

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/335102113>

# Bayesian uncertainty estimation for full waveform inversion: A numerical study

Conference Paper · August 2019

DOI: 10.1190/segam2019-3216008.1

CITATION

1

READS

158

3 authors:



**Muhammad Izzatullah**

King Abdullah University of Science and Technology

10 PUBLICATIONS 6 CITATIONS

SEE PROFILE



**Tristan van Leeuwen**

Utrecht University

112 PUBLICATIONS 1,159 CITATIONS

SEE PROFILE



**Daniel Peter**

King Abdullah University of Science and Technology

102 PUBLICATIONS 1,092 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Understanding UQ for waveform inversion [View project](#)



SRVM-based Uncertainty Quantification in FWI [View project](#)

# Bayesian uncertainty estimation for full waveform inversion: A numerical study

Muhammad Izzatullah\*, Tristan van Leeuwen†, and Daniel Peter\*

\* King Abdullah University of Science and Technology, Saudi Arabia † Utrecht University, The Netherlands

## SUMMARY

Full waveform inversion enables us to obtain high-resolution subsurface images. However, estimating the associated uncertainties is not trivial. Hessian-based method gives us an opportunity to assess the uncertainties around a given estimate based on the inverse of the Hessian, evaluated at that estimate. In this work we study various algorithms for extracting information from this inverse Hessian based on a low-rank approximation. In particular, we compare the Lanczos method to the randomized singular value decomposition. We demonstrate that the low-rank approximation may lead to a biased conclusion.

## INTRODUCTION

Full-waveform inversion (FWI) was proposed to estimate physical parameters of the subsurface by fully exploiting information embedded in seismic waveforms and it enables us to obtain high resolution subsurface images (Pratt et al., 1998; Virieux and Operto, 2009; Virieux et al., 2017). However, quantifying the uncertainties in the reconstructed model is still a challenge. The studies and discussions related to uncertainty quantification and resolution analysis (Fichtner and Trampert, 2011; Martin et al., 2012; Trampert et al., 2013; Fichtner and Leeuwen, 2015) are relatively few compared to the discussions on efficiency and accuracy in searching the best-fitting models. However, without proper uncertainty analysis, the inverted subsurface models are difficult to interpret and assess.

Recently, uncertainty quantification has become an active of research area within exploration geophysics especially to estimate model uncertainties associated with FWI within the Bayesian inference framework (Tarantola, 2004). Based on Tarantola (2004), in the vicinity of the global minimum of an optimization problem, the inverse Hessian operator could be taken as the posterior covariance matrix. This leads to the recent technique which has been implemented to estimate model uncertainties based on a low-rank approximation of the Hessian (Martin et al., 2012; Bui-Thanh et al., 2013; Chen and Ghattas, 2018; Zhu et al., 2016).

In this abstract, we numerically study the performance of the low-rank approximation of the posterior covariance operator. In particular, we compare the randomized singular value decomposition (SVD) (Halko et al., 2011; Halko, 2012) to the Lanczos algorithm (Parlett, 1998; Golub and van Loan, 2013). We study the effect of the truncation rank on the estimated uncertainty.

The outline of this work is as follows. First, we discuss the theoretical foundation of uncertainty estimation within the Bayesian inference framework and its relationship to the non-linear least squares problem i.e., FWI. Secondly, we briefly discuss the

randomized SVD and Lanczos algorithm used to construct the low-rank approximation of the posterior covariance operator. Next, we present numerical examples in frequency domain FWI based on these two algorithms. Finally, we discuss the results and present conclusions.

## THEORY

In FWI, we assume that the observations,  $\mathbf{d}$ , are given by

$$\mathbf{d} = F(\mathbf{m}) + \varepsilon, \quad (1)$$

where  $\mathbf{m}$  denotes the medium parameters,  $F$  is the known forward operator and  $\varepsilon \sim \mathcal{N}(0, \mathbf{C}_D)$  is additive, normally distributed noise with zero mean and covariance  $\mathbf{C}_D$ .

