J. Pendleton, and D. Roweth. 1987. "Hybrid Monte Carlo" *Physics Letters B* 195, no. 2: 216–222. https://doi.org/10.1016/0370-2693(87)91197-X

Eckart, C., and G. Young. 1936. "The Approximation of One Matrix by Another of Lower Rank" *Psychometrika* 1, no. 3: 211–218. https://doi.org/10.1007/BF02288367

Eikrem, K. S., G. Nævdal, and M. Jakobsen. 2019. "Iterated Extended Kalman Filter Method for Time-Lapse Seismic Full-Waveform Inversion" *Geophysical Prospecting* 67, no. 2: 379–394. https://doi.org/10.1111/1365-2478.12730

Ely, G., A. Malcolm, and O. V. Poliannikov. 2018. "Assessing Uncertainties in Velocity Models and Images With a Fast Nonlinear Uncertainty Quantification Method" *Geophysics* 83, no. 2: R63–R75. https://doi.org/10.1190/geo2017-0321.1

Engquist, B., and Y. Yang. 2022. "Optimal Transport Based Seismic Inversion: Beyond Cycle Skipping" *Communications on Pure and Applied Mathematics* 75, no. 10: 2201–2244. https://doi.org/10.1002/cpa.21990

Fichtner, A., and J. Trampert. 2011a. "Hessian Kernels of Seismic of Seismic Data Functional Based Upon Adjoint Techniques" *Geophysical Journal International* 185, no. 2: 775–798. https://doi.org/10.1111/j.1365-246X.2011.04966.x

Fichtner, A., and J. Trampert. 2011b. "Resolution Analysis in Full Waveform Inversion" *Geophysical Journal International* 187, no. 3: 1604–1624. https://doi.org/10.1111/j.1365-246X.2011.05218.x

Fichtner, A., and T. vanLeeuwen. 2015. "Resolution Analysis By Random Probing" *Journal of Geophysical Research: Solid Earth* 120, no. 8: 5549–5573. https://doi.org/10.1002/2015JB012106

Fichtner, A., and A. Zunino. 2019. "Hamiltonian Nullspace Shuttles" *Geophysical Research Letters* 46, no. 2: 644–651. https://doi.org/10.1029/2018GL080931

Fino, B. J., and V. R. Algazi. 1976. "Unified Matrix Treatment of the Fast Walsh-Hadamard Transform" *IEEE Transactions on Computers* C-25, no. 11: 1142–1146. https://doi.org/10.1109/TC.1976.1674569

Gfeller, D., and P. D. L. Rios. 2007. "Spectral Coarse Graining of Complex Networks" *Physical Review Letters* 99: 038701. https://doi.org/10.1103/PhysRevLett.99.038701

Gibson, R. L., K. Gao, E. Chung, and Y. Efendiev. 2014. "Multiscale Modeling of Acoustic Wave Propagation in 2D Media" *Geophysics* 79, no. 2: T61–T75. https://doi.org/10.1190/geo2012-0208.1

Gradshteyn, I. S., and I. M. Ryzhik. 2000. *Table of Integrals, Series, and Products*. 6th ed. Academic Press.

Guo, P., S. Singh, V. A. Vaddineni, G. Visser, I. Grevemeyer, and E. Saygin. 2020. "Nonlinear Full Waveform Inversion of Wide-Aperture OBS Data for Moho Structure Using a Trans-Dimensional Bayesian Method" *Geophysical Journal International* 224, no. 2: 1056–1078. https://doi.org/10.1093/gji/ggaa505

Hackbusch, W. 1985. *Multi-Grid Methods and Applications*. Springer. https://doi.org/10.1007/978-3-662-02427-0

Hak, B., and W. A. Mulder. 2010. "Migration for Velocity and Attenuation Perturbations" *Geophysical Prospecting* 58, no. 6: 939–952. https://doi.org/10.1111/j.1365-2478.2010.00866.x

Hanson, A. J. 1995. II.4 - 4 Rotations for *N*-Dimensional Graphics. In: Paeth A. W., ed. *Graphics Gems V*, pp. 55–64. Academic Press. https://doi.org/10.1016/B978-0-12-543457-7.50017-6

Hawkins, R., and M. Sambridge. 2015. "Geophysical Imaging Using Trans-Dimensional Trees" *Geophysical Journal International* 203, no. 2: 972–1000. https://doi.org/10.1093/gji/ggv326

Hoffmann, A., R. Brossier, L. Métivier, and A. Tarayoun. 2024. "Local Uncertainty Quantification for 3-D Time-Domain Full-Waveform Inversion With Ensemble Kalman Filters: Application to a North Sea OBC Data Set" *Geophysical Journal International* 237, no. 3: 1353–1383. https://doi.org/10.1093/gji/ggae114

Huang, X. 2023. "Full Wavefield Inversion with Multiples: Nonlinear Bayesian Inverse Multiple Scattering Theory Beyond the Born Approximation" *Geophysics* 88, no. 6: T289–T303. https://doi.org/10.1190/geo2022-0604.1

Huang, X., K. S. Eikrem, M. Jakobsen, and G. Nævdal. 2020. "Bayesian Full-Waveform Inversion in Anisotropic Elastic Media Using the Iterated Extended Kalman Filter" *Geophysics* 85, no. 4: C125–C139. https://doi.org/10.1190/geo2019-0644.1

Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata. 1990. "Whole Mantle P-Wave Travel Time Tomography" *Physics of the Earth and Planetary Interiors* 59, no. 4: 294–328. https://doi.org/10.1016/0031-9201(90)90236-Q

Izzatullah, M., T. Alkhalifah, J. Romero, M. Corrales, N. Luiken, and M. Ravasi. 2023. "Plug-and-Play Stein Variational Gradient Descent for Bayesian Post-Stack Seismic Inversion." In *Proceedings of the 84th EAGE Annual Conference & Exhibition*, vol. 2023, 1–5. European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.202310177

Jiao, K., D. Sun, X. Cheng, and D. Vigh. 2015. "Adjustive Full Waveform Inversion." In *SEG Technical Program Expanded Abstracts 2015*, 1091–1095. Society of Exploration Geophysicists. https://doi.org/10.1190/segam2015-5901541.1

Keating, D., and K. A. Innanen. 2021. "Null-Space Shuttles for Targeted Uncertainty Analysis in Full-Waveform Inversion" *Geophysics* 86, no. 1: R63–R76. https://doi.org/10.1190/GEO2020-0192.1

