Estimating large-scale uncertainty in the context of full-waveform inversion

Mulder, W.; Kuvshinov, B.

DOI
10.3997/2214-4609.202310304

Publication date
2023

Document Version
Final published version

Citation (APA)
Mulder, W., & Kuvshinov, B. (2023). *Estimating large-scale uncertainty in the context of full-waveform inversion*. Paper presented at 84th EAGE ANNUAL Conference and Exhibition 2023, Vienna, Austria. https://doi.org/10.3997/2214-4609.202310304

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

2023 EAGE ANNUAL
84TH CONFERENCE & EXHIBITION
VIENNA | AUSTRIA

Estimating large-scale uncertainty in the context of full-waveform inversion

W. Mulder[1,2], B. Kuvshinov[1]

[1] Shell Global Solutions International BV; [2] Delft University of Technology

Summary

The uncertainty of model parameters obtained by full-waveform inversion can be determined from the hessian of the least-squares error functional. 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. A synthetic 2-D isotropic elastic example illustrates how conditional and marginal uncertainties can be estimated for the properties per geological unit by themselves and in relation to other units. A discussion on how the chosen length scale affects the result is included.

Estimating large-scale uncertainty in the context of full-waveform inversion

Introduction

When using a L-BFGS method for gradient minimization of the least-squares data misfit for full-waveform inversion (FWI), it seems natural to use the estimated hessian for uncertainty quantification. However, this only provides the least uncertain components of the model, as already shown and explained by Deal and Nolet (1996) on a linear problem. The computation of the full hessian is feasible for small problems (Pratt et al., 1998, e.g.) but too costly in general. Instead of determining the full hessian for perturbations of each model parameter in each point of a finite-difference grid, a segmented representation is proposed with perturbations of each model parameter per geological unit. The question how the length scale of a unit affects the uncertainty estimates is also addressed.

Method

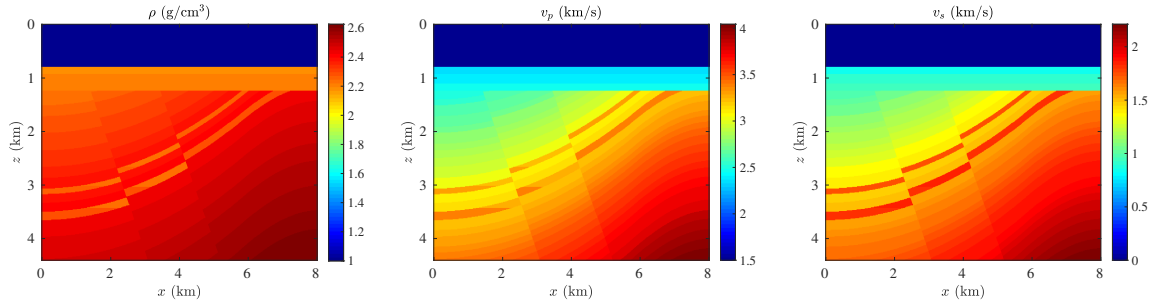
Full-waveform inversion attempts to minimize the least-squares error functional $J = \frac{1}{2}\|\mathbf{F}(\overline{m}) - \mathbf{d}_{\text{obs}}\|^2$, describing the difference between modelled data $\mathbf{F}(\overline{m})$ for a model $\overline{m}$ and observed seismic data $\mathbf{d}_{\text{obs}}$. The model parameters $\mathbf{m}$ can be defined relative to a local minimum $\overline{m}_0$, perhaps coinciding with the global minimum, as $m_{j,k} = (\overline{m}_{j,k} - \overline{m}_{0,j,k})/\overline{m}_{0,j,k}$, with k enumerating all the model components and j all grid points or element nodes, assuming numerical modelling by a finite-difference method or finite-element method. Then, $J = J_0 + \mathbf{g}^{\mathsf{T}}\mathbf{m} + \frac{1}{2}\mathbf{m}^{\mathsf{T}}\mathbf{H}\mathbf{m} + \ldots$ with gradient $\mathbf{g} = \nabla_{\mathbf{m}} J \simeq 0$ near the minimum. The remainder $J_1 = J - \frac{1}{2}\mathbf{m}^{\mathsf{T}}\mathbf{H}\mathbf{m} = J_0 + \mathbf{g}^{\mathsf{T}}\mathbf{m} + \ldots$ contains the noise energy in the data, the noise in the numerical modelling, the inability to reach the global minimum ($\mathbf{g} \neq 0$), deviations from quadratic behaviour near the minimum in the form of higher-order terms, inadequacies of the chosen model, ignored and unknown physics, and so on. The gaussian distribution related to the maximum likelihood, ignoring physical bounds and assuming a 'temperature' T, for a hessian of size n_{h} is (Tarantola, 2005)