With the prior assumption that  $\mathbf{m} \sim \mathcal{N}(0, (\alpha \mathbf{L}^T \mathbf{L})^{-1})$  we can define the posterior density as

$$\rho_{post}(\mathbf{m}|\mathbf{d}) \propto \exp[-V(\mathbf{m})], \quad (2)$$

where  $V(\mathbf{m})$  is the misfit function defined as

$$V(\mathbf{m}) = \frac{1}{2} \|F(\mathbf{m}) - \mathbf{d}\|_2^2 + \frac{\alpha}{2} \|\mathbf{Lm}\|_2^2. \quad (3)$$

The misfit function defined above resembles the Tikhonov regularization least-squares approach where the first and second terms on the right hand side are equivalent to our likelihood and prior PDFs, both of which are Gaussian. The posterior is not Gaussian, however, due to the non-linearity of  $F$ . A Gaussian approximation of the posterior may be constructed however, by linearizing around the maximum a posteriori model estimate,  $\mathbf{m}_{MAP}$  (Martin et al., 2012; Bui-Thanh et al., 2013), and the approximated posterior PDF can be defined as

$$\rho_{post}(\mathbf{m}|\mathbf{d}) \propto \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_{MAP})^T \mathbf{C}_{post}^{-1}(\mathbf{m} - \mathbf{m}_{MAP})\right]. \quad (4)$$

The maximum a posteriori model,  $\mathbf{m}_{MAP}$ , can be determined by solving a non-linear least-squares problem. Here,  $\mathbf{C}_{post}$  is the posterior covariance operator, evaluated at the maximum a posteriori model  $\mathbf{m}_{MAP}$  which can be described as

$$\mathbf{C}_{post} = (\mathbf{H}(\mathbf{m}_{MAP}) + \alpha \mathbf{L}^T \mathbf{L})^{-1} \quad (5)$$

where  $\mathbf{H}(\mathbf{m}) = \mathbf{J}(\mathbf{m})^T \mathbf{J}(\mathbf{m})$  with  $\mathbf{J}$  the Jacobian of  $F$  is the Gauss-Newton approximation of the Hessian (Pratt et al., 1998) evaluated at  $\mathbf{m}_{MAP}$ .

## Low-rank approximation of the Hessian

Here, we will review the low-rank approximation of the posterior covariance proposed by (Martin et al., 2012; Bui-Thanh

## Bayesian uncertainty estimation for FWI

et al., 2013). The Gauss-Newton Hessian  $\mathbf{H}$  is a symmetric positive semi-definite matrix and admits an eigen-value decomposition,

$$\mathbf{H} = \mathbf{U}_r \Lambda_r \mathbf{U}_r^T \quad (6)$$

where  $\mathbf{U}$  contains the eigenvectors and  $\Lambda$  is a diagonal matrix containing the eigenvalues. In the formulation above, the subscript  $r$  is refer to the first  $r$  most dominant eigenvectors, this giving a rank  $r$  approximation of the Gauss-Newton Hessian  $\mathbf{H}$ .

The corresponding low-rank approximation of the posterior covariance matrix is given by Bui-Thanh et al. (2013)

$$\mathbf{C}_{post} = \mathbf{L}^{-1} \left( \alpha^{-1} \mathbf{I} - \alpha^{-2} \mathbf{L}^{-1} \mathbf{U}_r \Sigma_r \mathbf{U}_r^T \mathbf{L}^{-T} \right) \mathbf{L}^{-T} + \mathcal{O} \left( \sum_{i=r+1}^n \frac{\lambda_i}{\lambda_i + 1} \right). \quad (7)$$

where  $\Sigma_r := \text{diag}(\lambda_1/(\lambda_1 + 1), \dots, \lambda_r/(\lambda_r + 1))$ . Equation (7) also shows the truncation error due to the low rank approximation based on the first  $r$  eigenvalues.

### The randomized SVD and Lanczos algorithms

Randomized SVD is used to reduce the computational cost when computing the several most dominant eigenpairs of the Hessian  $\mathbf{H}$  since it is considered more efficient compared to the classical SVD algorithm in terms of computational time (Halko et al., 2011). Meanwhile, the Lanczos algorithm (Golub and van Loan, 2013; Parlett, 1998) is a deterministic algorithm that is an adaptation of power method to find the  $r$  most dominant eigenvalues and eigenvectors for symmetric matrices.