Lévêque, J.-J., L. Rivera, and G. Wittlinger. 1993. "On the Use of the Checker-Board Test to Assess the Resolution of Tomographic Inversions" *Geophysical Journal International* 115, no. 1: 313–318. https://doi.org/10.1111/j.1365-246X.1993.tb05605.x

Liu, Q., and D. Peter. 2019a. "Square-Root Variable Metric Based Elastic Full-Waveform Inversion–Part 1: Theory and Validation" *Geophysical Journal International* 218, no. 2: 1121–1135. https://doi.org/10.1093/gji/ggz188

Liu, Q., and D. Peter. 2019b. "Square-Root Variable Metric Based Elastic Full-Waveform Inversion–Part 2: Uncertainty Estimation" *Geophysical Journal International* 218, no. 2: 1100–1120. https://doi.org/10.1093/gji/ggz137

Liu, Q., and D. Wang. 2016. "Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm." In *Advances in Neural Information Processing Systems* edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, vol. 29, 2378–2386. Curran Associates. https://dl.acm.org/doi/10.5555/3157096.3157362

Loris, I., G. Nolet, I. Daubechies, and F. A. Dahlen. 2007. "Tomographic Inversion Using ℓ_1-Norm Regularization of Wavelet Coefficients" *Geophysical Journal International* 170, no. 1: 359–370. https://doi.org/10.1111/j.1365-246X.2007.03409.x

Malinverno, A., and S. Leaney. 2005. "A Monte Carlo Method to Quantify Uncertainty in the Inversion of Zero-Offset VSP Data." In *SEG Technical Program Expanded Abstracts 2000*, 2393–2396. SEG. https://doi.org/10.1190/1.1815943

Martin, J., L. C. Wilcox, C. Burstedde, and O. Ghattas. 2012. "A Stochastic Newton MCMC Method for Large-Scale Statistical Inverse Problems With Application to Seismic Inversion" *SIAM Journal on Scientific Computing* 34, no. 3: A1460–A1487. https://doi.org/10.1137/110845598

Meju, M. A. 2009. "Regularized Extremal Bounds Analysis (REBA): An Approach To Quantifying Uncertainty in Nonlinear Geophysical Inverse Problems" *Geophysical Research Letters* 36, no. L03304: 1–5. https://doi.org/10.1029/2008GL036407

Métivier, L., A. Allain, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux. 2018. "Optimal Transport for Mitigating Cycle Skipping in Full-Waveform Inversion: A Graph-Space Transform Approach" *Geophysics* 83, no. 5: R515–R540. https://doi.org/10.1190/geo2017-0807.1

Métivier, L., R. Brossier, J. Virieux, and S. Operto. 2013. "Full Waveform Inversion and the Truncated Newton Method" *SIAM Journal on Scientific Computing* 35, no. 2: B401–B437. https://doi.org/10.1137/16M1093239

Minkoff, S. E. 1996. "A Computationally Feasible Approximate Resolution Matrix for Seismic Inverse Problems" *Geophysical Journal International* 126, no. 2: 345–359. https://doi.org/10.1111/j.1365-246X.1996.tb05295.x

Mulder, W. A. 2021. "Numerical Methods, Multigrid." In *Encyclopedia of Solid Earth Geophysics, Part 12*, edited by H. K. Gupta, 895–900. Encyclopedia of Earth Sciences Series. Springer. https://dx.doi.org/10.1007/978-90-481-8702-7_153

Mulder, W. A., and B. N. Kuvshinov. 2023. "Estimating Large-Scale Uncertainty in the Context of Full-Waveform Inversion." In *Proceedings of the 84th EAGE Annual Conference & Exhibition*, vol. 2023, 1–5. European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.202310304

Mulder, W. A., and B. N. Kuvshinov. 2025. "Accelerating Target-Oriented Multi-Parameter Elastic Full-Waveform Uncertainty Estimation By Reciprocity" *Geophysical Prospecting* 73, no. 1: 38–48. https://doi.org/10.1111/1365-2478.13650

Østmo, S., W. A. Mulder, and R. Plessix. 2002. "Finite-Difference Iterative Migration By Linearized Waveform Inversion in the Frequency Domain." In *SEG Technical Program Expanded Abstracts 2002*, 1384–1387. Society of Exploration Geophysicists. https://doi.org/10.1190/1.1816917

Owhadi, H., and L. Zhang. 2008. "Numerical Homogenization of the Acoustic Wave Equations with a Continuum of Scales" *Computer Methods in Applied Mechanics and Engineering* 198, no. 3: 397–406. https://doi.org/10.1016/j.cma.2008.08.012

Petra, N., and E. W. Sachs. 2021. "Second Order Adjoints in Optimization." In *Numerical Analysis and Optimization*, edited by M. Al-Baali, A. P. Anton, and L. Grandinetti, 209–230. Springer International Publishing. https://doi.org/10.1007/978-3-030-72040-7_10

Piana, A. N., G. Giacomuzzi, and A. Malinverno. 2015. "Local Three-Dimensional Earthquake Tomography by Trans-Dimensional Monte Carlo Sampling" *Geophysical Journal International* 201, no. 3: 1598–1617. https://doi.org/10.1093/gji/ggv084

Plessix, R., and W. A. Mulder. 2004. "Frequency-Domain Finite-Difference Amplitude-Preserving Migration" *Geophysical Journal International* 157, no. 3: 975–987. https://doi.org/10.1111/j.1365-246X.2004.02282.x

Pratt, R. G., C. Shin, and G. J. Hicks. 1998. "Gauss-Newton and Full Newton Methods in Frequency-Space Seismic Waveform Inversion" *Geophysical Journal International* 133, no. 2: 341–362. https://doi.org/10.1046/j.1365-246X.1998.00498.x

Pringle, R. M., and A. A. Rayner. 1971. *Generalized Inverse Matrices With Application to Statistics*. Griffin.