$$p(\mathbf{m}) = \sqrt{\frac{\det(\mathbf{H})}{T(2\pi)^{n_{\text{h}}}}}\, e^{-\mathbf{m}^{\mathsf{T}}\mathbf{H}\mathbf{m}/(2T)}.$$

Integrating over all model parameters except m_ℓ and letting $T = 1$ results in the marginal distribution $p(m_\ell) = (\sigma_\ell\sqrt{2\pi})^{-1}\exp[-(m_\ell - m_{0,\ell})^2/(2\sigma_\ell^2)]$, with $\sigma_\ell = \sqrt{C_{\ell,\ell}}$ and covariance matrix $\mathbf{C} = \mathbf{H}^{\dagger}$, the pseudo-inverse of the hessian. The standard deviation σ_ℓ represents the worst-case uncertainty given the chosen model, but neglects the unknown unknowns. The marginal distribution for a subset of the model parameters is obtained by ignoring their complement in the covariance matrix, leaving a covariance matrix of smaller size.

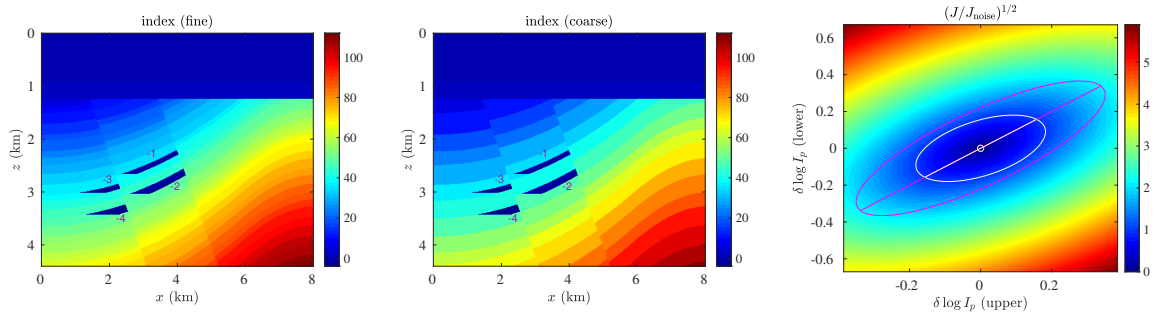
The hessian can be found from the data $\tilde{\mathbf{F}}_{j,k}$ that correspond to a perturbed model $\tilde{m}_{j',k'} = m_{0,j,k}(1 + \varepsilon\,\delta_{j=j'}\delta_{k=k'})$ with Kronecker delta $\delta_{j=j'}$, equalling $\mathbf{m}_0$ everywhere except for model component k at point j. The perturbation data are $\mathbf{F}_{j,k} = \lim_{\varepsilon\to 0}(\tilde{\mathbf{F}}_{j,k} - \mathbf{F}_{0,j,k})/\varepsilon$ and the hessian $\mathbf{H} = \mathbf{F}^{\mathsf{T}}\mathbf{F}$, summed over all receivers and sources. In practice, the perturbation parameter ε should be sufficiently small to make higher-order terms negligible, but large enough to avoid numerical round-off errors. The Born approximation of the elastic wave equation is an alternative, but typically involves the simultaneous solution of two sets of equations, one set of elastic equations for the background model and one for the scattered field. This is more costly than computing the data for the background model only once, but provides cleaner scattering data. In the current setting, the perturbation data contain free-surface and interbed multiples. The model perturbations can be expressed as $\delta\log m_{j,k} = \delta m_{j,k}/m_{0,j,k}$. In this form, the data $\mathbf{F}_{j,k}$ and hessian can be easily transformed to other model representations, for instance, perturbations in P- and S-impedances I_p and I_s and density ρ, or ρ, P-velocity v_p and S-velocity v_s.