The advantage of the randomized SVD and Lanczos algorithms is they only require matrix-vector products ("matvecs"), and therefore there is no need to explicitly form and store the dense Hessian in the memory.

In terms of the computational cost, as mentioned in Halko et al. (2011), for matrix-vectors multiplications, the randomized SVD algorithm has the same computational cost as Lanczos algorithm which is proportional to  $rT_{mult} + (m+n)r^2$ , where  $T_{mult}$  denotes the cost of a matrix-vector multiplication with the input matrix for approximating  $r$  leading singular vectors of the input matrix. Moreover, the advantage of randomized algorithm here compared to Lanczos algorithm is in the way of data is being accessed, where the  $r$  matrix-vector multiplies can be executed in parallel.

A point to highlight here that these algorithms are not computing the exact singular values as in SVD but an approximation. Moreover, an additional approximation error is incurred when using the dominant eigenvectors of the Hessian to approximate the posterior covariance, which is based on the *inverse* of the Hessian. For further discussions on the accuracy and computational cost of these algorithms, reader may refer to (Golub and van Loan, 2013; Parlett, 1998; Halko et al., 2011).

## NUMERICAL EXAMPLE - MARMOUSI MODEL

In this numerical example, we use the Marmousi model, depicted in the top in Figure 1 to test the proposed methodology. A standard finite-difference method is used to solve the Helmholtz equation (van Leeuwen, 2019). The grid size is  $61 \times 220$ , and grid spacing is  $50 \times 50m$ . 50 shots at every 100m and 100 receivers at every 50m are used in this numerical example. We use frequency content from 0.5 Hz to 3.5 Hz with frequency sampling of 0.5 Hz.

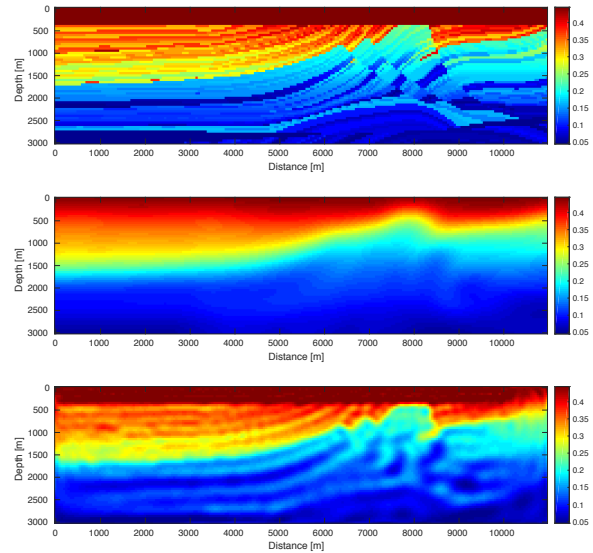


Figure 1: *Top*: True slowness-squared model. *Middle*: Initial slowness-squared model. *Bottom*: Reconstructed slowness-squared model ( $\mathbf{m}_{MAP}$ ) with 100 nonlinear CG iterations.

We performed 100 nonlinear conjugate gradient iterations, starting from the initial model depicted in the middle in Figure 1 to obtain the maximum a posteriori model  $\mathbf{m}_{MAP}$  shown in the bottom of Figure 1. As regularization, we use the Tikhonov regularization method with regularization operator  $\mathbf{L}$  as first order derivative operator and regularization parameter  $\alpha$  equals to 0.1.

In practice, the Hessian is not store explicitly in memory and only its matrix-vectors product are being computed. For analysis purposes, in this numerical example we explicitly compute the Hessian for comparison but all the computations through randomized SVD and Lanczos algorithm were performed using matvecs. We compute the first 150 eigenvalues and eigenvectors of  $\mathbf{H}$ .