Qu, L., M. Araya-Polo, and L. Demanet. 2024. "Uncertainty Quantification in Seismic Inversion Through Integrated Importance Sampling and Ensemble Methods." Preprint, arXiv, September 10. https://doi.org/10.48550/arXiv.2409.06840

Rawlinson, N., A. Fichtner, M. Sambridge, and M. K. Young. 2014. *Chapter One - Seismic Tomography and the Assessment of Uncertainty*, 1–76. vol. 55 Advances in Geophysics. Elsevier. https://doi.org/10.1016/bs.agph.2014.08.001

Ray, A., S. Kaplan, J. Washbourne, and U. Albertin. 2017. "Low Frequency Full Waveform Seismic Inversion within a Tree Based Bayesian Framework" *Geophysical Journal International* 212, no. 1: 522–542. https://doi.org/10.1093/gji/ggx428

Revelo Obando, B. 2018. "Full Waveform Inversion in a MCMC Framework." Master's thesis, Delft University of Technology, Delft, The Netherlands. http://resolver.tudelft.nl/uuid:3232eba7-453d-43a3-a20d-71ee4826f986

Rickett, J. E. 2003. "Illumination-Based Normalization for Wave-Equation Depth Migration" *Geophysics* 68, no. 4: 1371–1379. https://doi.org/10.1190/1.1598130

Riffaud, S., M. A. Fernández, and D. Lombardi. 2024. "A Low-Rank Solver for Parameter Estimation and Uncertainty Quantification in Time-Dependent Systems of Partial Differential Equations" *Journal of Scientific Computing* 99, no. 2. https://doi.org/10.1007/s10915-024-02488-3

Riyanti, C. D., W. A. Mulder, G. Baeten, and R.-E. Plessix. 2008. "An Application of a Novel Frequency-domain Finite-difference Solver to Compute 3D Amplitude-preserving Migration Weights." In *Proceedings of the 70th EAGE Conference and Exhibition incorporating SPE EUROPEC 2008*. European Association of Geoscientists & Engineers. https://doi.org/10.3997/2214-4609.20147692

Rizzuti, G., A. Siahkoohi, P. A. Witte, and F. J. Herrmann. 2020. "Parameterizing Uncertainty By Deep Invertible Networks: An Application To Reservoir Characterization." In *SEG Technical Program Expanded Abstracts 2020*, 1541–1545. Society of Exploration Geophysicists. https://doi.org/10.1190/segam2020-3428150.1

Sen, M. K., and R. Biswas. 2017. "Transdimensional Seismic Inversion Using the Reversible Jump Hamiltonian Monte Carlo Algorithm" *Geophysics* 82, no. 3: R119–R134. https://doi.org/10.1190/geo2016-0010.1

Siahkoohi, A., G. Rizzuti, R. Orozco, and F. J. Herrmann. 2023. "Reliable Amortized Variational Inference With Physics-Based Latent Distribution Correction" *Geophysics* 88, no. 3: R297–R322. https://doi.org/10.1190/geo2022-0472.1

Simons, F. J., I. Loris, G. Nolet, I. C. Daubechies, S. Voronin, J. S. Judd, P. A. Vetter, J. Charléty, and C. Vonesch. 2011. "Solving or Resolving Global Tomographic Models With Spherical Wavelets, and the Scale and Sparsity of Seismic Heterogeneity" *Geophysical Journal International* 187, no. 2: 969–988. https://doi.org/10.1111/j.1365-246X.2011.05190.x

Stuart, G. K., S. E. Minkoff, and F. Pereira. 2019. "A Two-Stage Markov Chain Monte Carlo Method for Seismic Inversion and Uncertainty Quantification" *Geophysics* 84, no. 6: R1003–R1020. https://doi.org/10.1190/GEO2018-0893.1

Sun, J., K. Innanen, and C. Huang. 2021. "Physics-Guided Deep Learning for Seismic Inversion With Hybrid Training and Uncertainty Analysis" *Geophysics* 86, no. 3: R303–R317. https://doi.org/10.1190/geo2020-0312.1

Tarantola, A. 1984. "Linearized Inversion of Seismic Reflection Data" *Geophysical Prospecting* 32, no. 6: 998–1015. https://doi.org/10.1111/j.1365-2478.1984.tb00751.x

Tarantola, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM. https://doi.org/10.1137/1.9780898717921

Thompson, A. 2017. "The Cascading Haar Wavelet Algorithm for Computing the Walsh-Hadamard Transform" *IEEE Signal Processing Letters* 24, no. 7: 1020–1023. https://doi.org/10.1109/LSP.2017.2705247

Thurin, J., R. Brossier, and L. Métivier. 2019. "Ensemble-Based Uncertainty Estimation in Full Waveform Inversion" *Geophysical Journal International* 219, no. 3: 1613–1635. https://doi.org/10.1093/gji/ggz384

van Leeuwen, T., and W. A. Mulder. 2008. "Velocity Analysis Based on Data Correlation" *Geophysical Prospecting* 56, no. 6: 791–803. https://doi.org/10.1111/j.1365-2478.2008.00704.x

Vasco, D. W. 2007. "Invariance, Groups, and Non-Uniqueness: The Discrete Case" *Geophysical Journal International* 168, no. 2: 473–490. https://doi.org/10.1111/j.1365-246X.2006.03161.x

Vasco, D. W., L. R. Johnson, and O. Marques. 2003. "Resolution, Uncertainty, and Whole Earth Tomography" *Journal of Geophysical Research: Solid Earth* 108, no. B1: ESE 9–1–ESE 9–26. https://doi.org/10.1029/2001JB000412

Wang, W., G. A. McMechan, and J. Ma. 2023. "Reweighted Variational Full-Waveform Inversions" *Geophysics* 88, no. 4: R499–R512. https://doi.org/10.1190/geo2021-0766.1

Warner, M., and L. Guasch. 2016. "Adaptive Waveform Inversion: Theory" *Geophysics* 81, no. 6: R429–R445. https://doi.org/10.1190/geo2015-0387.1

Zhang, X., and A. Curtis. 2020. "Variational Full-Waveform Inversion" *Geophysical Journal International* 222, no. 1: 406–411. https://doi.org/10.1093/gji/ggaa170

Zhao, Z., and M. K. Sen. 2021. "A Gradient-Based Markov Chain Monte Carlo Method for Full-Waveform Inversion and Uncertainty Analysis" *Geophysics* 86, no. 1: R15–R30. https://doi.org/10.1190/geo2019-0585.1

Zhu, D., and R. Gibson. 2018. "Seismic Inversion and Uncertainty Quantification Using Transdimensional Markov Chain Monte Carlo Method" *Geophysics* 83, no. 4: R321–R334. https://doi.org/10.1190/geo2016-0594.1

Zhu, H., S. Li, S. Fomel, G. Stadler, and O. Ghattas. 2016. "A Bayesian Approach to Estimate Uncertainty for Full-Waveform Inversion Using A Priori Information From Depth Migration" *Geophysics* 81, no. 5: R307–R323. https://doi.org/10.1190/geo2015-0641.1

APPENDIX A: Calculation of the Hessian

We derive an expression for the Hessian in Hilbert space. To emphasise that we consider a general case that describes continuum or discretised scalar or vector fields and operators or matrices, we do not use boldface symbols as elsewhere in the paper.