Because the hessian is generally too costly to compute and too large to be stored, we make the following simplification. Instead of a dense grid of perturbations, larger regions are considered that we call geological units. Inside a unit, the unperturbed or background model may vary with position $\mathbf{x}$, but the perturbations are considered as piecewise constant, with the same relative change over the whole unit. The geological units can be obtained by segmenting a full-waveform inversion result. This leads to the question how sensitive the result is to the chosen length scale of the geological units. Before addressing that subject, an example is presented.



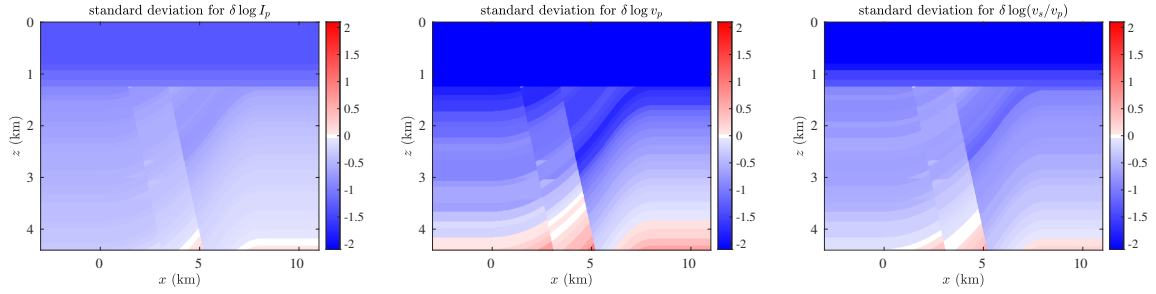
(a) (b) (c)

Figure 1 *Isotropic elastic model with (a) density, (b) P- and (c) S-wave velocity.*



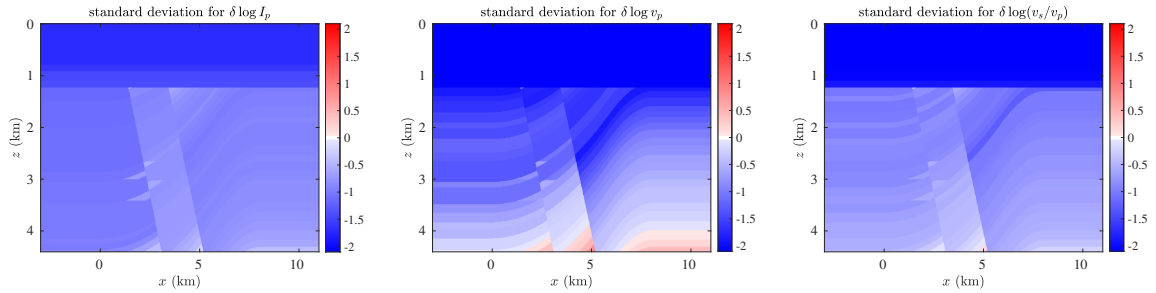
(a) (b) (c)

Figure 2 *The index map (a) defines the piecewise constant values per layer for the model parameters in Figure 1. The negative indices correspond to 4 reservoirs. A coarser version (b) is obtained by pairwise combinations of layers, excluding sea water, top layer, and reservoirs. Ellipses (c) for the conditional distribution of 2 parameters in two adjacent layers.*



(a) (b) (c)

Figure 3 *Standard deviations on a logarithmic scale for the marginal distributions for components (a) $\delta \log I_p$, (b) $\delta \log v_p$, and (c) $\delta \log(v_s/v_p)$. Some values are clipped at the extrema of the colour scale.*