In Figure 2, we show the first 150 eigenvalues of  $\mathbf{H}$  computed by these two algorithms. The eigenvalues computed by Lanczos algorithm are very close to the exact eigenvalues computed from the explicit Hessian using SVD. This result is aligned with one reported by Chen and Saad (2009). However, the eigenvalues computed by the randomized SVD are a good approximation for the first 10 eigenvalues while the rest are underestimated. The percentage error between exact and approximated eigenvalues is computed by  $\frac{\|\lambda_{real} - \lambda_{approx}\|}{\|\lambda_{real}\|} \times 100$ . The

## Bayesian uncertainty estimation for FWI

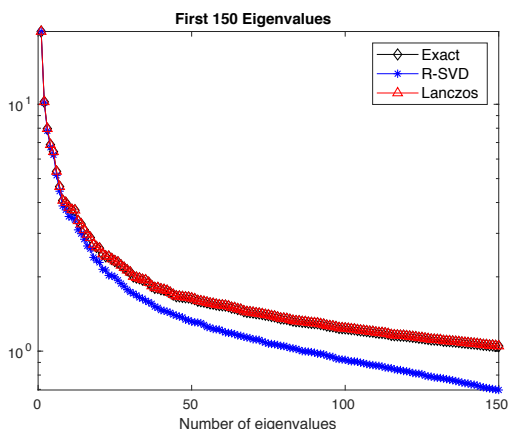


Figure 2: Comparison of first 150 eigenvalues obtained from randomized SVD and Lanczos Algorithm with the exact eigenvalues computed using SVD of explicit Hessian.

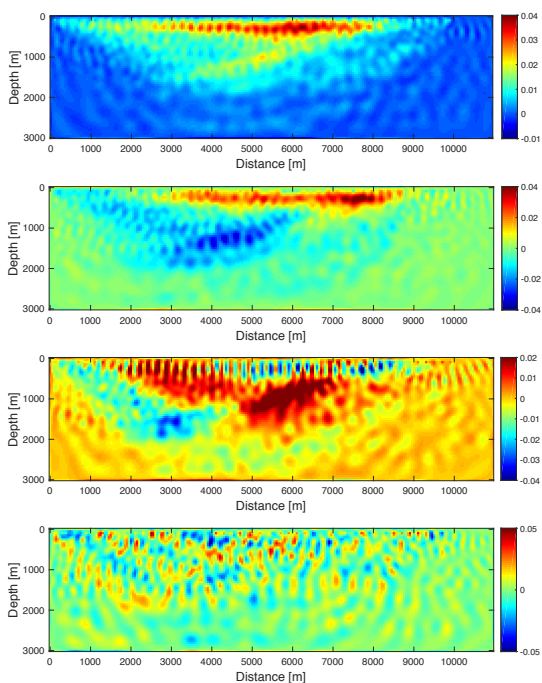


Figure 3: Comparison of (a) 1st, (b) 2nd, (c) 5th, and (d) 25th singular vectors obtained from randomized SVD.

percentage error for randomized SVD and Lanczos algorithm are 8.702% and  $7.4571 \times 10^{-13}\%$ , respectively. This variations in eigenvalues could leads to different results in sampling the posterior distribution.

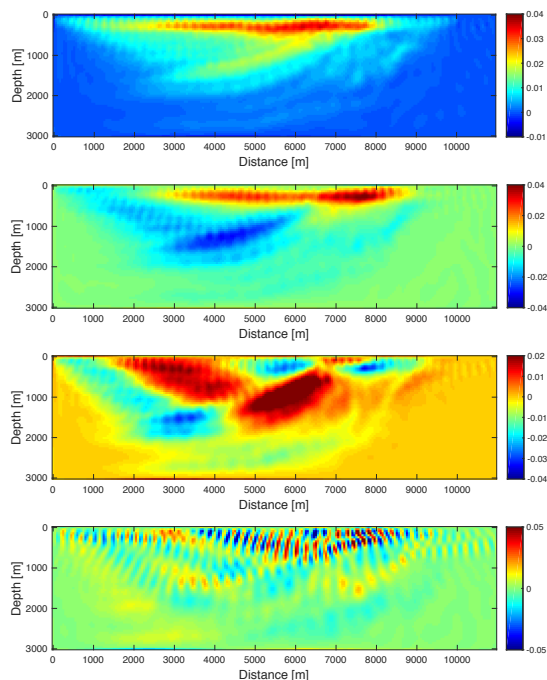


Figure 4: Comparison of (a) 1st, (b) 2nd, (c) 5th, and (d) 25th eigenvectors obtained from Lanczos algorithm.