A functional $\mathcal{X}(u, m)$, where $u \in U$ and $m \in M$, is minimised under the constraint $s = 0$, where $s \in S$ is defined by the map $\mathcal{F}(u, m)$ from $U \times M \to S$, and U, M and S are inner product (Hilbert) spaces. We assume that for each set of parameters m there exists a unique $u(m)$ that satisfies the above constraint. Then, the problem considered reduces to minimisation of the functional $\mathcal{X}(m) = \mathcal{X}(u(m), m)$.

Following the Lagrangian formalism, or the adjoint state method, we introduce the augmented functional

$$\mathcal{X}_{\text{aug}} = \mathcal{X} + \langle v, s \rangle_S, \tag{A.1}$$

which coincides with $\mathcal{X}$ if $u = u(m)$. Here, $v \in S$ is the so-called adjoint field that plays the role of the Lagrangian multiplier. Unconstrained perturbations δm and δu cause perturbations $\delta \mathcal{X} = \partial_u \mathcal{X} \delta u + \partial_m \mathcal{X} \delta m$ and $\delta s = \partial_u \mathcal{F} \, \delta u + \partial_m \mathcal{F} \delta m$, where ∂_u and ∂_m denote Fréchet derivatives with respect to u and m, respectively. The Riesz representation theorem states that for each linear functional $\mathcal{L}$ on a Hilbert space X there exists a unique element $x_L \in X$ such that $\mathcal{L}x = \langle x_L, x \rangle_X$ for each $x \in X$. Here, the angular brackets $\langle \cdot, \cdot \rangle$ denote the inner product and the subscript 'X' indicates that this inner product is taken in X-space. Taking into account that the Fréchet derivative is a linear functional, applying the Riesz theorem and involving the adjoint relation, we obtain

$$\delta \mathcal{X}_{\text{aug}} = \left\langle \frac{\delta \mathcal{X}}{\delta u} + (\partial_u \mathcal{F})^* v, \delta u \right\rangle_U + \left\langle \frac{\delta \mathcal{X}}{\delta m} + (\partial_m \mathcal{F})^* v, \, \delta m \right\rangle_M. \tag{A.2}$$

Here, $\delta \mathcal{X}/\delta u$ and $\delta \mathcal{X}/\delta m$ are Riesz representations of $\partial_u \mathcal{X}$ and $\partial_m \mathcal{X}$ and '$*$' denotes the adjoint operator. With the choice

$$(\partial_u \mathcal{F})^* v = -\frac{\delta \mathcal{X}}{\delta u}, \tag{A.3}$$

Equation (A.2) reduces to $\delta \mathcal{X}_{\text{aug}} = \langle \delta \mathcal{X}/\delta m, \delta m \rangle_M$, where

$$\frac{\delta \mathcal{X}}{\delta m} = (\partial_m \mathcal{F})^* v + \frac{\delta \mathcal{X}}{\delta m}. \tag{A.4}$$

Since $\delta \mathcal{X}_{\text{aug}} = \delta \mathcal{X}$ for perturbations constrained by the condition $\mathcal{F} = 0$, the value $\delta \mathcal{X}/\delta m$ given by Equation (A.4) is the Riesz representation of $\partial_m \mathcal{X}$. Equations (A.3) and (A.4) constitute the first-order adjoint-state method. Specifying parameters m and solving the equation $\mathcal{F}(u, m) = 0$ for u, one finds the derivatives of $\mathcal{X}$ and $\mathcal{F}$ on the right-hand sides of Equations (A.3) and (A.4). Equation (A.3) can be solved for v and then Equation (A.4) to find $\delta \mathcal{X}/\delta m$.

Further perturbations of m by $\delta m'$, which are independent of δm, cause second-order changes of $\mathcal{X}$ equal to

$$\delta^2 \mathcal{X} = \partial_{mm} \mathcal{X}(m', m) = \left\langle \partial_m \left(\frac{\delta \mathcal{X}}{\delta m} \right) \delta m', \, \delta m \right\rangle_M. \tag{A.5}$$

The second Fréchet derivative $\partial_{mm} \mathcal{X}$ is a bilinear operator, which is called the Hessian. The first term in the angular brackets in Equation (A.5) describes the action of the Hessian on $\delta m'$ in the Riesz representation, and it is the same as $\mathbf{H} \, \delta \mathbf{m}'$ in the notation used elsewhere in the paper. Differentiation of Equation (A.4) provides

$$\partial_m \left(\frac{\delta \mathcal{X}}{\delta m} \right) \delta m' = (\partial_m \mathcal{F})^* \delta v^{\star\prime} + \partial_u \left(\frac{\delta \mathcal{X}}{\delta m} \right) \delta u'$$
$$+ \partial_m \left(\frac{\delta \mathcal{X}}{\delta m} \right) \delta m' + (\partial_{mu} \mathcal{F} \, \delta u')^* v \tag{A.6}$$
$$+ (\partial_{mm} \mathcal{F} \, \delta m')^* v.$$

Here, $\delta v^{\star\prime} = \partial_m v^\star \, \delta m'$ and $\delta u' = \partial_m u \, \delta m'$ are changes in $v^\star$ and u associated with $\delta m'$, $\partial_{mm} \mathcal{X} = \partial_m (\delta \mathcal{X}/\delta m)$ and $\partial_{mu} \mathcal{X} = \partial_m (\delta \mathcal{X}/\delta u)$. Differentiating Equation (A.3), one finds the governing equation for the secondary adjoint field $\delta v^{\star\prime}$,

$$(\partial_u \mathcal{F})^* \delta v^{\star\prime} = -\partial_u \left(\frac{\delta \mathcal{X}}{\delta u} \right) \delta u' - \partial_m \left(\frac{\delta \mathcal{X}}{\delta u} \right) \delta m'$$
$$- (\partial_{um} \mathcal{F} \, \delta m')^* v^\star - (\partial_u \mathcal{F} \, \delta u')^* v^\star. \tag{A.7}$$

Equations (A.6) and (A.7) constitute the second-order adjoint-state method. Such equations have been previously derived by Fichtner and Trampert (2011a) and Métivier et al. (2013) assuming that $\mathcal{X}$ does not depend on m and using the linear approximation for the operator $\mathcal{F}$. Equation (50) of Fichtner and Trampert (2011a) for the correspondent 'Hessian kernel' is recovered by substituting Equation (A.6) into Equation (A.5), involving the adjoint relation and specifying the inner product in the S-space as the integral over space and time. Petra and Sachs (2021) present a more general derivation, applicable to Banach spaces. In Hilbert spaces, their results reduce to Equations (A.5)–(A.7). The values $\delta \mathcal{X}/\delta u$ and v are small near the extrema of $\mathcal{X}$. Neglecting such terms in Equations (A.6) and (A.7) produces the Gauss–Newton approximation for the Hessian action.