(a) (b) (c)

Figure 4 *Standard deviations on a logarithmic scale for the conditional distributions for components (a) $\delta \log I_p$, (b) $\delta \log v_p$, and (c) $\delta \log(v_s/v_p)$. Some values are clipped at the extrema of the colour scale.*

Example

Figure 1 displays a 2-D piecewise constant isotropic elastic model, with density ρ in 1(a), v_p in 1(b), and v_s in 1(c), all governed by the index map in 2(a). The sea water has index 0 and four reservoirs are numbered by negative indices. For simplicity, a geological unit is taken to be identical to the region defined by one index in the index map, in this case leading to constant material parameters per unit. This is not necessary for the method, which allows for variations of the model parameters inside a unit. Only the relative perturbations should be constant per unit. The model was extended laterally by piecewise constant extrapolation, perpendicular to the boundary. A marine acquisition has 199 source position x_s between -2900 and 7000 m at a 50-m interval and 10-m depth with, using a 15-Hz Ricker wavelet. A receiver line has offsets $x_r - x_s$ between 100 and 6000 m with a 25-m interval at 8-m depth. The recording time is 5 s. The associated hessian has 3×93 rows and columns, for 93 geological units with 3 parameters each, and with one zero row and column for the water layer where $v_s = 0$. The ellipsoid given by $J_{\text{unc}}(\mathbf{m}) = J_{\text{noise}}$ describes the uncertainty, where $J_{\text{unc}} = \frac{1}{2}\mathbf{m}^{\mathsf{T}}\mathbf{H}\mathbf{m}$ and $J_{\text{noise}} = \frac{1}{2}\|\mathbf{d}_{\text{noise}}\|^2$ is the noise energy. The singular value decomposition $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathsf{T}}$, $\mathbf{S} = \text{diag}(\mathbf{s})$, provides $J_{\text{unc}} = \frac{1}{2}\mathbf{y}_{\text{H}}^{\mathsf{T}}\mathbf{y}_{\text{H}}$ with $\mathbf{y}_{\text{H}} = \mathbf{S}^{1/2}U^{\mathsf{T}}\mathbf{m}$ and $\mathbf{m} = \mathbf{U}(\mathbf{S}^{\dagger})^{1/2}\mathbf{y}_{\text{H}}$. The ellipsoid follows from $\mathbf{y}_{\text{H}}$ on a high-dimensional sphere with radius $(2J_{\text{noise}})^{1/2}$.

Figure 2(c) shows the conditional distribution for all parameters fixed except the P-impedances for units with index 41 and 44 below it, corresponding to the central part of the model in between the two faults and the reservoirs with index -1 and -2. The white ellipse is the uncertainty for a noise energy taken as 0.1% of the data energy. The image depicts $(J_{\text{unc}}/J_{\text{noise}})^{1/2}$ as a function of the two model parameters, assuming all others to be zero. Similarly, the covariance matrix $\mathbf{C} = \mathbf{U}\mathbf{S}^{\dagger}\mathbf{U}^{\mathsf{T}}$ determines ellipses defined by constant values of $\mathbf{m}^{\mathsf{T}}\mathbf{C}\mathbf{m} = \mathbf{y}_{\text{C}}^{\mathsf{T}}\mathbf{y}_{\text{C}}$, with $\mathbf{y}_{\text{C}} = (\mathbf{S}^{\dagger})^{1/2}\mathbf{U}^{\mathsf{T}}\mathbf{m}$ on a hypersphere and $\mathbf{m} = \mathbf{U}\mathbf{S}^{1/2}\mathbf{y}_{\text{C}}$. The magenta ellipse in Figure 2(c) corresponds to a 2×2 subset of $\mathbf{C}$ and represents the marginal distribution of the two parameters. The line segments will be explained later on. Figure 3 shows the standard deviations with the same noise energy as above for the marginal distributions of each model parameter, where the other variables are integrated away, implying that no a priori conditions are imposed other than that they have a gaussian distribution. Only quantities in the darker blue part can be reconstructed with any confidence. On the one hand, these estimates are too pessimistic because no a priori information is used. On the other hand, the contribution of small-scale variations and ignored parameters describing, for instance, anisotropy and attenuation, will tend to increase the standard deviations. The piecewise constant assumption has a regularizing effect. The uncertainty in $\delta \log I_p$ appears to be larger than that in $\delta \log v_p$. This can be understood in terms of the layered structure: not only the amplitude-versus-offset (AVO) effect of the top of a layer plays a rôle, but also the travel time to the bottom and back.