We illustrate several dominant eigenvectors of the Hessian computed by these two algorithms in Figures 3 and 4. We observe as the order of the eigenvector increases, the energy distribution in the eigenvectors moves gradually from the top to the bottom of the subsurface model. This gives us information about certainty where the region with high energy distribution contains the most information. For example, the first eigenvector in Figures 3 and 4 shows that the shallow region has the most energy distribution and we are certain that this region contains the most information because it is close to the acquisition setup.

For numerical comparison of eigenvectors, the eigenvectors computed by Lanczos algorithm is much "cleaner" compared to randomized SVD where noise is obviously visible in the randomized SVD singular vectors. Furthermore, in this comparison we sample the random matrix  $\Omega$  with maximum number of over-samples which equals to 10 as suggested in Halko et al. (2011), however, the presence of noise in the singular vectors is pretty obvious.

In Figure 5 and Figure 6, we demonstrate the standard deviations computed from diagonal elements of the low-rank approximation of the posterior covariance matrix constructed by using first 50, 100, and 150 eigenvectors and eigenvalues ob-

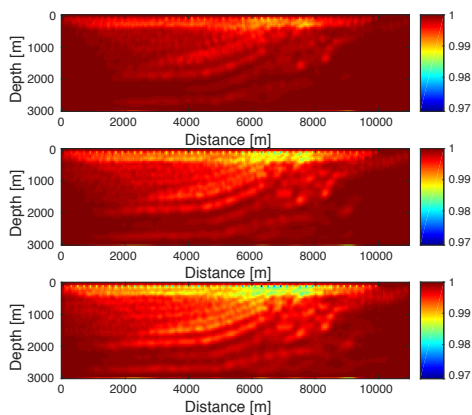


Figure 5: Comparison of standard deviations obtained from diagonal elements of posterior covariance matrix constructed with (a)50, (b) 100, and (c) 150 singular vectors of randomized SVD.

tained from R-SVD and Lanczos algorithms. Also, we illustrated the explicitly computed posterior covariance matrix for comparison purposes in Figure 7. As we observe, both algorithms gave different range and intensity of standard deviations. This is related back to the accuracy of the algorithms in approximating the eigenpairs of the Hessian. Also, the range of standard deviation is changing for different number of  $r$  eigenvectors and eigenvalues chosen. Furthermore, compared to the true standard deviation of posterior covariance matrix, the standard deviations obtained from low-rank approximation of posterior covariance matrix are underestimated. In this situations we could say that our standard deviations are biased.

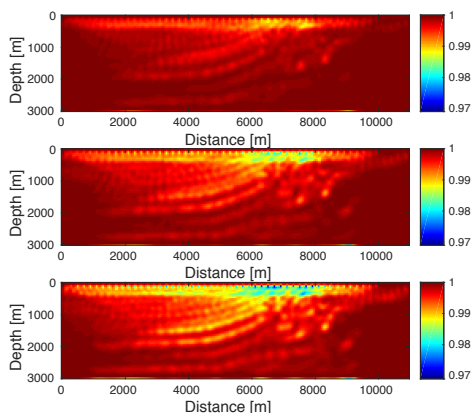


Figure 6: Comparison of standard deviations obtained from diagonal elements of posterior covariance matrix constructed with (a)50, (b) 100, and (c) 150 eigenvectors of Lanczos algorithm.

CONCLUSION

In conclusion, based on the numerical studies conducted, the Lanczos algorithm performs better with high accuracy than the randomized SVD in approximating the eigenpairs of the Hessian. The randomized SVD still has a computational advantage due to the ease of parallelization of the algorithm. A further study is needed to assess the trade-off between accuracy and computational efficiency.

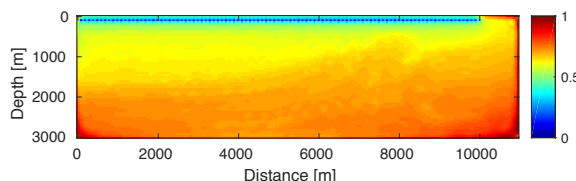


Figure 7: True standard deviation of posterior covariance matrix computed explicitly from equation 5.