APPENDIX B: Minimisation of Combined Cost Functions

Consider a cost function consisting of two terms

$$\mathcal{X}(\mathbf{m}) = \frac{1}{2} (\mathbf{m} - \mathbf{m}_{\text{pr}})^\top \mathbf{H}_{\text{pr}} (\mathbf{m} - \mathbf{m}_{\text{pr}}) + \frac{1}{2} (\mathbf{Fm} - \mathbf{d})^\top \mathbf{W} (\mathbf{Fm} - \mathbf{d}). \tag{B.1}$$

Here, $\mathbf{m}_{\text{pr}}$ is the prior value of the model parameter $\mathbf{m}$, $\mathbf{H}_{\text{pr}}$ is the prior precision matrix and the second term describes the misfit between the linearly modelled data $\mathbf{Fm}$ and the measured data $\mathbf{d}$. The matrices $\mathbf{H}_{\text{pr}}$ and $\mathbf{W}$ are symmetric positive semi-definite, and hence they can be represented as $\mathbf{H}_{\text{pr}} = \mathbf{U}_{\text{pr}}^\top \mathbf{U}_{\text{pr}}$ and $\mathbf{W} = \mathbf{U}_{\text{w}}^\top \mathbf{U}_{\text{w}}$. The gradient of the cost function (B.1) has the form

$$\nabla_\mathbf{m} \mathcal{X}(\mathbf{m}) = \mathbf{H}_\Sigma (\mathbf{m} - \mathbf{m}_{\text{pr}}) + \mathbf{F}^\top \mathbf{W} (\mathbf{F} \mathbf{m}_{\text{pr}} - \mathbf{d}), \tag{B.2}$$

where $\mathbf{H}_\Sigma = \mathbf{H}_{\text{pr}} + \mathbf{H}_I$ and $\mathbf{H}_I = \mathbf{F}^\top \mathbf{WF}$. The least-squares solution of the equation $\nabla_\mathbf{m} \mathcal{X}(\mathbf{m}_\Sigma) = 0$ is

$$\mathbf{m}_\Sigma = \mathbf{m}_{\text{pr}} - \mathbf{K} (\mathbf{F} \mathbf{m}_{\text{pr}} - \mathbf{d}), \tag{B.3}$$

where

$$\mathbf{K} = \mathbf{H}_\Sigma^\dagger \mathbf{F}^\mathsf{T} \mathbf{W} = \mathbf{H}_\Sigma^\dagger \mathbf{U}_I^\mathsf{T} \mathbf{U}_w \tag{B.4}$$

is the Kalman gain and $\mathbf{U}_I = \mathbf{U}_w \mathbf{F}$ is the square root of matrix $\mathbf{H}_I$. In fact, $\nabla_\mathbf{m} \mathcal{X}$ vanishes exactly at $\mathbf{m} = \mathbf{m}_\Sigma$. Since $\mathbf{H}_{pr}$ and $\mathbf{H}_I$ are symmetric positive semi-definite, the null space of $\mathbf{H}_\Sigma$ lies inside the null space of $\mathbf{H}_I$: $\mathcal{N}(\mathbf{H}_\Sigma) \subset \mathcal{N}(\mathbf{H}_I)$. Hence, the range of $\mathbf{H}_I$ lies inside the range of $\mathbf{H}_\Sigma$: $\mathcal{R}(\mathbf{H}_I) \subset \mathcal{R}(\mathbf{H}_\Sigma)$. The ranges of matrices have the properties $\mathcal{R}(\mathbf{AB}) \subset \mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A}^\mathsf{T}\mathbf{A}) = \mathcal{R}(\mathbf{A}^\mathsf{T})$. From these relations, it follows that $\mathcal{R}(\mathbf{F}^\mathsf{T}\mathbf{W}) \subset \mathcal{R}(\mathbf{F}^\mathsf{T}\mathbf{U}_w^\mathsf{T}) = \mathcal{R}(\mathbf{H}_I) \subset \mathcal{R}(\mathbf{H}_\Sigma)$. Since the range of $\mathbf{F}^\mathsf{T}\mathbf{W}$ lies inside the range of $\mathbf{H}_\Sigma$, this matrix does not change after applying the projection operator $\mathbf{P}_\Sigma = \mathbf{H}_\Sigma \mathbf{H}_\Sigma^\dagger$ onto $\mathcal{R}(\mathbf{H}_\Sigma)$. Substitution of $\mathbf{F}^\mathsf{T}\mathbf{W} = \mathbf{P}_\Sigma \mathbf{F}^\mathsf{T}\mathbf{W}$ in Equation (B.2) shows that $\nabla_\mathbf{m}\mathcal{X}(\mathbf{m}_\Sigma) = 0$. We take into account that $\mathbf{H}_I = \mathbf{U}_I^\mathsf{T}\mathbf{U}_I$ and assume that $\mathbf{U}_I^\mathsf{T}$ lies in the column space of $\mathbf{H}_{pr}$, that is, $\mathcal{R}(\mathbf{H}_I) \subset \mathcal{R}(\mathbf{H}_{pr})$ and $\mathbf{H}_{pr}\mathbf{H}_{pr}^\dagger \mathbf{U}_I^\mathsf{T} = \mathbf{U}_I^\mathsf{T}$. In that case, the formula of the pseudo-inverse of a sum of symmetric matrices (Pringle and Rayner 1971), which is a generalisation of the Sherman–Morrison–Woodbury matrix identity, is applicable, and we have

$$\mathbf{H}_\Sigma^\dagger = \left(\mathbf{H}_{pr} + \mathbf{U}_I^\mathsf{T}\mathbf{U}_I\right)^\dagger = \mathbf{H}_{pr}^\dagger - \mathbf{H}_{pr}^\dagger \mathbf{U}_I^\mathsf{T}\left(\mathbf{I} + \mathbf{U}_I\mathbf{H}_{pr}^\dagger\mathbf{U}_I^\mathsf{T}\right)^{-1}\mathbf{U}_I\mathbf{H}_{pr}^\dagger \tag{B.5}$$