Figure 4 depicts the standard deviations for the conditional distribution of each parameter, given all others. The uncertainties are smaller than those in Figure 3. The conditional distribution given all parameters except those in a single unit is given by a 3×3 block of the hessian and is useful for multi-parameter inversion. Figure 5 shows the heatmap of its inverse, the covariance matrix, for the assumed noise level. The diagonal corresponds to squared standard deviations, the off-diagonals describe the coupling to the other two parameters.

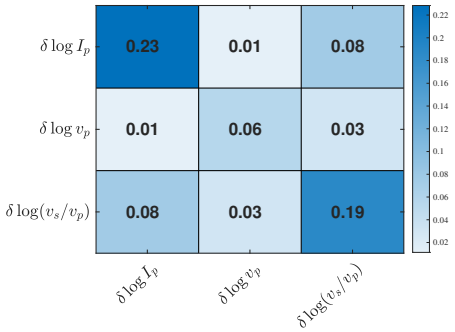
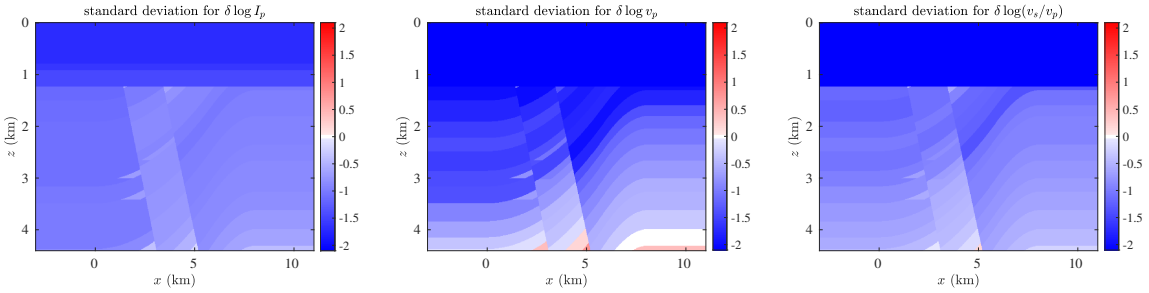


Figure 5 *Conditional covariance matrix for the reservoir with index* -1.

Fine and coarse

How does the length scale of the geological units affect the uncertainty estimates? To describe the relation between a fine and coarse representation of the model perturbations, we adopt the multigrid nomenclature (Hackbusch, 1985, e.g.). The prolongation operator $\mathbf{I}_{\text{f}}$ interpolates model parameters from the coarse to the fine grid and the restriction operator $\mathbf{I}_{\text{c}}$ projects them from fine to coarse. The term 'grid' is not entirely appropriate in our setting, but will be used anyway to describe a finer-scale representation of geological units, each with a constant model perturbation for each parameter, or a coarser representation obtained by combining a small number of them. The restriction and prolongation operators are related by $\mathbf{I}_{\text{c}} = \mathbf{D}_{\text{c}}\mathbf{I}_{\text{f}}^{\mathsf{T}}$, with $\mathbf{D}_{\text{c}}$ a diagonal matrix that scales $\mathbf{I}_{\text{f}}^{\mathsf{T}}$ such that the row sum of $\mathbf{I}_{\text{c}}$ is 1, and $\mathbf{I}_{\text{f}}^{\mathsf{T}}$ is the trans-



(a) (b) (c)