In estimating the uncertainty, the important issue need to be taken seriously here is that by taking low-rank approximation of the Hessian in analyzing the uncertainties, we need to pay the price that our conclusion is biased. This is a serious issue because for different numbers of  $r$  taken in the low-rank approximation, the variance and the standard deviation would be different as well and this leads to a biased conclusion. The cause for this bias is the relationship between the inverse Hessian and the posterior covariance matrix. As commonly understood, the most dominant singular vectors of Hessian are the least dominant singular vectors of posterior covariance matrix. Further analysis of this approach will be included in our future works.

ACKNOWLEDGMENTS

This research work was performed under supervision of Tristan van Leeuwen during the first author’s research visit to Mathematical Institute of Utrecht University, Netherlands from 20th January to 5th February 2019. The first author would like to thank Tristan van Leeuwen for his guidance throughout the visit. The research visit and the work reported here was supported by funding from King Abdullah University of Science and Technology (KAUST). We would like to thank Nick Luiken, Ajinkya Kadu, and Sarah Gaaf from Mathematical Institute of Utrecht University, Netherlands and members of the Seismic Modeling and Inversion (SMI) group at KAUST especially Qiancheng Liu, Armando Espindola and Diaz Urozayev for constructive discussions.

## REFERENCES

- Bui-Thanh, T., O. Ghattas, J. Martin, and G. Stadler, 2013, A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion: *SIAM Journal on Scientific Computing*, **35**, A2494–A2523, doi: <https://doi.org/10.1137/12089586X>.
- Chen, J., and Y. Saad, 2009, Lanczos vectors versus singular vectors for effective dimension reduction: *IEEE Transactions on Knowledge and Data Engineering*, **21**, 1091–1103, doi: <https://doi.org/10.1109/TKDE.2008.228>.
- Chen, P., and O. Ghattas, 2018, Hessian-based sampling for high-dimensional model reduction: 1–19, arXiv:1809.10255.
- Fichtner, A., and T. V. Leeuwen, 2015, Resolution analysis by random probing: *Journal of Geophysical Research: Solid Earth*, **120**, 5549–5573, doi: <https://doi.org/10.1002/2015JB012106>.
- Fichtner, A., and J. Trampert, 2011, Resolution analysis in full waveform inversion: *Geophysical Journal International*, **187**, 1604–1624, doi: <https://doi.org/10.1111/j.1365-246X.2011.05218.x>.
- Gerhard Pratt, R., C. Shin, G. J. Hick, 1998, Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi: <https://doi.org/10.1046/j.1365-246X.1998.00498.x>.
- Golub, G. H., and C. F. van Loan, 2013, *Matrix computations*, 4th ed.: JHU Press.
- Halko, N., P. Martinsson, and J. Tropp, 2011, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions: *SIAM Review*, **53**, 217–288, doi: <https://doi.org/10.1137/090771806>.
- Halko, N. P., 2012, Randomized methods for computing low-rank approximations of matrices: Thesis, [https://doi.org/https://amath.colorado.edu/faculty/martinss/Pubs/2012\\_halko\\_dissertation.pdf](https://doi.org/https://amath.colorado.edu/faculty/martinss/Pubs/2012_halko_dissertation.pdf).
- 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**, A1460–A1487, doi: <https://doi.org/10.1137/110845598>.
- Parlett, B. N., 1998, *The symmetric eigenvalue problem*: Prentice-Hall, Inc.
- Tarantola, A., 2004, *Inverse problem theory and methods for model parameter estimation*: Society for Industrial and Applied Mathematics.
- Trampert, J., A. Fichtner, and J. Ritsema, 2013, Resolution tests revisited: The power of random numbers: *Geophysical Journal International*, **192**, 676–680, doi: <https://doi.org/10.1093/gji/ggs057>.
- van Leeuwen, T., 2019, <https://doi.org/https://github.com/tristanvanleeuwen/simplefwi>.
- Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, 6. An introduction to full waveform inversion, in V. Grechka and K. Wapenaar, eds., *Encyclopedia of exploration geophysics*: Society of Exploration Geophysicists, R1-1–R1-40.
- Virieux, J., and S. Operto, 2009, An overview of FWI in exploration geophysics: *Geophysics*, **74**, no. 6, WCC127–WCC152, doi: <https://doi.org/10.1190/1.3238367>.
- 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, doi: <https://doi.org/10.1190/geo2015-0641.1>.