with identity matrix $\mathbf{I}$. Then $\mathbf{H}_\Sigma^\dagger \mathbf{U}_I^\mathsf{T} = \mathbf{H}_{pr}^\dagger \mathbf{U}_I^\mathsf{T}\left(\mathbf{I} + \mathbf{U}_I\mathbf{H}_{pr}^\dagger\mathbf{U}_I^\mathsf{T}\right)^{-1}$ and $\mathbf{K} = \mathbf{H}_{pr}^\dagger \mathbf{U}_I^\mathsf{T}\left(\mathbf{I} + \mathbf{U}_I\mathbf{H}_{pr}^\dagger\mathbf{U}_I^\mathsf{T}\right)^{-1}\mathbf{U}_w$. Equation (B.5) can be written in terms of the Kalman gain and the covariance matrices as

$$\mathbf{C}_\Sigma = (\mathbf{I} - \mathbf{K}\mathbf{F})\mathbf{C}_{pr}, \tag{B.6}$$

where $\mathbf{C}_\Sigma = \mathbf{H}_\Sigma^\dagger$ and $\mathbf{C}_{pr} = \mathbf{H}_{pr}^\dagger$. Assuming that $\mathbf{W} = \mathbf{U}_w^\mathsf{T}\mathbf{U}_w$ is non-singular, the Kalman gain can be cast into the standard form

$$\mathbf{K} = \mathbf{C}_{pr}\mathbf{F}^\mathsf{T}\left(\mathbf{W}^{-1} + \mathbf{F}\mathbf{C}_{pr}\mathbf{F}^\mathsf{T}\right)^{-1}. \tag{B.7}$$

APPENDIX C: Partitioning of Symmetric, Positive Semi-definite Matrices

Any symmetric, positive semi-definite matrix $\mathbf{H}$ allows for the representation $\mathbf{H} = \mathbf{U}^\mathsf{T}\mathbf{U}$, where $\mathbf{U}$ is a positive semi-definite matrix, which is called the square root of $\mathbf{H}$. Using the partitioning $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2,)$ we cast $\mathbf{H}$ in the block form,

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1^\mathsf{T}\mathbf{U}_1 & \mathbf{U}_1^\mathsf{T}\mathbf{U}_2 \\ \mathbf{U}_2^\mathsf{T}\mathbf{U}_1 & \mathbf{U}_2^\mathsf{T}\mathbf{U}_2 \end{pmatrix}. \tag{C.1}$$

Taking into account the properties of pseudo-inverse

$$\left(\mathbf{A}^\mathsf{T}\mathbf{A}\right)^\dagger = \mathbf{A}^\dagger\left(\mathbf{A}^\mathsf{T}\right)^\dagger, \quad \mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\mathsf{T}, \quad \mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}, \tag{C.2}$$

one finds $\mathbf{H}_{11}\mathbf{H}_{11}^\dagger = \mathbf{U}_1^\mathsf{T}\mathbf{U}_1(\mathbf{B}_1^\mathsf{T}\mathbf{U}_1)^\dagger = \left(\mathbf{U}_1^\mathsf{T}\mathbf{U}_1\mathbf{U}_1^\dagger\right)\left(\mathbf{U}_1^\mathsf{T}\right)^\dagger = \mathbf{U}_1^\mathsf{T}\left(\mathbf{U}_1^\mathsf{T}\right)^\dagger$ and (see Theorem 9.1.6 of Albert 1972)

$$\mathbf{H}_{11}\mathbf{H}_{11}^\dagger\mathbf{H}_{12} = \left[\mathbf{U}_1^\mathsf{T}\left(\mathbf{U}_1^\mathsf{T}\right)^\dagger\mathbf{U}_1^\mathsf{T}\right]\mathbf{U}_2 = \mathbf{U}_1^\mathsf{T}\mathbf{U}_2 = \mathbf{H}_{12}. \tag{C.3}$$

Equation (C.3) implies $\mathcal{R}(\mathbf{H}_{12}) \subset \mathcal{R}(\mathbf{H}_{11})$. Using Equation (C.3) one checks that the matrix $\mathbf{H}$ is block-diagonalised by the transformation

$$\overline{\mathbf{H}} = \mathbf{S}^\mathsf{T}\mathbf{H}\mathbf{S} = \begin{pmatrix} \mathbf{H}_{11} & 0 \\ 0 & \mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^\dagger\mathbf{H}_{12} \end{pmatrix}, \tag{C.4}$$

where $\mathbf{S}$ is a non-singular matrix,

$$\mathbf{S} = \begin{pmatrix} \mathbf{I} & -\mathbf{H}_{11}^\dagger\mathbf{H}_{12} \\ 0 & \mathbf{I} \end{pmatrix}, \quad \mathbf{S}^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{H}_{11}^\dagger\mathbf{H}_{12} \\ 0 & \mathbf{I} \end{pmatrix}. \tag{C.5}$$

From Equation (C.4), it follows that $\mathbf{m}^\mathsf{T}\mathbf{H}\mathbf{m} = \overline{\mathbf{m}}^\mathsf{T}\overline{\mathbf{H}}\,\overline{\mathbf{m}}$, where $\overline{\mathbf{m}} = \mathbf{S}^{-1}\mathbf{m}$. Partitioning the vector $\mathbf{m}$ as $\mathbf{m} = (\mathbf{m}_1, \mathbf{m}_2)^\mathsf{T}$ we get

$$\mathbf{m}^\mathsf{T}\mathbf{H}\mathbf{m} = \overline{\mathbf{m}}_1^\mathsf{T}\mathbf{H}_{11}\overline{\mathbf{m}}_1 + \mathbf{m}_2^\mathsf{T}\overline{\mathbf{H}}_{22}\mathbf{m}_2. \tag{C.6}$$

Here, $\overline{\mathbf{m}}_1 = \mathbf{m}_1 + \mathbf{H}_{11}^\dagger\mathbf{H}_{12}\mathbf{m}_2$ and $\overline{\mathbf{H}}_{22} = \mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^\dagger\mathbf{H}_{12}$ is the pseudo-Schur complement of $\mathbf{H}$ with respect to $\mathbf{H}_{11}$.

APPENDIX D: Back to the Modelling Grid

In general, there is no obvious relation between the conditional and marginal uncertainties obtained from the compressed Hessian and those from the Hessian for the modelling grid. Nevertheless, we may consider three options to map the compressed results back to the original modelling grid: based on $\mathbf{Q}^\mathsf{T}$, just copy, or using the projected $\mathbf{H}_p$. A pictorial description for the two-parameter case is provided to illustrate the difference between the resulting uncertainties.

Figure A1 shows the ellipses described by $\mathcal{X} = \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^\mathsf{T}\mathbf{H}(\mathbf{m} - \mathbf{m}_0) = \varepsilon_0$ for two parameters, $\mathbf{m} = (m_1, m_2)^\mathsf{T}$, with $\varepsilon_0 = 1$, and with Hessians given by, respectively,

$$\mathbf{H} = \begin{pmatrix} 0.63 & 0.36 \\ 0.36 & 0.90 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 1.05 & -0.60 \\ -0.60 & 1.50 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 8.0 & 5.5 \\ 5.5 & 4.3 \end{pmatrix},$$
$$\mathbf{H} = \begin{pmatrix} 8.0 & -3.0 \\ -3.0 & 1.5 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 0.4 & 0.2 \\ 0.2 & 4.0 \end{pmatrix}, \mathbf{H} = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.8 \end{pmatrix}. \tag{D.1}$$
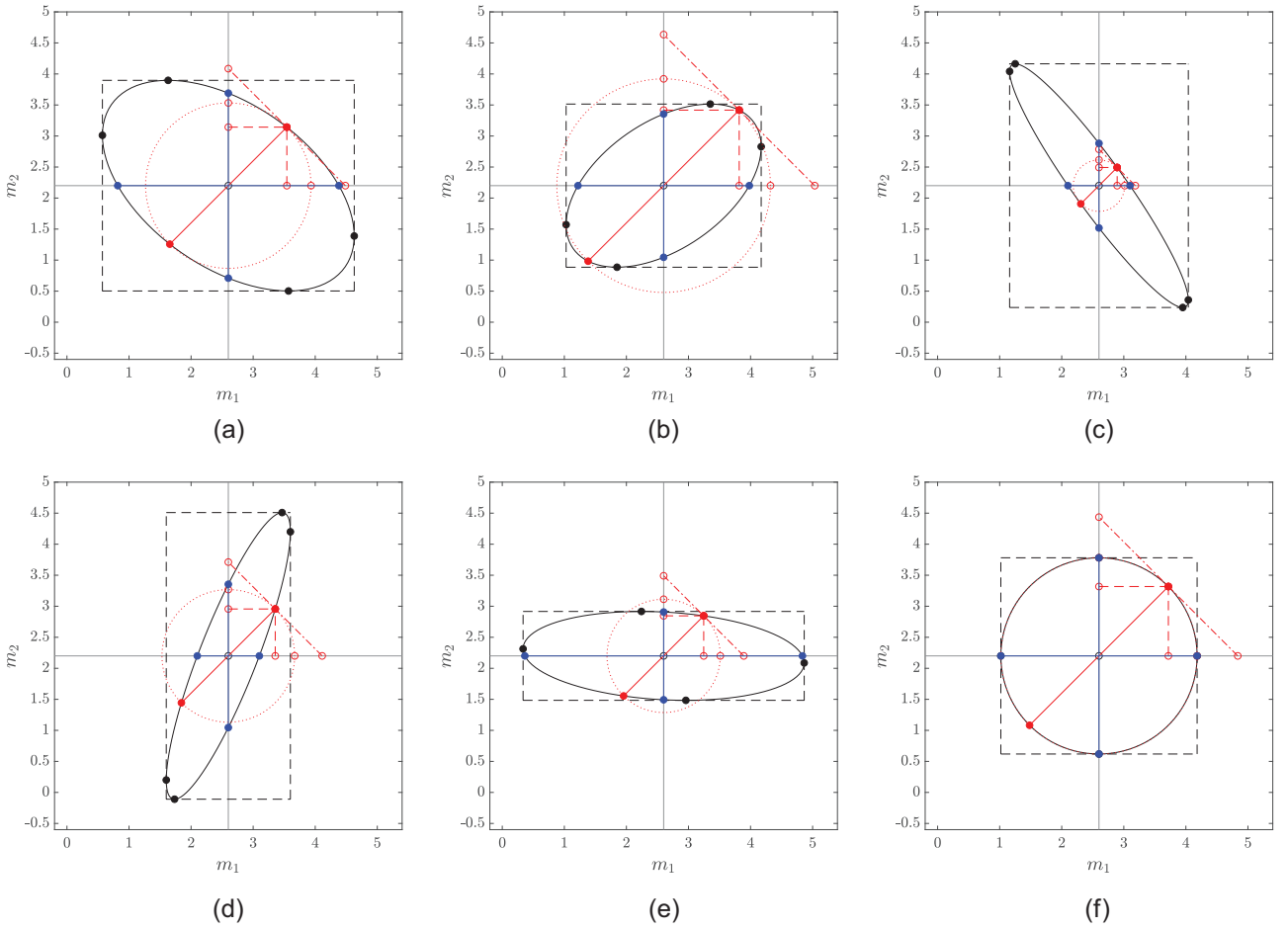
The conditional uncertainties defined by $\mathcal{X} \leq \varepsilon_0$ are $|m_j - m_{0,j}| \leq [2\varepsilon_0/\mathrm{diag}(\mathbf{H})]^{1/}$, $j = 1, 2$, and are drawn as blue line segments, ending at the blue dots. The marginal uncertainties are given by $|m_j - m_{0,j}| \leq [2\varepsilon_0\mathrm{diag}(\mathbf{H}^\dagger)]^{1/}$. The corresponding line segments are the horizontal and vertical lines through the minimum at the centre of the ellipse, bounded by the dashed black bounding box of the ellipse. The intersection points of the bounding box with the ellipse are marked by black dots.

The compression matrix for 2 points is $\mathbf{Q} = (1, 1)/\sqrt{2}$. Its application to the model parameters, relative to the minimum, yields a line at 45°. The compressed Hessian is $\mathbf{H}_c = \mathbf{Q}\mathbf{H}\mathbf{Q}^\mathsf{T}$. Since it is a 1×1 matrix in this example, the related conditional and marginal uncertainty are the same: $|m_c - m_{c,0}| \leq [2\varepsilon_0/\mathrm{diag}(\mathbf{H}_c)]^{1/} = [2\varepsilon_0\mathrm{diag}(\mathbf{H}_c^\dagger)]^{1/}$. In the figures, they are indicated by the drawn red line segments. The endpoints always lie inside the bounding box defined by the marginal uncertainties.

How can these results be mapped back to the original coordinates m_1 and m_2? The red endpoints of the line segments can originate from any ellipse passing through them with its origin at the midpoint of the segment and, therefore, this question cannot be answered. However, three options can be considered and are sketched as examples in Figure A1.

(1) Take the coordinates of the compressed result, marked by the red dots: $\mathbf{m} - \mathbf{m}_0 = \mathbf{Q}^\mathsf{T}(\mathbf{m}_c - \mathbf{m}_{c,0})$. This is sketched by the dashed red lines ending at the red open circles. If the original ellipse happens to be identical to the line segment, with a short axis at −45° of zero length, these coordinates determine the marginal uncertainty and the conditional uncertainty vanishes.

Figure A1 | Ellipses defining the uncertainty regions in the 2-parameter case. Conditional uncertainties are indicated by blue line segments, ending at blue dots. The marginal uncertainty in each model parameter determines the bounding boxes, drawn as dashed black rectangles. The red line segments indicate the uncertainty after compression. The endpoints, marked by red dots, are the intersections between the ellipse and a line at 45°. To map the result back to the original parameters, three options can be considered: coordinates of the intersection point following the red dashed line segments, intersection points of a red dotted circle centred at the minimum with the horizontal and vertical lines through the minimum, or their intersections with a red dash-dotted line perpendicular to the one at 45° through the compression result. The first always ends up inside the bounding box of the original ellipse, and the other two may lie inside or outside. In terms of m_1, the three estimates may all lie outside (d) or inside (e) the original ellipse, or some inside and some outside, and similarly for m_2.

(2) Assume the ellipse to be a circle. This is sketched by the dotted circle, intersecting the horizontal and vertical lines through the minimum at the small red open circles. Not all intersections are drawn.

(3) The conditional uncertainty can be based on $\mathbf{H}_p = \mathbf{Q}^\top \mathbf{H}_c \mathbf{Q} = \mathbf{PHP}$ with $\mathbf{P} = \mathbf{Q}^\top \mathbf{Q}$. This is represented by the dash-dotted red lines ending at the red open circles.

The three maps result in endpoints of the conditional uncertainty ranges that are progressively larger. Those for the first option lie always inside the bounding box of the original ellipse. For the other two, they can be inside or outside. Figure A1 depicts several cases, the last one a circle. The points may all end up outside (d) or inside the ellipse (e). With only two parameters, the three estimates differ by a factor of $\sqrt{2}$. With n_f parameters, the factor is $\sqrt{n_f}$ and becomes quite large for $n_f \gg 1$.

APPENDIX E: Low-Storage Use of the Projected Hessian

At first sight, the projected Hessian $\mathbf{H}_p$ requires as much storage as the Hessian on the modelling grid. Here, we explain how this can be avoided

by just using the compressed Hessian $\mathbf{H}_c$. The description is for the single-parameter case, but the generalisation to the multi-parameter is straightforward.

The set of n grid points is partitioned into m segments or geological units. The indicator matrix $\mathbf{X}$ has $x_{j,i} = 1$ if point i lies in segment j, for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Otherwise, $x_{j,i} = 0$. The vector $\mathbf{n}_f$ contains the number of grid points $n_{f,j} = \sum_{i=1}^{n} x_{j,i}$ contained in each segment and $\sum_{j=1}^{m} n_{f,j} = m$. The simplest prolongation operator is a piecewise constant interpolation, which implies just copying and is represented by $\mathbf{X}^\top$. The related restriction operator is $\mathbf{R} = \mathbf{N}_f^{-1} \mathbf{X}$, where $\mathbf{N}_f$ is a diagonal matrix with $\text{diag}(\mathbf{N}_f) = \mathbf{n}_f$ and zero otherwise, representing the arithmetic mean. Its orthonormal version is denoted by $\mathbf{Q} = \mathbf{N}_f^{-1/2} \mathbf{X}$. For an $n \times n$ non-negative symmetric matrix $\mathbf{H}$, the compressed version is defined as $\mathbf{H}_c = \mathbf{QHQ}^\top$ and the projected version as $\mathbf{H}_p = \mathbf{Q}^\top \mathbf{H}_c \mathbf{Q} = \mathbf{PH}_c\mathbf{P}$ with projection matrix $\mathbf{P} = \mathbf{Q}^\top \mathbf{Q}$. With the current restriction and prolongation operators, the operator $\mathbf{P}$ applied to model parameters $\mathbf{m}$ in the single-parameter case replaces the $n_{f,j}$ values in unit j by their average $n_{f,j}^{-1} \sum_{i(j)}^{n_{f,j}} m_i$, where $i(j)$ enumerates the $n_{f,j}$ grid points contained in segment j. In this way, the projection $\mathbf{P}$ acts as a spatial high-cut filter, removing shorter wavelengths. If we define $\mathbf{H}_r = \mathbf{R}\,\mathbf{H}\,\mathbf{R}^\top =$

$\mathbf{N}_f^{-1}\mathbf{X}\mathbf{H}\mathbf{X}^\mathsf{T}\mathbf{N}_f^{-1} = \mathbf{N}_f^{-1/2}\mathbf{H}_c\,\mathbf{N}_f^{-1/2}$ with $\mathbf{H}_c = \mathbf{N}_f^{-1/2}\mathbf{X}\mathbf{H}\mathbf{X}^\mathsf{T}\mathbf{N}_f^{-1/2}$, then $\mathbf{H}_p = \mathbf{X}^\mathsf{T}\mathbf{H}_r\,\mathbf{X}$, showing that $\mathbf{H}_p$ consists of blocks each with $n_{f,j}$ copies of $\mathbf{H}_r$. To find the conditional uncertainty for a single parameter, we fix all parameters, select one value on the main diagonal of the Hessian and solve σ_i from $\frac{1}{2}\sigma_i h_{i,i}\sigma_i = \epsilon'\mathcal{X}_0$, where $\mathcal{X}_0$ is the data energy in the reference or background model. Let the diagonal element $h_{i,i}$ of the original Hessian $\mathbf{H}$ be replaced by that of the projected version $\mathbf{H}_p$. The diagonal of $\mathbf{H}_p$ has groups of $n_{f,j}$ identical values, equal to $(\mathbf{H}_r)_j$. We can, therefore, determine the above conditional σ-values for $\mathbf{H}_r$ instead of $\mathbf{H}_p$ and interpolate them to the original grid with $\mathbf{X}^\mathsf{T}$, that is, $\sigma_i = [2\epsilon'\mathcal{X}_0/(\mathbf{X}^\mathsf{T}\text{diag}(\mathbf{H}_r))_i]^{1/}$. As $\text{diag}(\mathbf{H}_r)$ is a vector, storage requirements for this operation are low.