Figure 6 *As Figure 4 but for coarser units. The uncertainty is reduced.*

pose of the non-square matrix $\mathbf{I}_\mathrm{f}$. A symmetric transfer operator is $\mathbf{J}_\mathrm{c} = \mathbf{D}_\mathrm{c}^{1/2}\mathbf{I}_\mathrm{f}^\mathrm{T}$, with $\mathbf{J}_\mathrm{c}\mathbf{J}_\mathrm{c}^\mathrm{T} = \mathbf{I}_\mathrm{c}\mathbf{I}_\mathrm{f} = \mathbf{I}$. The reverse $\mathbf{I}_\mathrm{f}\mathbf{I}_\mathrm{c}$ is a filter operator that takes fine-grid values to the coarse grid by averaging them and then puts the result back on the fine grid. Its application to the fine model $\mathbf{m}_f$ produces $\tilde{\mathbf{m}}_f = \mathbf{I}_\mathrm{f}\mathbf{I}_\mathrm{c}\mathbf{m}_f$ and the associated hessian becomes $\tilde{\mathbf{H}}_\mathrm{f} = \mathbf{I}_\mathrm{c}^\mathrm{T}\mathbf{I}_\mathrm{f}^\mathrm{T}\mathbf{H}_\mathrm{f}\mathbf{I}_\mathrm{f}\mathbf{I}_\mathrm{c} = \mathbf{I}_\mathrm{c}^\mathrm{T}\mathbf{H}_\mathrm{c}\mathbf{I}_\mathrm{c}$, where the hessian $\mathbf{H}_\mathrm{c}$ corresponds to the coarse model $\mathbf{m}_c = \mathbf{I}_\mathrm{c}\mathbf{m}_f$. This shows that the fine-grid hessian can be obtained from the coarser one if the fine grid model is filtered by removing high spatial frequencies corresponding to fine-grid details. For a fractal Earth, this statement is clearly false, but with band-limited data, the approach is useful up to the resolved scale. Another way of looking at restriction and prolongation is in terms of a rotation operator. Instead of a non-square matrix $\mathbf{J}_\mathrm{c}$, a square block diagonal matrix can be constructed. Each block j has the size of the number $n_{\mathrm{f},j}$ of fine-scale units contained inside the coarse one. Starting from an identity matrix, the first row of each block can be replaced by a subset of row j of $\mathbf{J}_\mathrm{c}$, containing ones divided by $\sqrt{n_{\mathrm{f},j}}$. The remaining rows are orthonormalized relative to this row. These rows will lead to differences of fine-scale model parameters, reminiscent of a Haar (1910) wavelet basis. Such a rotation preserves the uncertainty, but restricted to one direction.

In the example of Figure 2(c), unit 41 and 44 below it can be taken together, leaving the direction along the summed parameters. The white line segment then describes the range of uncertainty for the conditional distribution. It covers a part of the magenta segment for the marginal distribution. The scaling by $1/\sqrt{2}$ puts the endpoints on the ellipse. For the restriction of the hessian, however, this scaling should not be applied, increasing the length of the line segment. This reflects the fact that the uncertainty decreases by $\sqrt{n_{\mathrm{f},j}}$ if larger units are considered.

Conclusions

The computation of large-scale uncertainties in the context of full-waveform inversion is feasible if the recovered subsurface model is segmented into a limited number of geological units. Uncertainties at smaller scales can be roughly estimated by assuming that they are inversely proportional to the square root of the number of parameters under consideration.

Acknowledgements

This study benefited from discussions with Wei Dai, Fabian Ernst, Sijmen Gerritsen, Vanessa Goh, Mark Huiskes, Michael Kiehn, John Kimbro, Gautam Kumar, Colin Perkins, René-Édouard Plessix, Richard Shipp and Yi Yang.

References

Deal, M.M. and Nolet, G. [1996] Null-space shuttles. *Geophysical Journal International*, **124**(2), 372–380.

Haar, A. [1910] Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, **69**(3), 331–371.

Hackbusch, W. [1985] *Multi-Grid Methods and Applications*. Springer, Berlin, Heidelberg.

Pratt, R.G., Shin, C. and Hicks, G.J. [1998] Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, **133**(2), 341–362.

Tarantola, A. [2005] *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia.