Geophysical Journal International

Geophys. J. Int. (2019) **218,** 1100–1120
Advance Access publication 2019 May 22
GJI Seismology

doi: 10.1093/gji/ggz137

Square-root variable metric based elastic full-waveform inversion—Part 2: uncertainty estimation

Qiancheng Liu and Daniel Peter

King Abdullah University of Science and Technology (KAUST), Division of Physical Sciences and Engineering, Thuwal, Saudi Arabia.
E-mail: daniel.peter@kaust.edu.sa

Received 2019 February 20; in original form 2018 October 4

SUMMARY

In our first paper (Part 1) about the square-root variable metric (SRVM) method we presented the basic theory and validation of the inverse algorithm applicable to large-scale seismic data inversions. In this second paper (Part 2) about the SRVM method, the objective is to estimate the resolution and uncertainty of the inverted resulting geophysical model. Bayesian inference allows estimating the posterior model distribution from its prior distribution and likelihood function. These distributions, when being linear and Gaussian, can be mathematically characterized by their covariance matrices. However, it is prohibitive to explicitly construct and store the covariance in large-scale practical problems. In Part 1, we applied the SRVM method to elastic full-waveform inversion in a matrix-free vector version. This new algorithm allows accessing the posterior covariance by reconstructing the inverse Hessian with memory-affordable vector series. The focus of this paper is on extracting quantitative and statistical information from the inverse Hessian for quality assessment of the inverted seismic model by FWI. To operate on the inverse Hessian more efficiently, we compute its eigenvalues and eigenvectors with randomized singular value decomposition. Furthermore, we collect point-spread functions from the Hessian in an efficient way. The posterior standard deviation quantitatively measures the uncertainties of the posterior model. 2-D Gaussian random samplers will help to visually compare both the prior and posterior distributions. We highlight our method on several numerical examples and demonstrate an uncertainty estimation analysis applicable to large-scale inversions.

Key words: Waveform inversion; Computational seismology; Seismic tomography.

1 INTRODUCTION

Full-waveform inversion (FWI) addresses the geophysical inverse problem of estimating subsurface model parameters from observed waveform data. In most geophysical applications, FWI is introduced as an iterative, local optimization problem that attempts to minimize the least-squares residuals between observed and synthetic data. Mathematically, the inverse problem is ill-posed, leading to a non-uniqueness of the solutions. It remains challenging to solve inverse problems pratically due to limitations in data acquisition, measurement uncertainties and the non-uniqueness of the solution. Several methods in geophysics have been proposed to attack this FWI challenge. We can divide them mainly into two categories: deterministic methods such as gradient optimization-based ones (see, e.g. Lailly 1983; Tarantola 1984, 1987; Pratt 1999; Virieux & Operto 2009), and probabilistic approaches such as Bayesian inversions (see, e.g. Gouveia & Scales 1998; Käufl *et al.* 2013; Biswas & Sen 2017).

Deterministic methods on FWI have been well developed in tackling several difficult issues, such as convergence rate, cycle-skipping and multiparameter trade-offs (Bamberger *et al.* 1979; Tarantola 1987; Crase *et al.* 1990; Bunks *et al.* 1995; Igel *et al.* 1996; Tromp *et al.* 2005; Virieux & Operto 2009; Warner & Guasch 2016). Probabilistic inversions of FWI are in principle preferred solutions to an inverse problem as they additionally provide posterior distributions for assessing statistical informations about the solutions, but often become prohibitive due to computational costs. From the viewpoint of Bayesian inference, the solution of the inverse problem should not be limited to a single set of inverted values, but be represented by a probability density function (PDF) to quantify the posterior model uncertainty (Tarantola 2005).

Estimations of the resolution or uncertainty in seismic inversions have a long history in geophysics and can be analysed with mathematical tools such as the posterior covariance matrix (Tarantola & Valette 1982). The posterior covariance matrix is closely related to the Hessian matrix (Fichtner & Trampert 2011a,b; Fichtner & van Leeuwen 2015; Zhu *et al.* 2016). However, for practical problems with millions of parameters it is unfeasible to store such vast matrices. To handle large-scale inverse problems, Zhang & McMechan (1995) modify classic inversion algorithms with least-squares QR factorization. An (2012) evaluates the spatial resolution lengths

with a Gaussian approximation to the resolution matrix. Trampert *et al.* (2012) sample the tomographic models for resolution lengths with random probing. Fichtner & van Leeuwen (2015) analyse the direction-dependent resolution lengths of waveform tomography by autocorrelating the randomly sampled Hessian. We refer to Rawlinson *et al.* (2014) to provide a recent review about uncertainty estimations.

With the development of matrix probing theories in applied mathematics, randomized singular-value decomposition (SVD) attracted attention also to geophysicists (Halko *et al.* 2011; Liberty *et al.* 2007). Bui-Thanh *et al.* (2013) formulate waveform tomography in a Bayesian inference workflow, deriving an approximation to the posterior covariance matrix by decomposing the data-misfit Hessian into eigenvalues and eigenvectors with randomized SVD. Zhu *et al.* (2016) improve the efficiency of the Hessian computation by exploiting point-spread function (PSF) tests. Similar research can also be found in Mosegaard & Tarantola (1995), Gouveia & Scales (1998), Sambridge & Mosegaard (2002) and Osypov *et al.* (2013). Besides, more recent ensemble-based approaches using Kalman Filter (KF) theory (Kalman *et al.* 1960; Evensen 1994) have been applied to tomography problems by Jordan (2015) and Thurin *et al.* (2017).

In this second paper about the square-root variable metric (SRVM) approach in FWI, we are focusing on uncertainty estimation based on the Hessian-related approximation. As mentioned in the first part of our research (Liu *et al.* 2018), we approximate the inverse Hessian more directly by using the SRVM method (Morf & Kailath 1975; Williamson 1975; Hull & Tapley 1977). Theoretically, the SRVM method is capable of collecting the information of the second-order derivative from the initial model to the final, inverted model. We write the SRVM algorithm into a vector version to make it memory-affordable for large-scale problems. We then implement the vector-version SRVM into elastic FWI (Liu *et al.* 2018). The focus of this paper is to retrieve the second-order derivative information from the stored SRVM scalar and vector series in a recursive manner, after convergence of the inversion. The vector-version SRVM algorithm approaches the inverse Hessian in a low-rank form. To make this method computationally more flexible and efficient, we furthermore incorporate the randomized SVD to the SRVM approach.

In the following, we first review the theory of Bayesian inference for FWI. Next, we discuss how to reconstruct the inverse Hessian with the vector-version SRVM algorithm and relate the reconstructed matrix to the covariance matrix. Subsequently, we introduce randomized SVD into SRVM. Also, we discuss the possibility of a novel approach to get the PSFs of the Hessian and attempt to draw random samples from the posterior covariance. Finally, we verify our methods above on numerical examples to demonstrate the information gain of uncertainty analysis for seismic inversions.

2 INVERSE THEORY

For seismic applications, different concepts can be used to explain the inversion of seismic data to determine subsurface structures. In the theory of probabilistic inversions, the solution of an inverse problem is defined as a marginal probability distribution in the model parameter space (Tarantola 2005). This definition is very general and elegant. However, to complete the characterization of such a probability distribution, we are usually required to evaluate a large number of samples over the parameter space. Therefore, the feasibility of probabilistic theory to practical problems strongly depends on the size of the model parameter space.

2.1 Review of the Bayesian inference

Bayesian inference is a method in which the prior probability is updated as ever more information becomes available (Tarantola 2005). When applied to FWI workflows, Bayesian inference allows us to incorporate the prior information into waveform tomography to estimate posterior uncertainties of the inverted results. To review, we start from a short glance at the forward problem

$$\mathbf{d} = \mathbf{g}(\mathbf{m}), \tag{1}$$

which predicts the values of the observables $\mathbf{d}$ for a given model $\mathbf{m}$ via the (usually nonlinear) operator $\mathbf{g}$. The forward problem calculates what should be observed for a particular model, whereas the inverse problem calculates what the particular model should be for a set of observations. Eq. (1) denotes the forward-modelling process which relates the model parameters and data observables. Herein, it represents the numerical solution of the elastic wave equation based on a spectral-element method (Seriani & Priolo 1994; Faccioli *et al.* 1996; Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999).

Assuming the model prior is Gaussian distributed (Gauss 1809), we express the prior probability distribution function (PDF) as

$$\rho(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_{prior})\right)^T \mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_{prior})), \tag{2}$$

where $\mathbf{m}_{prior}$ is the prior model with its mean being the (initial) model $\mathbf{m}_0$, and $\mathbf{C}_m$ is the prior model covariance matrix. $\rho(\mathbf{m})$ can be inferred from geophysical surveys, for example, well logs. The pragmatic choice of Gaussian priors is very difficult to avoid because of the central limit theorem. However, note that such a Gaussian prior assumption can be invalid for some seismological applications, for example, when involving parameter discontinuities.

For eq. (1), even when the targeted parameter model $\mathbf{m}$ happens to coincide with the true parameter model, the synthetic data may still differ from observations due to data noise and measurement errors. Assuming the inherent discrepancy between the observed and synthetic data to be distributed as Gaussian noise, we represent the likelihood function of the data as

$$\rho(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{g}(\mathbf{m})\right)^T \mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{g}(\mathbf{m}))), \tag{3}$$

where $\mathbf{C}_d$ is the data covariance matrix indicating the data uncertainties. According to Bayesian inference, the solution to an inverse problem yields the posterior PDF, which is defined as

$$\rho(\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{d}|\mathbf{m})\rho(\mathbf{m}), \tag{4}$$

in which $\rho(\mathbf{m}|\mathbf{d})$ indicates the posterior probability, $\rho(\mathbf{d}|\mathbf{m})$ the likelihood and $\rho(\mathbf{m})$ the prior probability. Here, $\rho(\mathbf{m})$ is related to the model misfit and $\rho(\mathbf{d}|\mathbf{m})$ is related to the data misfit. Combining eqs (2) and (3) into eq. (4), we have

$$\rho(\mathbf{m}|\mathbf{d}) \propto \exp(-f(\mathbf{m})), \tag{5}$$

where $f(\mathbf{m})$ is the misfit function in the least-square sense, namely

$$f(\mathbf{m}) = \frac{1}{2}((\mathbf{d}-\mathbf{g}(\mathbf{m}))^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{g}(\mathbf{m})) + (\mathbf{m}-\mathbf{m}_{prior})^T\mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_{prior})). \tag{6}$$

Note that even with the Gaussian assumption for the PDFs of data error and model prior, the posterior PDF might be non-Gaussian

due to the non-linearity of the forward modeling operator $\mathbf{g}(\mathbf{m})$) in the likelihood function of eq. (3). One simplistic approach is to linearize the forward modeling operator around the maximum a posteriori (MAP) model (Gouveia & Scales 1998; Bui-Thanh *et al.* 2013), which gives the posterior PDF as

$$\rho(\mathbf{m}|\mathbf{d}) \propto \exp\left(-\frac{1}{2}(\mathbf{m}-\widetilde{\mathbf{m}})\right)^T \mathbf{C}_M^{-1}(\mathbf{m}-\widetilde{\mathbf{m}})), \tag{7}$$

with the MAP model being

$$\widetilde{\mathbf{m}} = \mathbf{m}_{prior} + (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1})^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{G}\mathbf{m}_{prior}), \tag{8}$$

and the posterior covariance being

$$\mathbf{C}_M = (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1})^{-1} = (\mathbf{H}_d + \mathbf{C}_m^{-1})^{-1}. \tag{9}$$

In eqs (8) and (9), $\mathbf{G} = \partial\mathbf{g}(\mathbf{m})/\partial\mathbf{m}$ is the Fréchet derivative. In eq. (9), we have $\mathbf{H}_d = \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G}$, which is known as the Gauss–Newton approximation to the data-misfit Hessian (Pratt 1999; Virieux & Operto 2009). The direct calculation of the Fréchet derivative, that is, the sensitivity kernel according to its definition is impossible in realistic models. An efficient solution to this problem is the adjoint method (Tarantola 1984, 1987; Tromp *et al.* 2005). This method allows us to obtain the first-order derivative by solving the forward problem once and its adjoint problem once. In this study with the absence of attenuation, the wave equation is self-adjoint, so there is in principle no need for automatic differentiation techniques (Sambridge *et al.* 2007; Tan *et al.* 2010; Vlasenko *et al.* 2016) to solve the adjoint problem, instead the same implementation as for the forward solution can be used in reversed time. From eq. (9), we realize that the key towards the uncertainty estimation is to efficiently quantify the second-order derivatives $\mathbf{H}_d$.

2.2 Constructing the inverse Hessian via SRVM vector series

In Newton's method, the model perturbation $\Delta\mathbf{m}$ is related to the gradient $\mathbf{g}$ via the Hessian $\mathbf{H}$ by

$$\Delta\mathbf{m} \approx -\mathbf{H}^{-1}\mathbf{g} = -\mathbf{B}\mathbf{g}, \tag{10}$$

where $\mathbf{B} \equiv \mathbf{H}^{-1}$. The standard Davidon–Fletcher–Powell (DFP) method (Fletcher & Powell 1963) gives an iterative approach to update the inverse Hessian, namely

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k\Delta\mathbf{g}_k\Delta\mathbf{g}_k^T\mathbf{B}_k}{\Delta\mathbf{g}_k^T\mathbf{B}_k\Delta\mathbf{g}_k} + \frac{\Delta\mathbf{m}_k\Delta\mathbf{m}_k^T}{\Delta\mathbf{g}_k^T\Delta\mathbf{m}_k}. \tag{11}$$

A stable and practical algorithmic form modified from the DFP method is the vector-version SRVM, whose detailed workflow can be found in Algorithm (2) in Liu *et al.* (2018). Considering eq. (11), we see that at convergence of the algorithm, the matrix $\mathbf{B}_{k+1}$ is a positive definite approximation to collect the information about the inverse Hessian over all the n iterations (Tarantola 2005) as follows:

$$\mathbf{B}_{n+1} = \mathbf{B}_0 + \sum_{k=0}^{n}\left(\frac{\Delta\mathbf{m}_k\Delta\mathbf{m}_k^T}{\Delta\mathbf{g}_k^T\Delta\mathbf{m}_k} - \frac{\mathbf{B}_k\Delta\mathbf{g}_k\Delta\mathbf{g}_k^T\mathbf{B}_k}{\Delta\mathbf{g}_k^T\mathbf{B}_k\Delta\mathbf{g}_k}\right), \tag{12}$$

where $\mathbf{B}_0$, the initial guess about the inverse Hessian, is usually chosen as the identity matrix $\mathbf{I}$ for convenience. Even when we have a good estimation of the diagonal Hessian, the estimated Hessian diagonals are imposed as point-wise weights over the model parameter space (Métivier *et al.* 2013), whereupon the identity matrix $\mathbf{I}$ remains. We can furthermore set $\mathbf{B}_0 = \mathbf{C}_m$, once given a prior covariance matrix $\mathbf{C}_m$ in a diagonal form. Also, $\mathbf{B}_0$ acts as a stabilizer to ensure a sufficiently stable $\mathbf{B}_k$ during its iterative update. Being a rank-2 updating approach, DFP (or SRVM) also has a self-correcting behaviour. For example, if $\mathbf{B}_k$ incorrectly estimates the curvature of the misfit function, $\mathbf{B}_k$ will tend to correct itself within a few successive steps (Nocedal & Wright 2006). The historical information about the gradients and model updates exists on the right-hand side (RHS) of eq. (12), so we take $\mathbf{B}_0 = \mathbf{I}$ as the stabilizer and retrieve the approximated inverse Hessian $\mathbf{H}^{-1}$ by

$$\mathbf{H}^{-1} = \mathbf{B}_{n+1} - \mathbf{I}. \tag{13}$$

The relation in eq. (13) resembles the Levenberg–Marquardt algorithm (damped least-squares inversion) (Nocedal & Wright 2006). After the SRVM-based elastic FWI converges, we can use the past scalar series $\frac{v_k}{P_k}$ and vector series $\mathbf{w}_k$ (see Algorithm (2) in Liu *et al.* 2018) to reconstruct $\mathbf{B}_{n+1}$ by:

Algorithm 1 $\mathbf{B}_{n+1}$ retrieval workflow

1: **for** $k \leftarrow 0$ to n **do**
2: $\quad\mathbf{S}_{k+1} = \mathbf{S}_k\left(\mathbf{I} - \frac{v_k}{P_k}\mathbf{w}_k\mathbf{w}_k^T\right)$
3: $\quad\mathbf{B}_{k+1} = \mathbf{S}_{k+1}\mathbf{S}_{k+1}^T = \mathbf{S}_k\left(\mathbf{I} - \frac{v_k}{P_k}\mathbf{w}_k\mathbf{w}_k^T\right)\left(\mathbf{I} - \frac{v_k}{P_k}\mathbf{w}_k\mathbf{w}_k^T\right)^T\mathbf{S}_k^T$
4: **end for**

The resulting $\mathbf{B}_{n+1}$ is symmetric positive definite, with its corresponding square-root matrix $\mathbf{S}_{n+1}$. In large-scale problems, the full reconstruction and storage of $\mathbf{B}_{n+1}$ in memory remain impractical. Fortunately, it is possible to extract the column (row) elements from $\mathbf{B}_{n+1}$ via a unit pulse probing vector (0,0,...,0,1,0,...,0,0), in which the unit pulse '1' locates at the target column (row) index i = 1, .., M, respectively. Here, M is the size of the model parameter space. Assuming $\hat{\mathbf{e}}_i$ is the probing vector, we probe $\mathbf{B}_{n+1}$ as following:

Algorithm 2 $\mathbf{B}_{n+1}$ probing

1: $\mathbf{b}_i = \hat{\mathbf{e}}_i$
2: **for** $k \leftarrow 0$ to n **do**
3: $\quad a = \frac{v_k}{P_k}\mathbf{b}_i\mathbf{w}_k^T$
4: $\quad\mathbf{b}_i = \mathbf{b}_i - a\mathbf{w}_k$
5: **end for**
6: **for** $k \leftarrow n$ to 0 **do**
7: $\quad a = \frac{v_k}{P_k}\mathbf{b}_i\mathbf{w}_k^T$
8: $\quad\mathbf{b}_i = \mathbf{b}_i - a\mathbf{w}_k$
9: **end for**

We fetch $\mathbf{B}_{n+1}\,\hat{\mathbf{e}}_i = \mathbf{b}_i$ when both loops in Algorithm (2) have finished. However, this probing method is not very efficient to extract arbitrary elements from $\mathbf{B}_{n+1}$. We can only extract one column (row) vector from $\mathbf{B}_{n+1}$ each time. To extract all diagonal elements, we need M such operations in Algorithm (2). Next, we will demonstrate how to make the probing of $\mathbf{B}_{n+1}$ more convenient.

2.3 Randomized SVD for the SRVM method

Algs (1) and (2) as well as eq. (13) present a low-rank approach to construct the inverse Hessian $\mathbf{H}^{-1}$. However, the probing method mentioned in the section above requires further improvement to be convenient for large-scale applications where the size M of the model parameter space is very large. Randomized

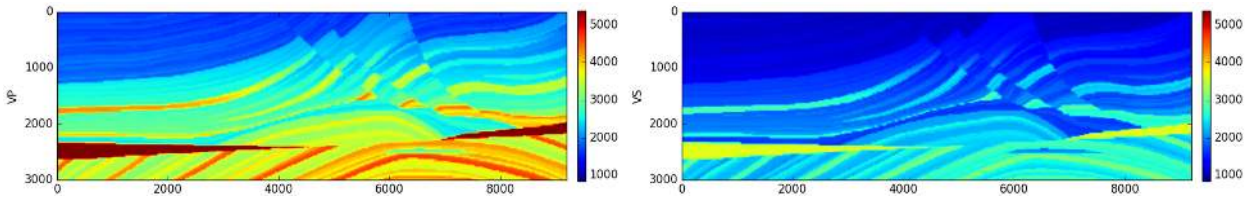

Figure 1. True V_P and V_S models used to generate observed data.

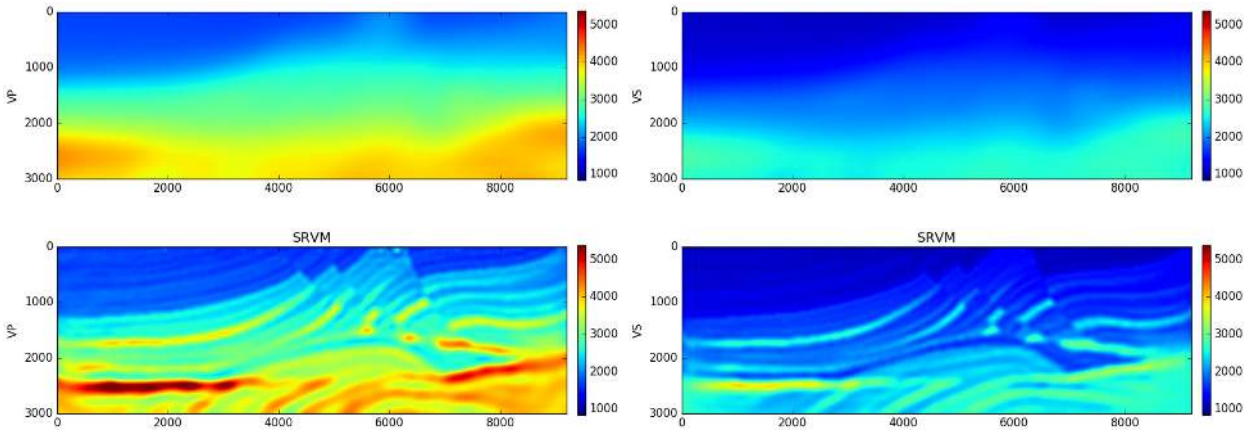

Figure 2. Initial and inverted V_P and V_S models. Top panels: prior V_P and V_S models obtained by smoothing the true models in Fig. 1. Bottom panels: inverted V_P and V_S models by SRVM-based elastic FWI after 46 iterations (details about data and model misfits can be found in Part 1 of our study).

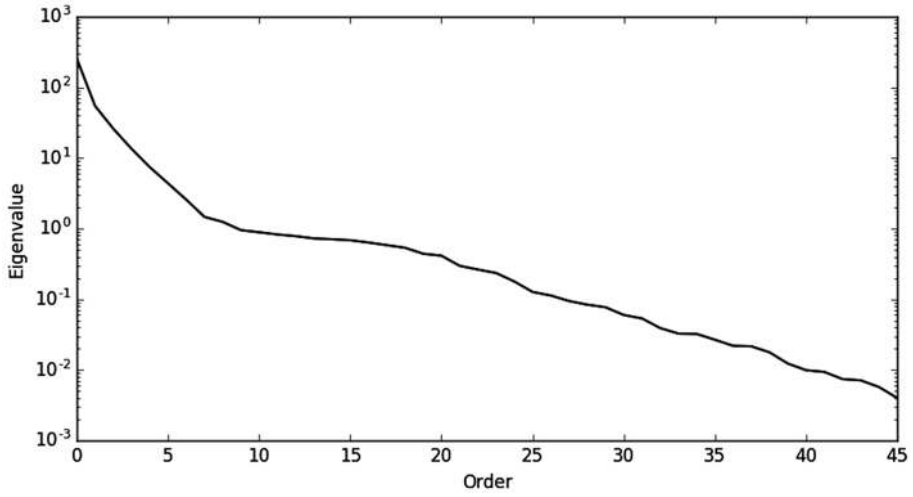

Figure 3. Eigenvalues, the $\mathbf{\Lambda}$ in eq. (14), extracted by randomized SVD algorithm from the stored 46 SRVM vectors.

SVD (Liberty *et al.* 2007; Halko *et al.* 2011) provides an efficient way to factorize large matrices into their corresponding eigenvalues and eigenvectors, especially for matrices with low-rank property. Compared with the classic randomized SVD (Liberty *et al.* 2007), the method proposed in Halko *et al.* (2011) probes a matrix only with one set of random vectors, which is much easier to implement.

Given a symmetric $M \times M$ matrix $\mathbf{Z}$ and a set of $M \times N_r$ random vectors $\mathbf{X}$, where M is the model size and N_r the estimated rank order of matrix $\mathbf{Z}$. As shown in Appendix A, the single-pass randomized SVD algorithm (Halko *et al.* 2011) processes the matrix $\mathbf{Z}$ as follows:

Algorithm 3 Single-pass randomised SVD algorithm

1: $\mathbf{Y} = \mathbf{ZX}$	▷ Sampling $\mathbf{Z}$ with $\mathbf{X}$
2: $\mathbf{QR} = \mathbf{Y}$	▷ QR decomposition on $\mathbf{Y}$
3: $\mathbf{A}(\mathbf{Q}^T\mathbf{X}) = \mathbf{Q}^T\mathbf{Y}$	▷ Solve for $\mathbf{A}$
4: $\mathbf{U\Lambda U}^T = \mathbf{A}$	▷ SVD on $\mathbf{A}$
5: $\mathbf{V} = \mathbf{QU}$	
6: $\mathbf{Z} = \mathbf{V\Lambda V}^T$	

For our purpose, we set $\mathbf{Z} = \mathbf{H}^{-1} = \mathbf{B}_{n+1} - \mathbf{I}$. With the reconstruction algorithm in Algorithm (1), it is still impossible to explicitly store $\mathbf{H}^{-1}$ for large-scale practical problems. Algorithm

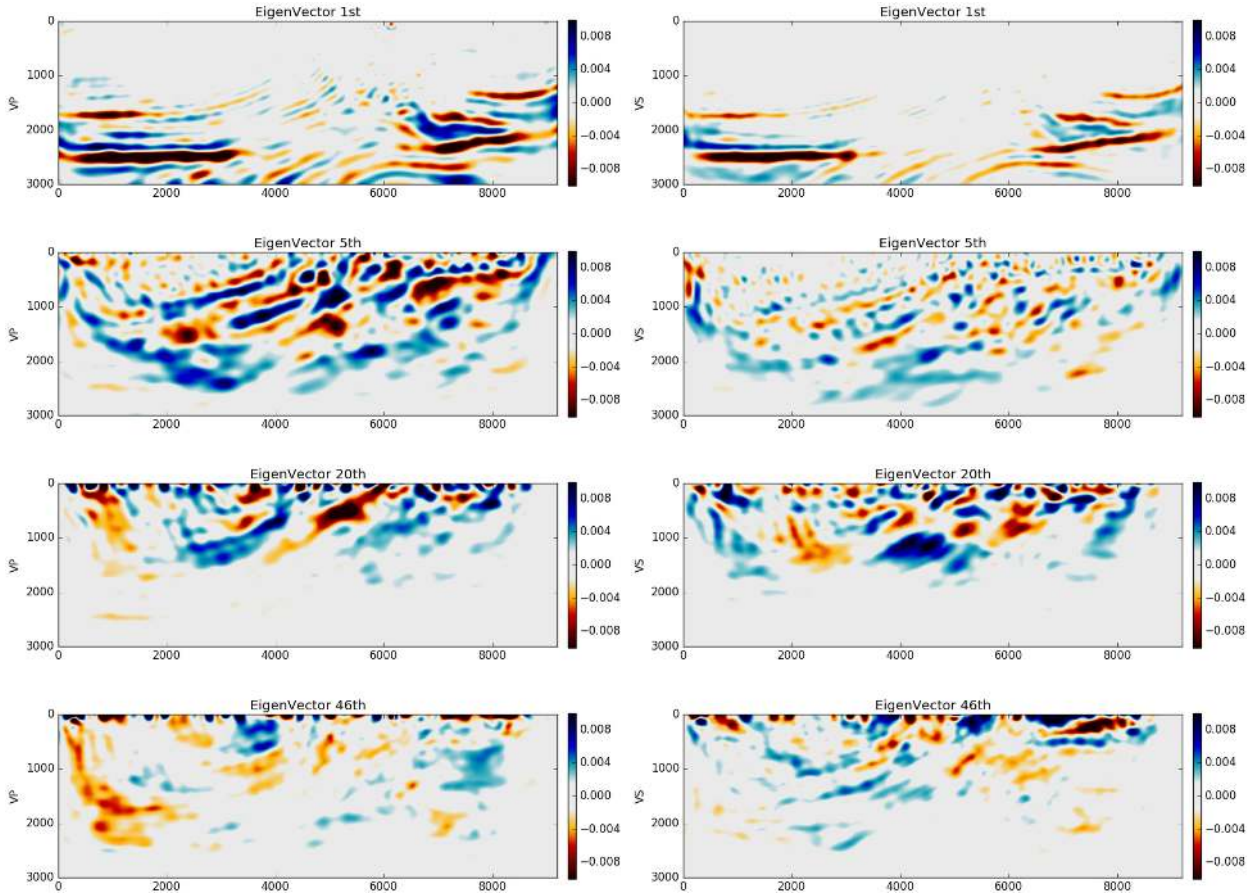

Figure 4. Eigenvectors, the $\mathbf{V}$ in eq. (14), factorized by randomized SVD algorithm from the stored 46 SRVM vectors. The left-hand and the right-hand columns are for V_P and V_S, respectively.

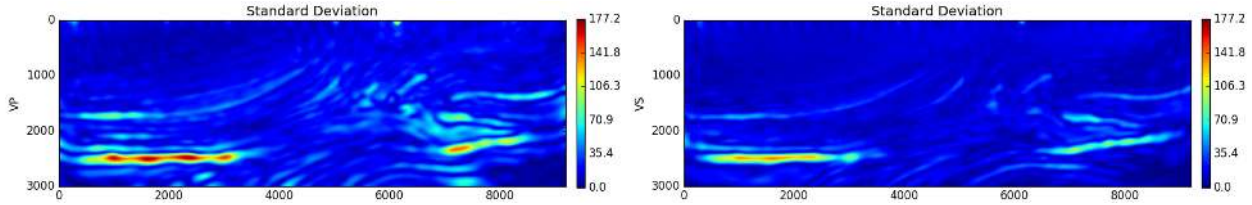

Figure 5. Standard deviation from the diagonals of the posterior covariance. Larger standard deviation values indicate higher uncertainty areas. The left and the right ones are for V_P and V_S, respectively.

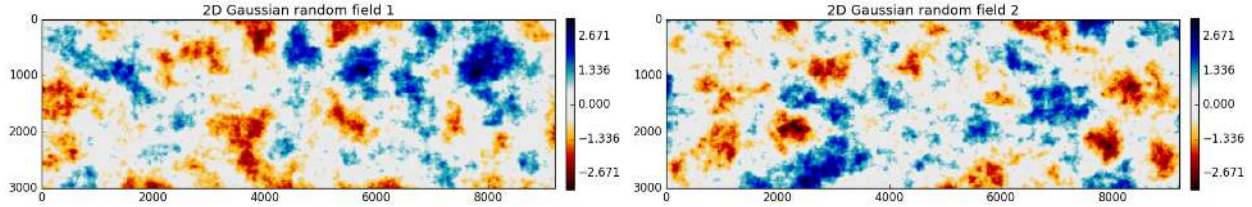

Figure 6. Examples of 2-D Gaussian random fields, with zero mean and unit variance. The left and the right ones are for V_P and V_S, respectively.

(2) works in probing $\mathbf{H}^{-1}$, but with poor efficiency. However, if we incorporate Algorithm (3) into Algorithm (2), then the probing of $\mathbf{H}^{-1}$ becomes efficient even for large M, as N_r is usually much smaller than M. We simply need to put the set of random vectors $\mathbf{X}$ in Algorithm (3) into Algorithm (2), and loop Algorithm (2) for N_r times to obtain $\mathbf{Y} = \mathbf{B}_{n+1}\mathbf{X} - \mathbf{X}$. Here, $\mathbf{X}$ comprises the N_r random vectors $\mathbf{x}_j$ ($j = 1, 2, ..., N_r$). $\mathbf{Y}$ is the same size as $\mathbf{X}$. From Algorithm (1), which can be taken as a low-rank form, we know that N_r equals the number of iterations. We finally have the inverse Hessian $\mathbf{H}^{-1}$ in a SVD form as

$$\mathbf{H}^{-1} = \mathbf{V}\Lambda\mathbf{V}^T, \tag{14}$$

in which $\mathbf{V}$ is a $M \times N_r$ eigenvector matrix, and $\mathbf{\Lambda}$ is the eigenvalue matrix only with N_r diagonal entries. Incorporating randomized SVD within the SRVM algorithm provides a convenient probing of

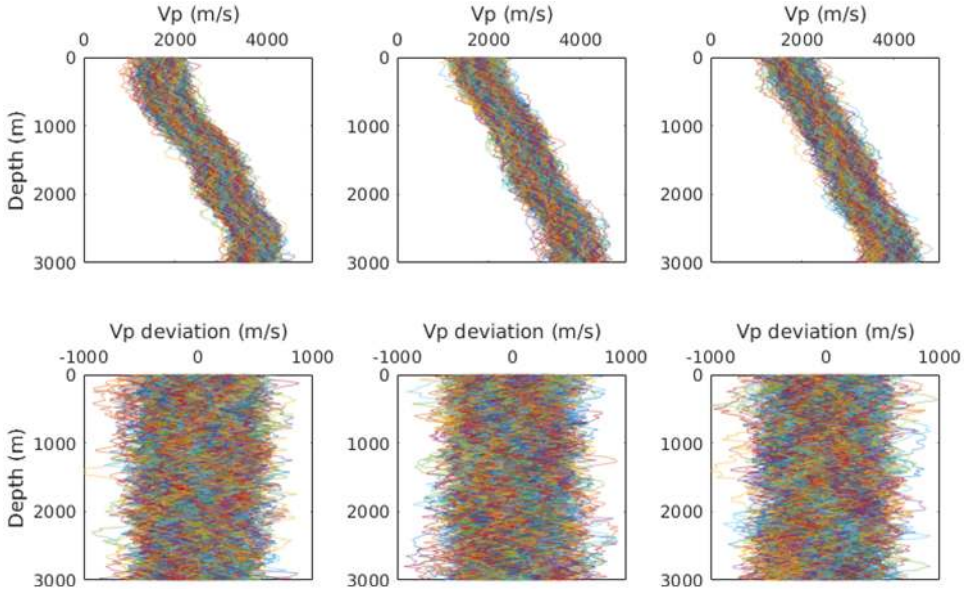

Figure 7. Prior distributions of V_P. Top row (from left- to right-hand) panel: prior distributions against depth of the V_P model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: prior distributions against depth of the V_P deviation at $X = 3100$, 4600, 6900 m.

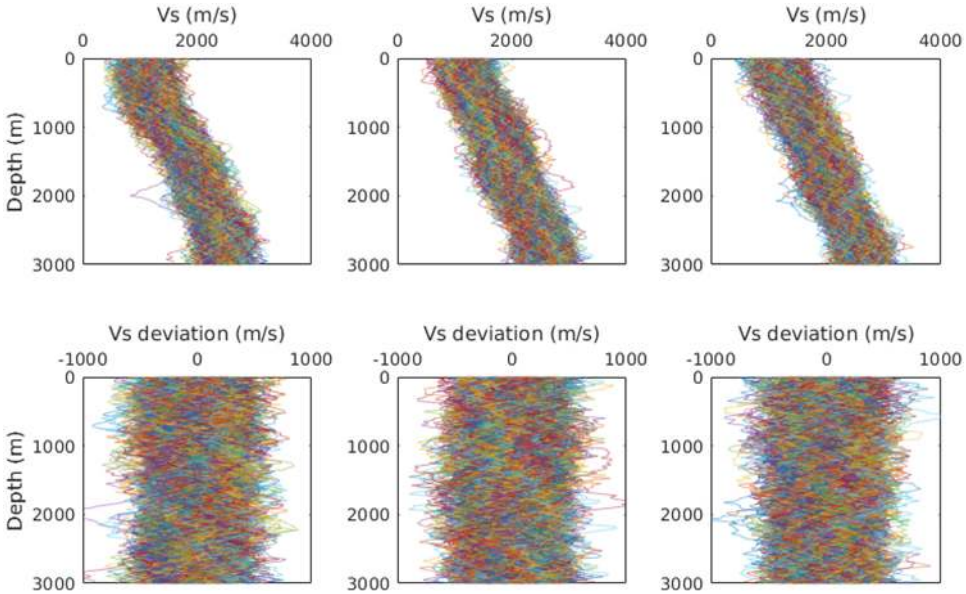

Figure 8. Prior distributions of V_S. Top row (from left- to right-hand) panel: prior distributions against depth of the V_S model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: prior distributions against depth of the V_S deviation at $X = 3100$, 4600, 6900 m.

the inverse Hessian $\mathbf{H}^{-1}$, both for column (row) elements as well as for diagonal elements.

2.4 SVD approach to point-spread function tests

Decomposing the spectrum by randomized SVD makes handling $\mathbf{H}^{-1}$ more convenient. Furthermore, after taking the inverse of $\mathbf{\Lambda}$ we can also access the Hessian:

$$\mathbf{H} = \mathbf{V}\Lambda^{-1}\mathbf{V}^T. \tag{15}$$

We recall that a single point-spread function (PSF) represents one column (row) of the Hessian. We can extract PSFs from eq. (15) for resolution analysis (Fichtner & Trampert 2011a,b; Bui-Thanh *et al.* 2013; Fichtner & van Leeuwen 2015). We can also extract diagonals from eq. (15) to show the source illumination, representing the data coverage of the acquisition geometry.

Inverse problems such as FWI suffer from being ill-conditioned, mostly due to a lack of data coverage. Therefore, the Hessian might be a singular matrix with a condition number tending towards infinity. Large condition numbers of the inverse Hessian lead to amplification of data noise being mapped to the model space. Thus, the truncated SVD here is equal to a norm damping regularization of the inverse problem. The truncation value for $\mathbf{\Lambda}$ could in principle being chosen by an L-curve assessment. For our purposes of probing the Hessian matrix, we prefer to sample $\mathbf{H}$ as exact as possible, thus choose a truncation value well below a perceivable norm damping effect.

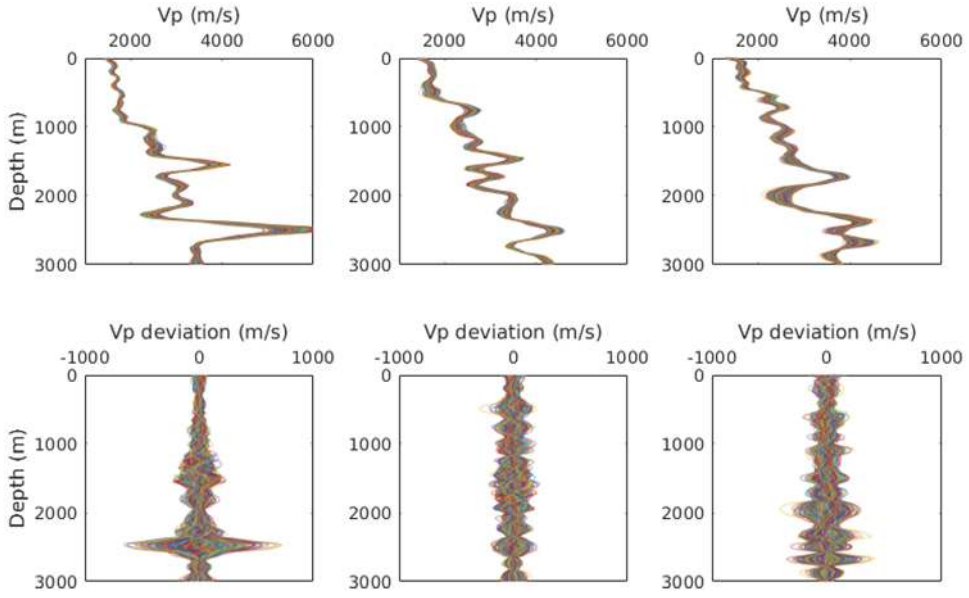

Figure 9. Posterior distributions of V_P. Top row (from left- to right-hand) panel: posterior distributions against depth of the V_P model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: posterior distributions against depth of the V_P deviation at $X = 3100$, 4600, 6900 m.

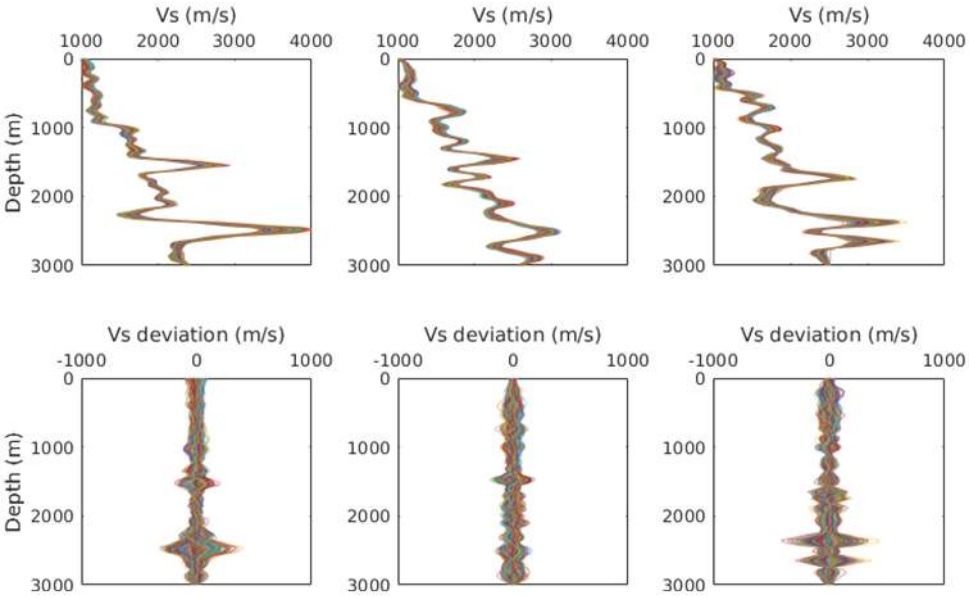

Figure 10. Posterior distributions of V_S. Top row (from left- to right-hand) panel: posterior distributions against depth of the V_S model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: posterior distributions against depth of the V_S deviation at $X = 3100$, 4600, 6900 m.

2.5 Practical assessment of the posterior covariance

Eq. (6) provides an elegant perspective on the objective function, connecting least-squares inversion with Bayesian inference. However, in waveform tomography, we do not use eq. (6) directly but rather use a generalized least-squares formulation in a more practical form (Rawlinson *et al.* 2014; Modrak & Tromp 2016; Modrak *et al.* 2018) as

$$f(\mathbf{m}) = \frac{1}{2}((\mathbf{d}-\mathbf{g}(\mathbf{m}))^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{g}(\mathbf{m})) + \varepsilon\,\delta\mathbf{m}^T\mathbf{C}_m^{-1}\delta\mathbf{m} + \eta\,\delta\mathbf{m}^T\mathbf{D}^T\mathbf{D}\delta\mathbf{m}), \quad (16)$$

where $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ is the model perturbation, $\mathbf{C}_d$ the data covariance matrix, and $\mathbf{C}_m$ the prior model covariance matrix, with $\mathbf{m}_0$ being the prior mean model (initial model) for a Gaussian distributed prior $\mathbf{m}_{prior}$. $\mathbf{D}$ is a second-order derivative smoothing operator (Trinh *et al.* 2017). ε and η are two constants supplied by the user to control the damping weight of the last two terms regarding the data misfit. When $\partial f(\mathbf{m})/\partial\mathbf{m} = 0$, the solution of eq. (16) occurs as

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \varepsilon\,\mathbf{C}_m^{-1} + \eta\,\delta\mathbf{D}^T\mathbf{D})^{-1} \times \mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{G}\mathbf{m}_0), \quad (17)$$

in which $\tilde{\mathbf{m}}$ is the inverted model (also known as the MAP model), and $\mathbf{G}$ the Fréchet derivative. $\mathbf{D}^T\mathbf{D}$ in eq. (17) works as the smoother, so we rewrite a practical form of the posterior covariance as

$$\mathbf{C}_M = (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \varepsilon\,\mathbf{C}_m^{-1})^{-1} = (\mathbf{H}_d + \varepsilon\,\mathbf{C}_m^{-1})^{-1}, \quad (18)$$

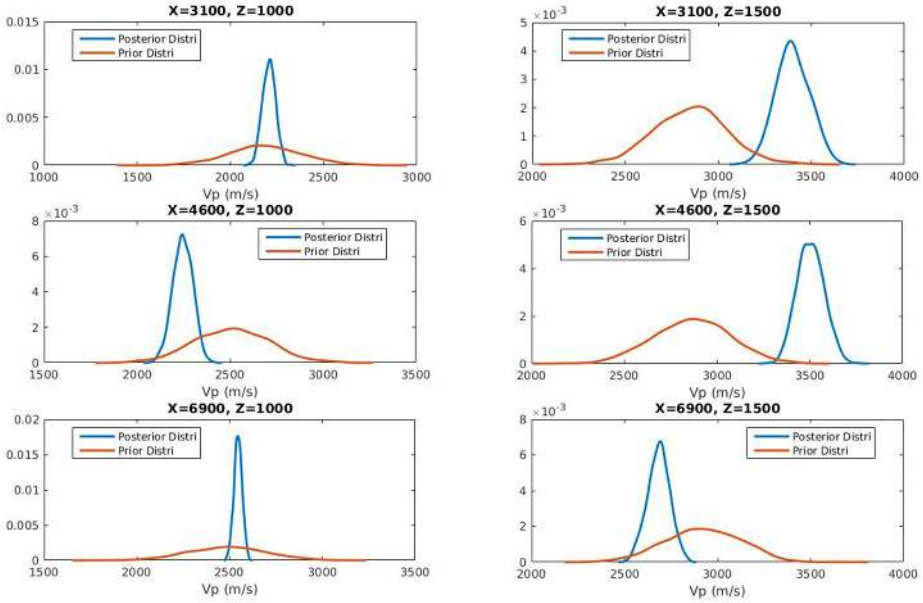

Figure 11. Samples for the comparisons of prior and posterior V_P distributions at different locations. Individual locations are titled on each subplot.

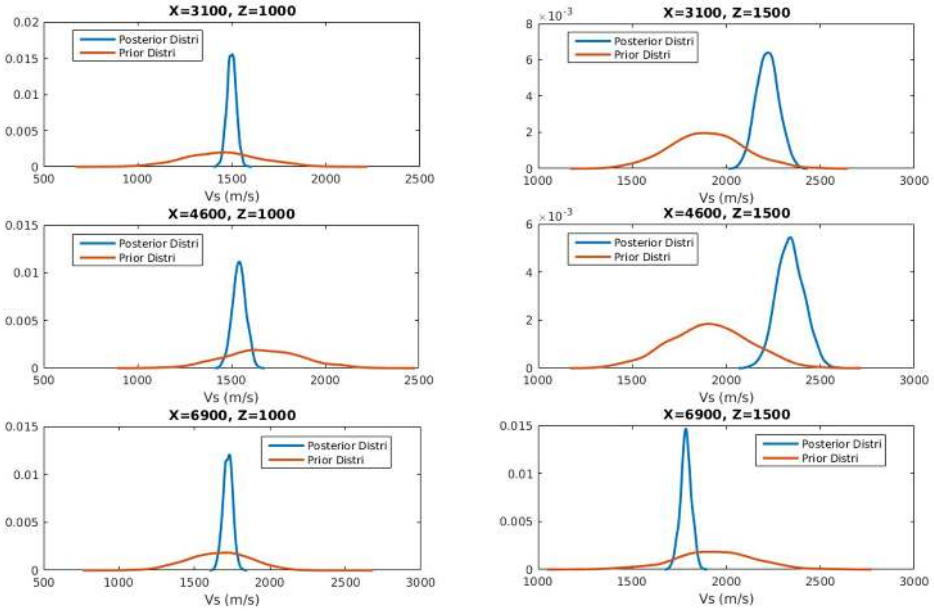

Figure 12. Samples for the comparisons of prior and posterior V_S distributions at different locations. Individual locations are titled on each subplot.

which differs from eq. (9) by an additional pre-factor ε. The approaches of Bui-Thanh *et al.* (2013) and Zhu *et al.* (2016) both begin by approximating the data-misfit Hessian and then evaluate the corresponding posterior covariance.

We notice that it is difficult to determine the relative magnitude between $\mathbf{H}_d$ and $\mathbf{C}_m^{-1}$ although ε acts as a tuning factor. When $\varepsilon \rightarrow \infty$, $\mathbf{C}_M \rightarrow 0$. The result is misleading. When the magnitude of $\varepsilon\,\mathbf{C}_m^{-1}$ is much higher than that of $\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G}$, $\mathbf{C}_M \rightarrow \mathbf{C}_m$. This is also misleading in that the model misfit dominates the inversion. When $\varepsilon \rightarrow 0$, we can take $\varepsilon\,\mathbf{C}_m^{-1}$ as a stabilizer such that $\mathbf{C}_M \rightarrow (\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G})^{-1}$. This last treatment is more reasonable and desirable than the previous ones (Rawlinson *et al.* 2014). Similarly, in the Bayesian transdimensional scheme by Bodin & Sambridge (2009), they mainly consider the data misfit function without the damping or smoothing terms. The idea that the Hessian is the inverse posterior covariance matrix in the vicinity of $\tilde{\mathbf{m}}$ can also be found in Fichtner & Trampert (2011b) and Fichtner & van Leeuwen (2015).

Therefore, we rewrite eq. (18) as

$$\mathbf{C}_M = (\mathbf{H}_d\,\mathbf{C}_m + \varepsilon\,\mathbf{I})^{-1}\mathbf{C}_m = (\mathbf{C}_m^{1/2}\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G}\mathbf{C}_m^{1/2} + \varepsilon\,\mathbf{I})^{-1}\mathbf{C}_m, \quad (19)$$

in which we take $\varepsilon\,\mathbf{I}$ as the stabilizer, $\mathbf{C}_m^{1/2}$ as spatial weights on the Fréchet derivative $\mathbf{G}$, and $\mathbf{H}^{-1} = (\mathbf{H}_d\,\mathbf{C}_m + \varepsilon\,\mathbf{I})^{-1}$ as a 'big filter' to connect $\mathbf{C}_M$ and $\mathbf{C}_m$. With $\mathbf{C}_m^{1/2}$ absorbed into $\mathbf{G}$, we approximately have

$$\mathbf{C}_M \approx \mathbf{H}^{-1}\mathbf{C}_m, \quad (20)$$

where $\mathbf{H}^{-1}$ has been approached in eq. (14). The prior covariance $\mathbf{C}_m$, which can be estimated from subsurface orientations or well logs (Fomel & Claerbout 2003), indicates the known information

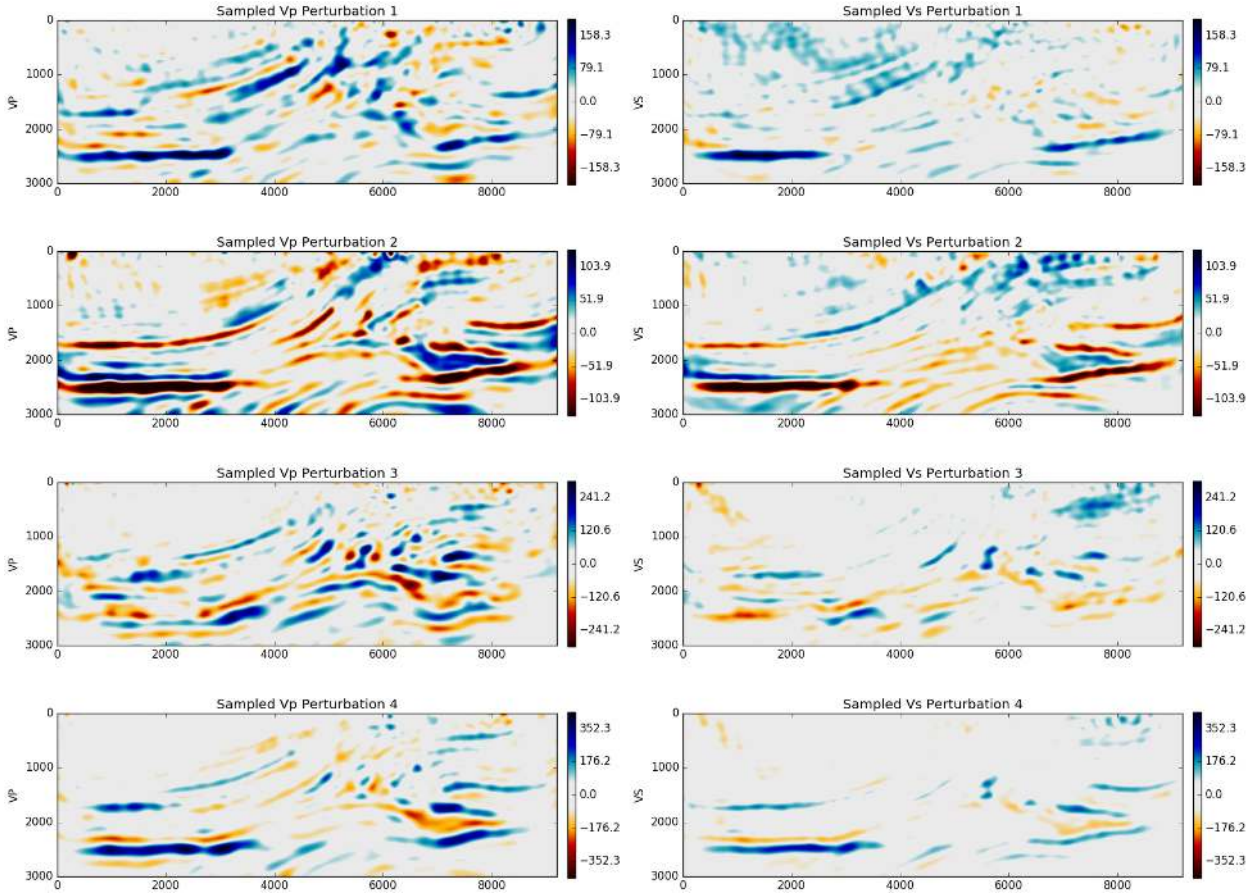

Figure 13. Different samples drawn from the posterior distributions about the V_P and V_S model perturbations.

before waveform tomography. Herein, we leave the precise estimation about prior covariance for our future research. We simply take the diagonal of $\mathbf{C}_m$ as constants in a vector form for the sake of simplified implementation. The square-root diagonal of $\mathbf{C}_M$, known as the standard deviation, provides a quantitative measure of the posterior distribution.

2.6 Random sampling on prior and posterior distributions

After obtaining a low-rank SVD approximation to the posterior covariance $\mathbf{C}_M$, we can draw and compare a Gaussian random sampling on $\mathbf{C}_m$ and $\mathbf{C}_M$. The sampling of the prior and posterior distributions can be expressed, respectively, as

$$\mathbf{m}_{\text{prior}} = \mathbf{m}_0 + \mathbf{C}_m^{1/2}\mathbf{n}, \tag{21}$$

$$\mathbf{m}_{\text{post}} = \widetilde{\mathbf{m}} + \mathbf{C}_M^{1/2}\mathbf{n}, \tag{22}$$

where $\mathbf{m}_0$ is the prior mean model (initial model) for a Gaussian distributed prior model $\mathbf{m}_{prior}$, $\mathbf{n}$ a 2-D Gaussian random sampler of zero mean and unit variance (Tarantola 2005) in the 2-D case, or a 3-D Gaussian random sampler in the 3-D case. Note that the independent 1-D normal random samplers are qualified enough for the sampling operations of randomized SVD in Algorithm (3), because the generation of 2-D/3-D Gaussian random samplers is computationally more costly. We compute the square root of the posterior covariance matrix in eq. (22) by

$$\mathbf{C}_M^{1/2} = \mathbf{H}^{-1/2}\mathbf{C}_m^{1/2} = \mathbf{V}\Lambda^{1/2}\mathbf{V}^T\mathbf{C}_m^{1/2}, \tag{23}$$

where $\mathbf{V}$ and $\mathbf{\Lambda}$ are the eigenvectors and eigenvalues of $\mathbf{H}^{-1}$, respectively. This way, we can assess the prior and posterior model uncertainties through visual comparisons of the random samplings on the prior and posterior distributions.

3 NUMERICAL EXAMPLES

The solution of an inverse problem depends on the chosen model parametrization. For elastic media, this could be *P*- and *S*-wave velocities and density, or bulk and shear modulus and density, or a parametrization involving bulk wave speed or impedances. In this paper, we focus on the inversion of the *P*- and *S*-wave velocities in elastic media, but keep the density fixed due to its low sensitivity in waveform tomography (Virieux *et al.* 2009). In the following, we use the 2-D elastic Marmousi model to demonstrate the applications of our methods. For our uncertainty analysis, the standard deviation will serve as a measure of uncertainty and the samplings of the prior and posterior distributions figure as visual inspections. Our demonstration comprises (i) SRVM plus randomized SVD, (ii) point-wise standard deviation of the posterior PDFs, (iii) a visual comparison between the prior and posterior distributions, (iv) 2-D randomly sampled V_P and V_S model perturbations from posterior PDFs and (v) Hessian diagonals and Hessian PSFs.

3.1 2-D elastic Marmousi model

In the 2-D elastic Marmousi (Martin *et al.* 2006) benchmark, the models are 9200 m long and 3000 m deep. The target and initial

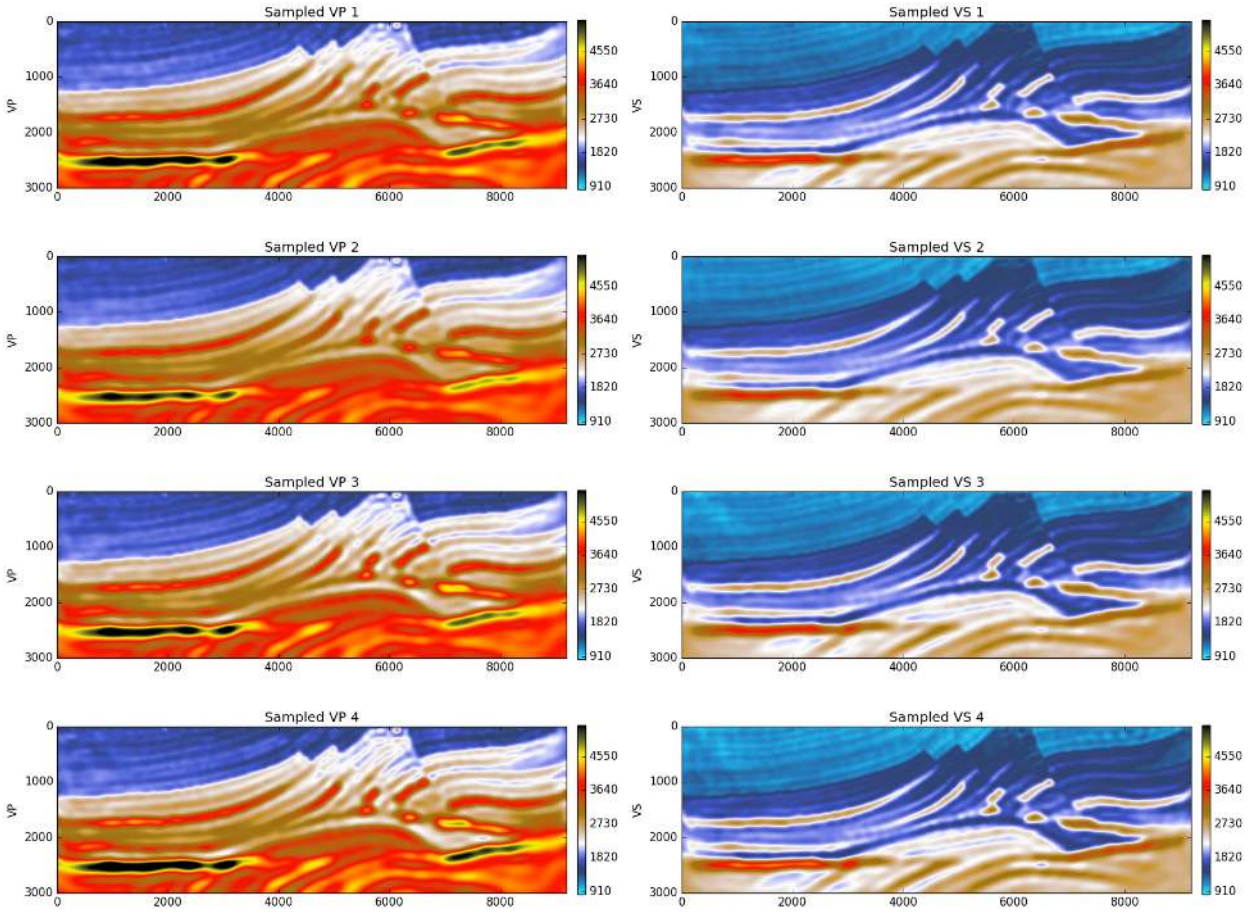

Figure 14. Different samples drawn from the posterior distributions about the V_P and V_S models (corresponding to the sampled perturbations in Fig. 13).

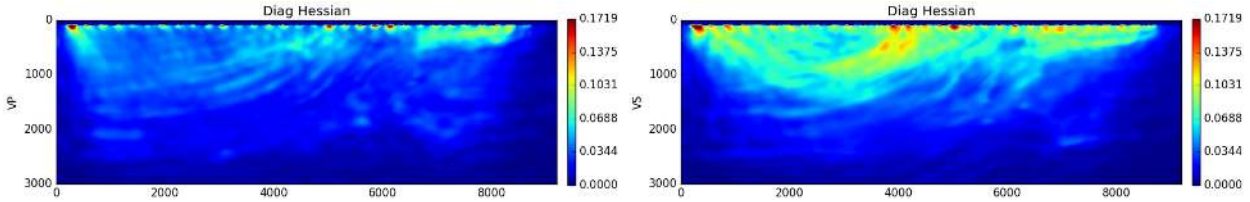

Figure 15. Square-root values of Hessian diagonals.

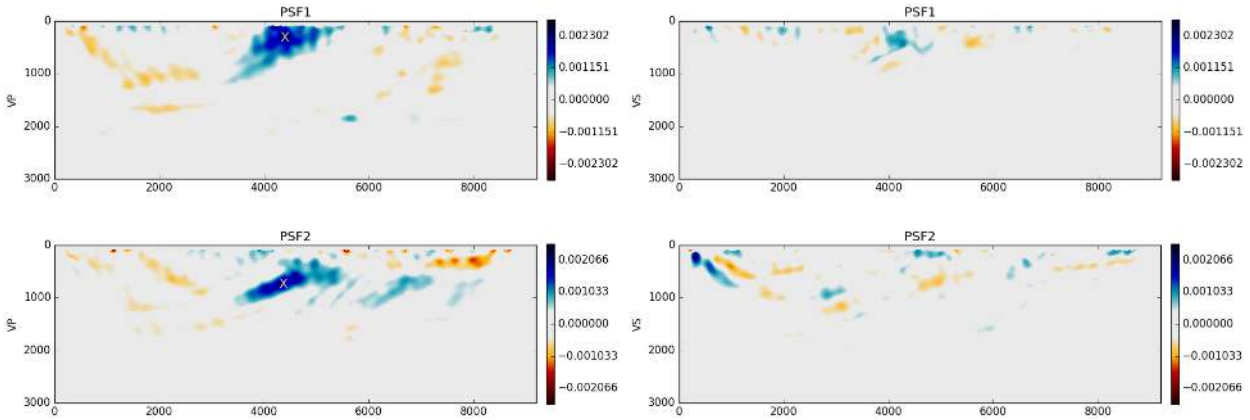

Figure 16. Two V_P PSFs from the estimated Hessian. Associated 'delta-like perturbations' are located in the V_P model, as indicated by the two 'crosses', respectively.

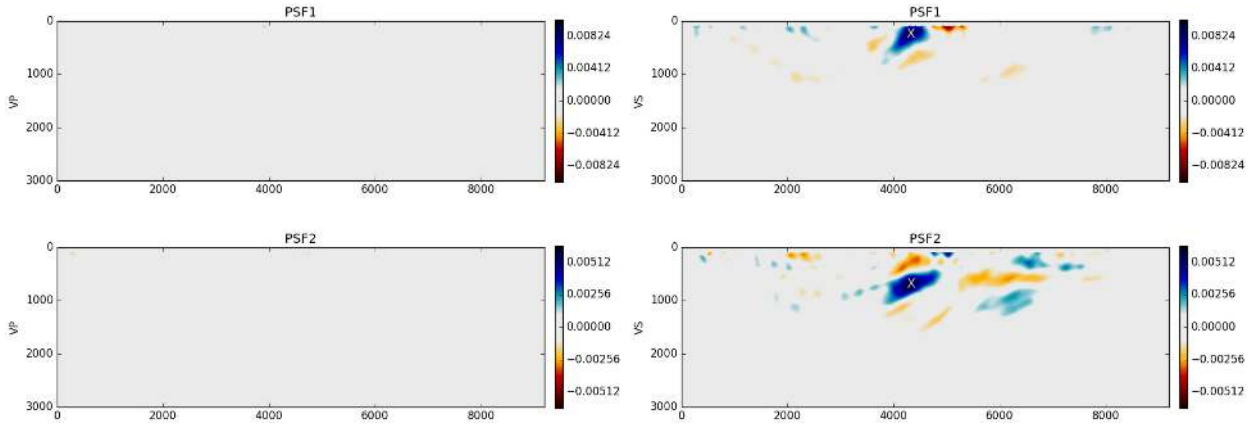

Figure 17. Two V_S PSFs from the estimated Hessian. Associated 'delta-like perturbations' are located at the V_S models, as indicated by the two 'crosses', respectively.

models are shown in Figs 1 and 2. We have an acquisition system placed 10 m below the surface, with 32 shots and 500 receivers distributed evenly. The source function is a Ricker wavelet with dominant frequency at 4 Hz. The elastic FWI meets the stopping criterion at the 46th iteration. Its normalized data misfit reduces to a value of 0.02. The inverted models are shown in Fig. 2. We exhibit more details about the comparisons between L-BFGS-based and SRVM-based elastic FWIs in different cases in Liu *et al.* (2018). Also, in Liu *et al.* (2018), after 46 iterations, we store 46 SRVM scalars and vectors in the elastic Marmousi test. We will use these stored scalars and vectors to reconstruct the inverse Hessian and to access to the posterior covariance and the Hessian.

We begin by evaluating the memory consumption of SRVM vectors in our method. The full (inverse) Hessian is prohibitive to build because of its vast size. In our method, assuming that the model size is M, the full Hessian is of size $M \times M$, whereas the required memory by SRVM is $M \times N_{\text{iter}}$. Here N_{iter} denotes the number of iterations. The memory requirement of the N_{iter} SRVM scalars is negligible; usually, N_{iter} is far less than M. Although Algorithm (2) can be used to probe the posterior covariance, it is not very efficient in, for example, extracting the diagonals. To facilitate the matrix probings and operations, we incorporate a randomized SVD (Halko *et al.* 2011) into the SRVM algorithm by combining Algorithm (2) with Algorithm (3). In randomized SVD, we employ one sampler set composed of 46 independent 1-D random vectors. We then represent the inverse Hessian in a SVD form. Note that the rank order of the factorized matrix by randomized SVD is the same as that of its original matrix.

The eigenvalues of the inverse Hessian $\mathbf{H}^{-1}$ are shown in Fig. 3. In Fig. 4 we display four eigenvectors associated to different eigenvalues retrieved, including the strongest and the weakest one. Note that the retrieved eigenvalues cover a range of ∼4 orders of magnitude while their eigenvectors are normalized. We observe from Fig. 4 that as the order of the eigenvector increases, the energy distribution in the eigenvectors moves gradually from the bottom to the top of the subsurface model. We understand this trend from the inverse of $\mathbf{H}^{-1}$, that is the Hessian, whose energy distribution should fade away from the surface to the bottom within this reflection acquisition setup. The trends in Fig. 4 also imply that it is possible to obtain the approximated Hessian by taking the inverse of $\mathbf{H}^{-1}$, which we will show next.

After obtaining the eigenvalues and eigenvectors, we can extract the diagonal of $\mathbf{H}^{-1}$ easily. The extracted square-root diagonals of the inverse Hessian are presented in Fig. 5 as a 2-D 'standard deviation map'. Both the magnitude and the appearance of our map resemble those of Fig. 5 of Trinh *et al.* (2017), who uses the Ensemble Kalman Filter (EnKF) method. This map provides a quantitative way to evaluate the uncertainty of inversion results. According to the physical interpretation of the Hessian, its energy decreases gradually with depth due to geometrical spreading effects. We can thus infer that the energy of the inverse Hessian should counteract this effect and basically increase with depth. We also note that certain high-velocity structures have high uncertainties associated to them. This phenomenon can be explained in that a high-velocity structure bends the wave path such that the wave energy is more likely to probe only the onset of the region but fails to sample the full area. In other words, it reflects the experience in FWI that subsurface areas with better data coverage will be better constrained. Note that for some very marginal areas (i.e. the absorbing layers), their uncertainties become zero. This occurs because in our method we can not adjust areas where the Fréchet kernels are always zero. Similar explanations are also presented in Kennett *et al.* (1988) and Rawlinson *et al.* (2014).

We also compare the standard deviation maps between V_P and V_S and notice that the map of V_P often has larger standard deviations than V_S. The difference between them can be explained from the perspective of radiation patterns of different elastic parameters as in Wu & Aki (1985); Tarantola (1986); Virieux *et al.* (2009). In their views, the radiation pattern of V_P is isotropic whereas that of V_S is far-offset dominated. This means, in the simultaneous elastic inversion, the V_S kernel possesses more low-wavenumber components than the V_P kernel. Low-wavenumber components however often have stronger amplitudes (Zhang & Sun 2009) and less chance to fall in a local minimum (Virieux *et al.* 2009).

Herein, we leave the estimation of the prior information as an open question and instead set the same standard deviation value of 250 m s^{-1} everywhere. With this value and under a Gaussian distribution, the term $\mathbf{C}_m^{1/2}\mathbf{n}$ in eq. (21) ranges between [−1000, 1000] m s^{-1}. The 2-D Gaussian random fields for V_P and V_S are shown in Fig. 6. We generate 1000 such random fields independently and use them for the sampling. We visualize prior distributions of V_P and V_S against depth at three different model positions in Figs 7 and 8, respectively. It is clear that in Figs 7 and 8 the same uncertainty occurs everywhere. We then infer the posterior distribution with eq. (22). The resulting posterior V_P and V_S distributions against depth can be found in Figs 9 and 10. We notice that the

visualized distributions are not homogeneous anymore, and reflect the different standard deviation values at different depth points. Thus, the heterogeneities indicate the point-wise uncertainties of the posterior model.

For detailed investigations, we pick out six points, respectively, from the prior and posterior V_P and V_S distributions, as shown in Figs 11 and 12. In these figures, both the prior and posterior distributions have Gaussian shapes, as expected due to the sampling and approximations used. Still, several point distributions nicely indicate a shift between the prior mean and posterior mean which is due to the elastic FWI and choice of initial model. We also notice that the posterior distributions are more 'concentrated', that is, have smaller standard deviations, than the prior ones, meaning that the posterior model involves less uncertainties than does the prior model. Therefore, the data inversion led to an information gain and better constrains the subsurface model parameters at these locations.

For a full 2-D plane view of how the $\mathbf{C}_M^{1/2}\mathbf{n}$ looks like in eq. (23), we draw four different samples as shown in Fig. 13. Each sample is obtained by a matrix-vector multiplication with different $\mathbf{n}$. We note that the posterior model perturbations are spatially continuous. Because the prior model perturbations are computed by $\mathbf{C}_m^{1/2}\mathbf{n}$, in which we make a homogeneous diagonal approximation to $\mathbf{C}_m^{1/2}$, the distribution of $\mathbf{C}_m^{1/2}\mathbf{n}$ should be similar to $\mathbf{n}$ but only differ in magnitude. Thus, they would be similar to those in Fig. 6 and we omit to plot them again. As for $\mathbf{C}_M^{1/2}\mathbf{n}$, an interesting aspect is that the posterior samples have the pattern of the MAP model. The variance of $\mathbf{C}_M^{1/2}\mathbf{n}$ between different samples helps us to analyse the posterior uncertainties. We also show the posterior samples denoted by eq. (22) in Fig. 14.

To utilize the second-order information by SRVM, we approximate the Hessian with eq. (15). The square-root diagonals of the Hessian are shown in Fig. 15. We take the square roots because according to its original definition, the Hessian contains four-fold geometrical spreading effects. Thus, the square-root form gives a better visualization of the spatial variance of Hessian diagonals. Fig. 15 underscores the physical interpretation: energy strength decreases from top to bottom, including effects of the elastic subsurface structure. Furthermore, this relates to the source illumination and will have an effect on the point-spread function analysis in the following.

We can extract point-spread functions from eq. (15) conveniently, where a point-spread function (PSF) represents one row (column) from the Hessian. To demonstrate such examples, two V_P and V_S PSFs are displayed in Figs 16 and 17, respectively, where the 'crosses' mark the row (column) indexes. These cross locations can be interpreted as the sites where a localized 'perturbation' to the prior model parameter would occur and the PSFs then reflect the 'smearing' of the perturbation due to an imperfect resolution operator in the inversion. In case the inversion resolution would be perfect, the same point location perturbation would be retrieved by the PSFs. We observe that the PSFs usually have some orientation-dependent shapes. Also, the plotted PSFs are concentrated locally close to the original 'perturbation' locations. Still, for such an exploration setup, it would be difficult to assign a confidence region in form of an ellipsoid around the PSF to estimate resolution lengths in horizontal/vertical directions. Since in our examples the chosen locations are somewhat close to the surface and thus in areas where the Hessian is dominated by the diagonal elements, the PSFs show little trade-offs between V_P and V_S parameters. By comparing V_P and V_S PSFs, we notice that the V_S PSFs are more focused and have weaker trade-offs compared to the V_P ones. This relates to the Hessian, as shown in Fig. 15, where the strengths of the diagonals help to explain why the V_S model is usually better resolved than the V_P model.

4 DISCUSSION

Results of geophysical inversions have inherent uncertainties due to limitations in measurements and theories. Bayesian inference provides an elegant description regarding uncertainties of such geophysical inversions in terms of the posterior PDF in the model space. For elastic FWI, however, a complete characterization of the posterior PDF using statistical sampling methods is prohibitive because of the highly multidimensional model space. One simplistic and feasible approach is to locate the minimum of the misfit function, which usually corresponds to the MAP model, under the assumption of linearized forward modeling operators and Gaussian distributed PDFs. Most of the gradient-optimization methods focus on finding the MAP model, but without uncertainty estimation.

In our elastic FWI study, and within the presented gradient-optimization framework, we can obtain the MAP model together with the posterior covariance for uncertainty estimation using a quasi-Newton algorithm named SRVM. To allow for large-scale applications, we formulate the SRVM algorithm in a vector version to alleviate the memory storage burden. The main advantage then of the SRVM algorithm is that after convergence of the elastic FWI, we can retrieve the information of the posterior covariance from the stored SRVM vector and scalar series. Incorporating the randomized SVD into our SRVM approach facilitates this information retrieval also for large-scale applications. Thus, we can estimate the uncertainty of the MAP model via analysis methods such as point-wise standard deviations, random samplings and point-spread function tests.

Our current study, however, is limited to the assumptions of linearized forward modeling operators and Gaussian distributed PDFs. Note that even with simplified Gaussian priors, the resulting posterior PDF may be non-Gaussian and highly complicated in shape due to the presence of model nonlinearities. Also, gradient-optimization methods often find a minimum in the neighborhood of the starting model. In situations such as FWI, where the misfit function has several secondary minima, one needs to either choose multiscale strategies or start close to the global minimum to overcome such shortcomings. We will further explore and address these limitations in future research.

5 CONCLUSIONS

This paper, Part 2 of our study, concerns the extension of the theories and methods on SRVM-based elastic FWI presented in Part 1 of our study. Herein, we estimate the uncertainty of the maximum *a posteriori* model, in which the challenges are high-dimensional model spaces, multiparameters, and computational limitations. We extract the information about the inverse Hessian from the SRVM vector series during elastic FWI. Since the SRVM algorithm is embedded in the elastic FWI, the SRVM vector series inherently captures the information about multiparameters. The computational costs are determined by the SRVM-based elastic FWI, which is similar cost-effective as an L-BFGS-based elastic FWI.

To analyse and manipulate the inverse Hessian efficiently, we factorize it into eigenvalues and eigenvectors using randomized SVD. Based on the randomized SVD results, we can also approximate

the Hessian, and thus have an elegant access to point-spread functions. Furthermore, we relate the inverse Hessian with the posterior covariance and assess the uncertainty of the posterior model parameters based on point-wise standard deviations, random samplings and point-spread function tests.

ACKNOWLEDGEMENTS

The authors are grateful to editor Jean Virieux and reviewer Andreas Fichtner and an anonymous reviewer for improving the initial manuscript. The authors are grateful to Carl Tape for inspiring discussions and valuable inputs to improve the manuscript. This work was supported by the King Abdullah University of Science & Technology (KAUST) Office of Sponsored Research (OSR) under award No. UAPN#2605-CRG4. Computational resources were provided by the Information Technology Division and Extreme Computing Research Center (ECRC) at KAUST.

REFERENCES

An, M., 2012. A simple method for determining the spatial resolution of a general inverse problem, *J. geophys. Int.,* **191**(2), 849–864.

Bamberger, A., Chavent, G. & Lailly, P., 1979. About the stability of the inverse problem in 1-D wave equationsapplication to the interpretation of seismic profiles, *Appl. Math. Opt.,* **5**(1), 1–47.

Biswas, R. & Sen, M., 2017. 2-D full-waveform inversion and uncertainty estimation using the reversible jump hamiltonian Monte Carlo, in *SEG Technical Program Expanded Abstracts 2017,* pp. 1280–1285, Society of Exploration Geophysicists.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *J. geophys. Int.,* **178**(3), 1411–1436.

Bui-Thanh, T., Ghattas, O., Martin, J. & Stadler, G., 2013. A computational framework for infinite-dimensional Bayesian inverse problems Part I: the linearized case, with application to global seismic inversion, *SIAM J. Scient. Comput.,* **35**(6), A2494–A2523.

Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics,* **60**(5), 1457–1473.

Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic nonlinear waveform inversion: aplication to real data, *Geophysics,* **55**(5), 527–538.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. geophys. Res.: Oceans,* **99**(C5), 10 143–10 162.

Faccioli, E., Maggio, F., Quarteroni, A. & Taghan, A., 1996. Spectral-domain decomposition methods for the solution of acoustic and elastic wave equations, *Geophysics,* **61**(4), 1160–1174.

Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques, *J. geophys. Int.,* **185**(2), 775–798.

Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *J. geophys. Int.,* **187**(3), 1604–1624.

Fichtner, A. & van Leeuwen, T., 2015. Resolution analysis by random probing, *J. geophys. Res.: Solid Earth,* **120**(8), 5549–5573.

Fletcher, R. & Powell, M.J., 1963. A rapidly convergent descent method for minimization, *Comput. J.,* **6**(2), 163–168.

Fomel, S. & Claerbout, J.F., 2003. Multidimensional recursive filter preconditioning in geophysical estimation problems, *Geophysics,* **68**(2), 577–588.

Gauss, C.F., 1809. *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium,* Vol. **7,** Perthes et Besser.

Gouveia, W.P. & Scales, J.A., 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, *J. geophys. Res.: Solid Earth,* **103**(B2), 2759–2779.

Halko, N., Martinsson, P.-G. & Tropp, J.A., 2011. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.,* **53**(2), 217–288.

Hull, D.G. & Tapley, B.D., 1977. Square-root variable-metric methods for minimization, *J. Opt. Theory Applicat.,* **21**(3), 251–259.

Igel, H., Djikpéssé, H. & Tarantola, A., 1996. Waveform inverstion of marine reflection seismograms for p impedance and poisson's ratio, *J. geophys. Int.,* **124**(2), 363–371.

Jordan, M., 2015. Estimation of spatial uncertainties in tomographic images, in *Proceedings of the 77th EAGE Conference and Exhibition - Workshops,* EAGE, doi:10.3997/2214-4609.201413555.

Kalman, R.E. *et al.*, 1960. A new approach to linear filtering and prediction problems, *J. Basic Eng.,* **82**(1), 35–45.

Käufl, P., Fichtner, A. & Igel, H., 2013. Probabilistic full waveform inversion based on tectonic regionalizationdevelopment and application to the australian upper mantle, *J. geophys. Int.,* **193**(1), 437–451.

Kennett, B., Sambridge, M. & Williamson, P., 1988. Subspace methods for large inverse problems with multiple parameter classes, *J. geophys. Int.,* **94**(2), 237–247.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *J. geophys. Int.,* **139**(3), 806–822.

Komatitsch, D. & Vilotte, J.P., 1998. The spectral-element method: an efficient tool to simulate the seismic response of 2-D and 3-D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Lailly, P., 1983. The seismic inverse problem as a sequence of before-stack migrations, in *Proceedings of the Conference on Inverse Scattering, Theory and Application Expanded Abstracts,* pp. 206–220, Society of Industrial and Applied Mathematics, Philadelphia, PA, USA.

Liberty, E., Woolfe, F., Martinsson, P.-G., Rokhlin, V. & Tygert, M., 2007. Randomized algorithms for the low-rank approximation of matrices, *Proc. Natl. Acad. Sci.,* **104**(51), 20 167–20 172.

Liu, Q., Peter, D. & Tape, C., 2018. Square-root variable metric based elastic full-waveform inversion – Part 1: theory and validation, *Geophys. J. Int.,* **218** (2), 1121–1135.

Martin, G.S., Wiley, R. & Marfurt, K.J., 2006. Marmousi2: an elastic upgrade for marmousi, *Leading Edge,* **25**(2), 156–166.

Métivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full waveform inversion and the truncated Newton method, *SIAM J. Scient. Comput.,* **35**(2), B401–B437.

Modrak, R. & Tromp, J., 2016. Seismic waveform inversion best practices: regional, global and exploration test cases, *J. geophys. Int.,* **206**(3), 1864–1889.

Modrak, R.T., Borisov, D., Lefebvre, M. & Tromp, J., 2018. Seisflowsflexible waveform inversion software, *Comput. Geosci.,* **115,** 88–95.

Morf, M. & Kailath, T., 1975. Square-root algorithms for least-squares estimation, *IEEE Trans. Automat. Contl.,* **20**(4), 487–497.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.: Solid Earth,* **100**(B7), 12 431–12 447.

Nocedal, J. & Wright, S., 2006. *Numerical Optimization,* Springer Series in Operations Research and Financial Engineering, 2nd edn, Springer-Verlag.

Osypov, K. *et al.*, 2013. Model-uncertainty quantification in seismic tomography: method and applications, *Geophys. Prospect.,* **61**(6), 1114–1134.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain—Part 1: theory and verification in a physical scale model, *Geophysics,* **64**(3), 888–901.

Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Chapter one-seismic tomography and the assessment of uncertainty, *Adv. Geophys.,* **55,** 1–76.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), doi:10.1029/2000RG000089.

Sambridge, M., Rickwood, P., Rawlinson, N. & Sommacal, S., 2007. Automatic differentiation in geophysical inverse problems, *Geophys. J. Int.,* **170**(1), 1–8.

Seriani, G.z. & Priolo, E., 1994. Spectral element method for acoustic wave simulation in heterogeneous media, *Finite Elem. Anal. Des.,* **16**(3), 337–348.

Tan, L., Brytik, V., Baumstein, A. & Hinkley, D., 2010. Verification of gradient and Hessian computation for full wavefield inversion using automatic

differentiation, in *SEG Technical Program Expanded Abstracts,* Society of Exploration Geophysicists, pp. 2762–2766.
Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49,** 1259–1266.
Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics,* **51**(10), 1893–1903.
Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation,* Elsevier Science Publ. Co., Inc.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics.
Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.,* **20**(2), 219–232.
Thurin, J., Brossier, R. & Metivier, L., 2017. Ensemble-based uncertainty estimation in full waveform inversion, in *Proceedings of the 79th EAGE Conference and Exhibition 2017,* EAGE, doi:10.3997/2214-4609.201701007.
Trampert, J., Fichtner, A. & Ritsema, J., 2012. Resolution tests revisited: the power of random numbers, *J. geophys. Int.,* **192**(2), 676–680.
Trinh, P., Brossier, R., Métivier, L., Virieux, J. & Wellington, P., 2017. Bessel smoothing filter for spectral-element mesh, *J. geophys. Int.,* **209**(3), 1489–1512.
Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *J. geophys. Int.,* **160**(1), 195–216.
Virieux, A. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.
Virieux, J., Operto, S., Ben-Hadj-Ali, H., Brossier, R., Etienne, V., Sourbier, F., Giraud, L. & Haidar, A., 2009. Seismic wave modeling for seismic imaging, *Leading Edge,* **28**(5), 538–544.
Vlasenko, A., Köhl, A. & Stammer, D., 2016. The efficiency of geophysical adjoint codes generated by automatic differentiation tools, *Comput. Phys. Commun.,* **199,** 22–28.
Warner, M. & Guasch, L., 2016. Adaptive waveform inversion: theory, *Geophysics,* **81**(6), R429–R445.
Williamson, W.E., 1975. Square-root algorithms for function minimization, *AIAA J.,* **13**(1), 107–109.
Wu, R.-S. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics,* **50**(4), 582–595.
Zhang, J. & McMechan, G.A., 1995. Estimation of resolution and covariance for large matrix inversions, *J. geophys. Int.,* **121**(2), 409–426.
Zhang, Y. & Sun, J., 2009. Practical issues of reverse time migration: True amplitude gathers, noise removal and harmonic-source encoding, in *Beijing International Geophysical Conference and Exposition 2009: Beijing 2009 International Geophysical Conference and Exposition, Beijing, China, 24–27 April 2009,* pp. 204–204, Society of Exploration Geophysicists.
Zhu, H., Li, S., Fomel, S., Stadler, G. & Ghattas, O., 2016. A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration, *Geophysics,* **81**(5), R307–R323.

APPENDIX A: SINGLE-PASS RANDOMIZED SVD

When an input matrix **Z** is too large, it might become infeasible for users to revisit it during randomized SVD. Therefore, Halko *et al.* (2011) developed the single-pass randomized SVD, which requires just one pass over the input matrix.

In our presented Algorithm (3), the input matrix Z is a symmetric $M \times M$ matrix. After convergence of the SRVM-based elastic FWI, we have N_r SRVM vectors to approximate the inverse data-misfit Hessian, with N_r being the total iteration number. Therefore, we use a set of $M \times N_r$ randoms samplers for the factorization of **Z**, whose rank can not be greater than N_r.

When the QR decomposition in Algorithm (3) is done, we can define a matrix **A** as

$$\mathbf{A} = \mathbf{Q}^T\mathbf{Z}\mathbf{Q} \tag{A1}$$

using the orthogonal matrix **Q** and symmetric matrix **Z**. Note that $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$. Post-multiplying both sides of eq. (A1) by $\mathbf{Q}^T\mathbf{X}$ yields

$$\mathbf{A}\mathbf{Q}^T\mathbf{X} = \mathbf{Q}^T\mathbf{Z}\mathbf{Q}\mathbf{Q}^T\mathbf{X} \tag{A2}$$

in which $\mathbf{Z} = \mathbf{Z}(\mathbf{Q}\mathbf{Q}^T)$ and $\mathbf{Y} = \mathbf{Z}\mathbf{X}$. From this, we have

$$\mathbf{A}(\mathbf{Q}^T\mathbf{X}) = \mathbf{Q}^T\mathbf{Y}. \tag{A3}$$

which is the third line in Algorithm (3). After solving for **A**, we can obtain a low-rank factorization

$$\mathbf{Z} = \mathbf{Q}\mathbf{A}\mathbf{Q}^T, \tag{A4}$$

which can lead to a SVD problem

$$\mathbf{Z} = \mathbf{V}\Lambda\mathbf{V}^T, \tag{A5}$$

with $\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^T$ and $\mathbf{V} = \mathbf{Q}\mathbf{U}$.

APPENDIX B: CLASSICAL RESOLUTION MODEL TEST CASE: 2-D ELASTIC CHECKERBOARD MODEL

In the 2-D elastic Checkerboard test, the models are 9000 m long and 3000 m deep. The true and initial models are shown in Figs B1 and B2. We have an acquisition system located 10 m below the surface, with 32 shots and 500 receivers distributed evenly. The source function is a Ricker wavelet with dominant frequency at 4 Hz. The elastic FWI converges at the 100th iteration. Its data misfit reduces to a value of 0.001. The inverted models are shown in Fig. B2. A detailed comparisons between L-BFGS and SRVM-based elastic FWIs for different cases can be found in Part 1 of our study (Liu *et al.* 2018).

We handle the 2-D elastic Checkerboard model similarly to the 2-D Marmousi model. One notable difference between the two models are their PDFs. The Checkerboard PDFs are more spread out. This may be caused by more reflection waves existing in the Checkerboard test. For better illustration, we show all related tests in Figs B1–B16.

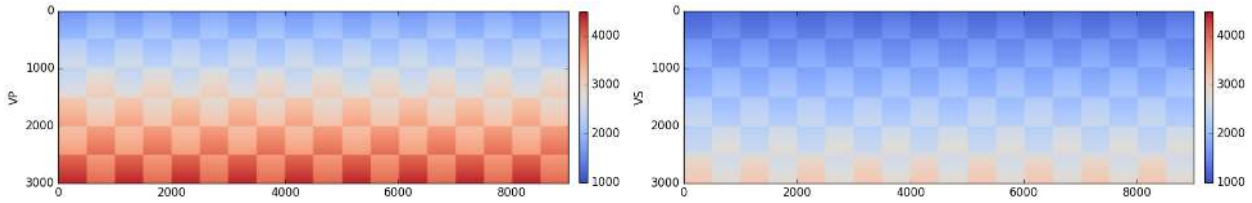

Figure B1. True V_P and V_S models used for data generation in the elastic Checkerboard test.

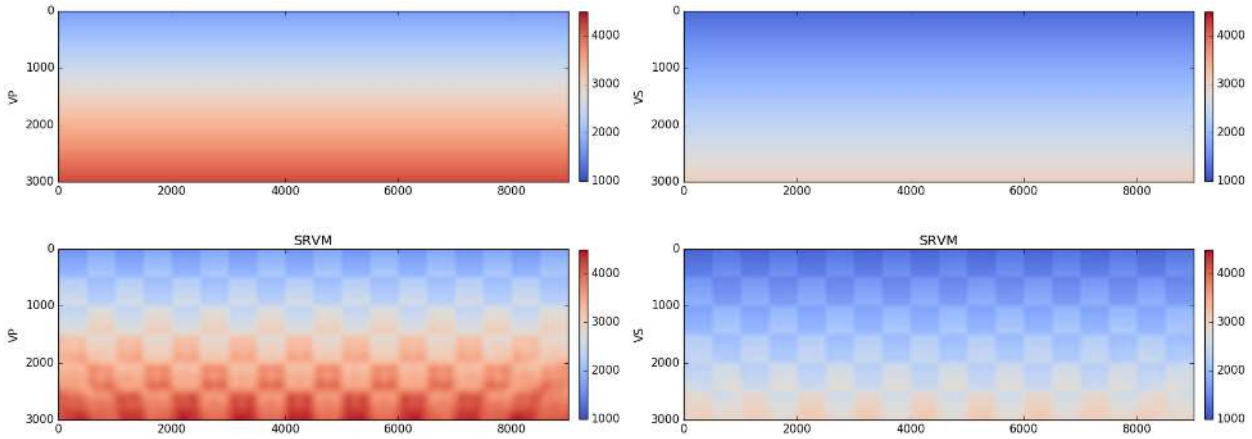

Figure B2. Initial and inverted V_P and V_S models in elastic Checkerboard test. Top panel: prior V_P and V_S models obtained by smoothing the true models in Fig. 1. Bottom panel: inverted V_P and V_S models by SRVM based elastic FWI after 100 iterations (details about data and model misfits can be found in Part 1 of our study).

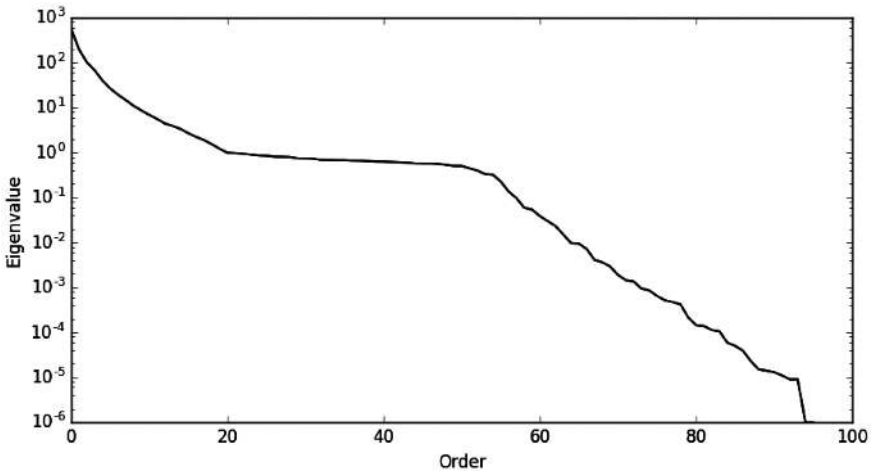

Figure B3. Eigenvalues extracted by randomized SVD algorithm from the stored 46 SRVM vectors in the elastic Checkerboard test.

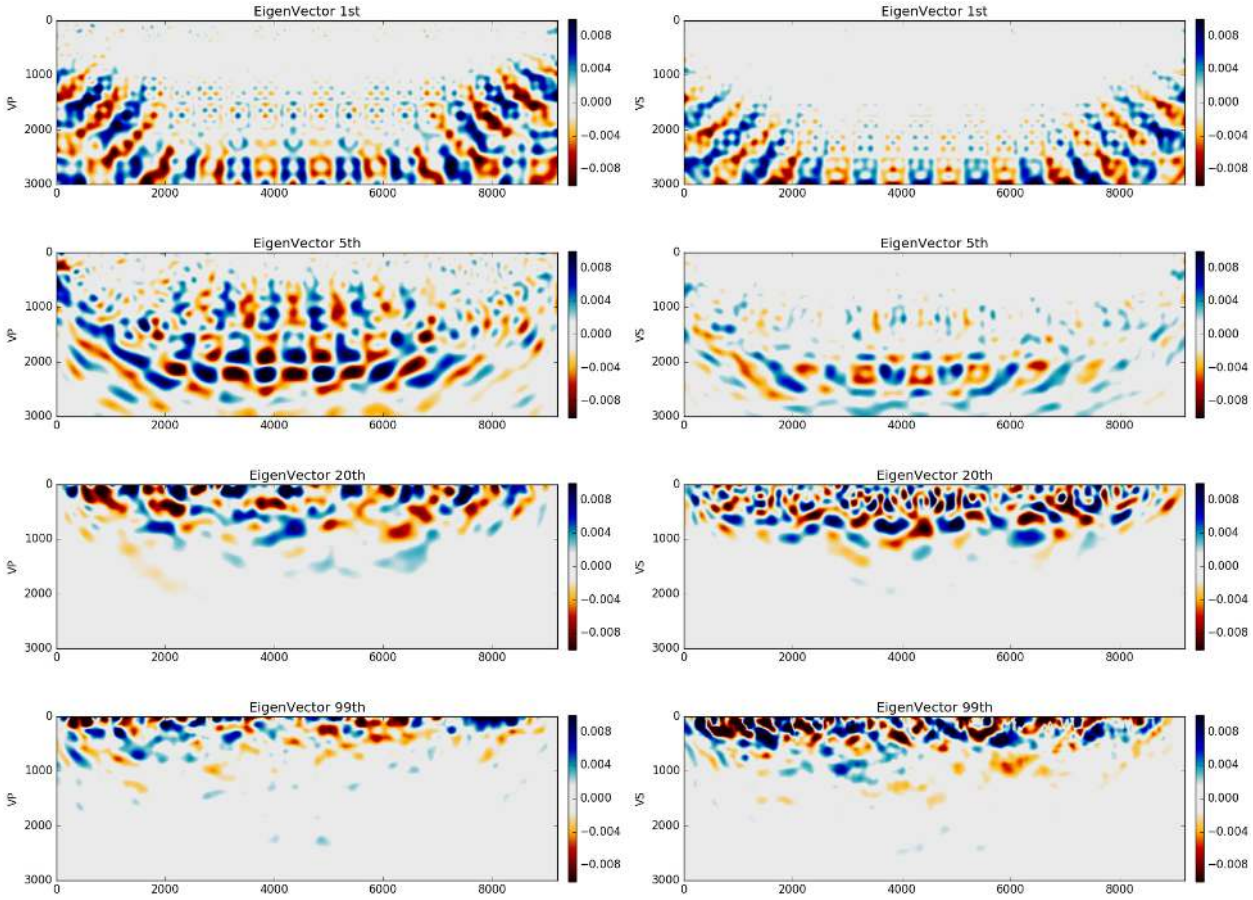

Figure B4. Eigenvectors factorized by randomized SVD algorithm from the stored 100 SRVM vectors in the elastic Checkerboard test.

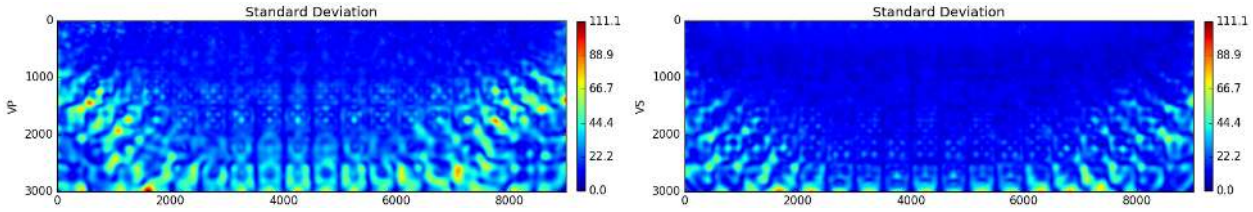

Figure B5. Standard deviation from the posterior covariance in the elastic Checkerboard test.

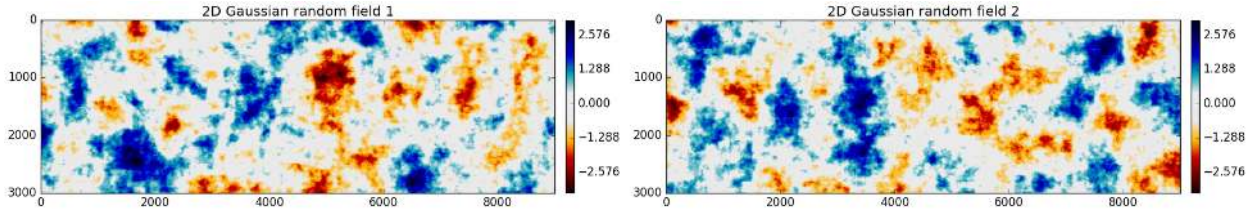

Figure B6. Examples of 2-D Gaussian random fields, with zero mean and unit variance. These are used to sample the prior and posterior distributions. Note that we draw 1000 such samplers, and each of them is independent of the others. The 2-D random fields are only for the sampling of distributions. As for the sampling in randomized SVD, 1-D randoms are qualified enough, the generation of which is more cost-effective.

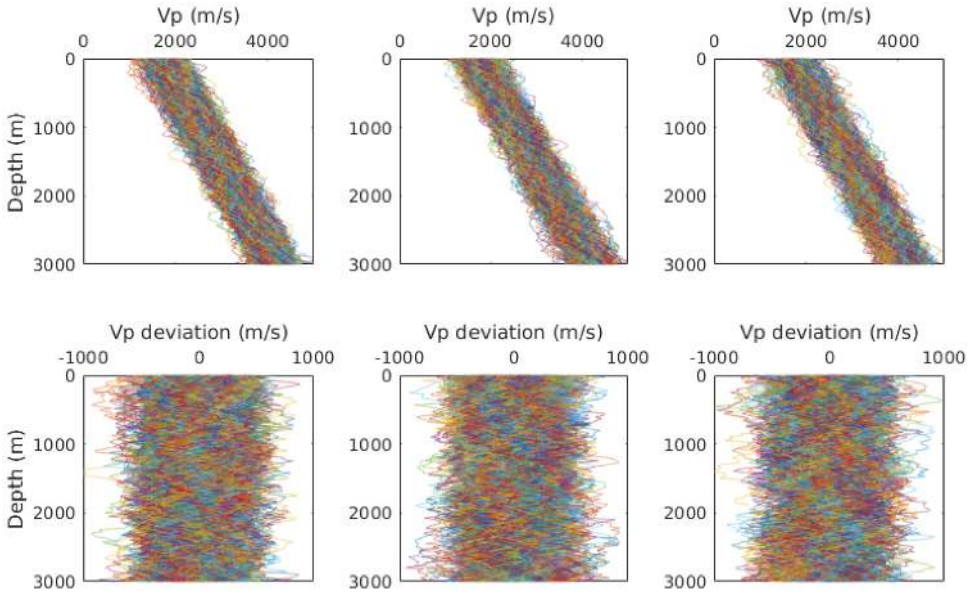

Figure B7. Prior distributions of V_P in the elastic Checkerboard test. Top row (from left- to right-hand) panel: prior distributions against depth of the V_P model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: prior distributions against depth of the V_P deviation at $X = 3100$, 4600, 6900 m.

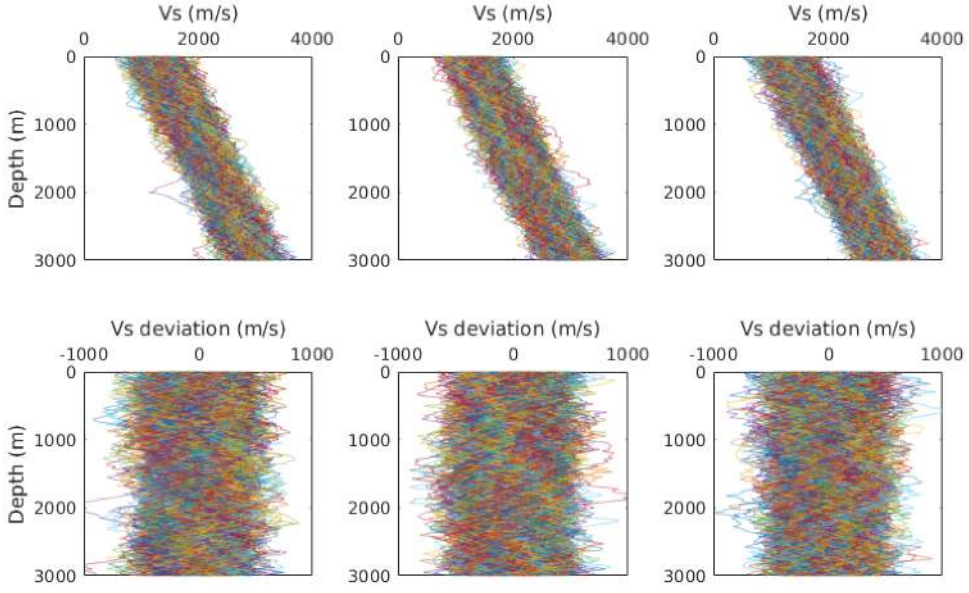

Figure B8. Prior distributions of V_S in the elastic Checkerboard test. Top row (from left- to right-hand) panel: prior distributions against depth of the V_S model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: prior distributions against depth of the V_S deviation at $X = 3100$, 4600, 6900 m.

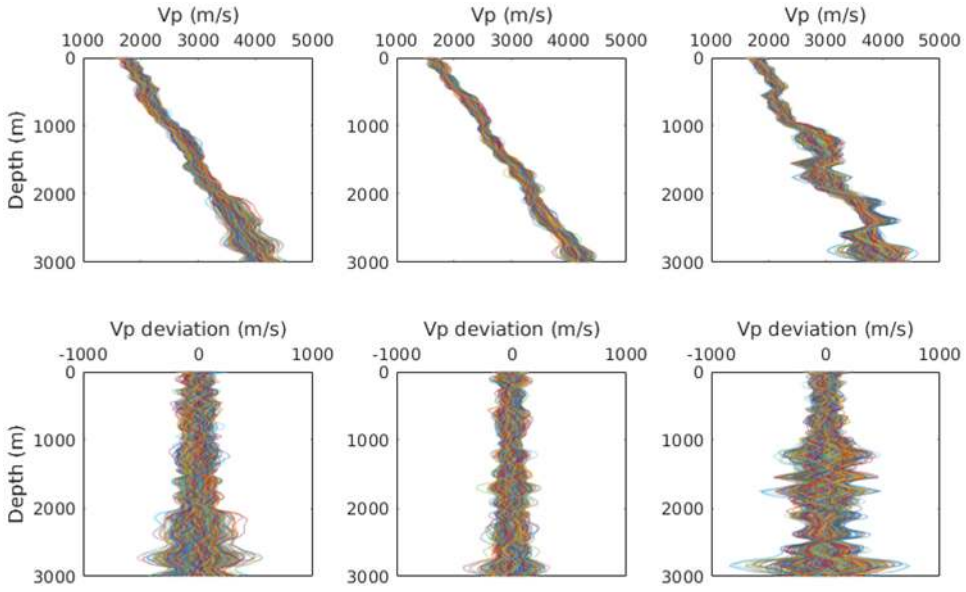

Figure B9. Posterior distributions of V_P in the elastic Checkerboard test. Top row (from left- to right-hand) panel: prior distributions against depth of the V_P model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: posterior distributions against depth of the V_P deviation at $X = 3100$, 4600, 6900 m.

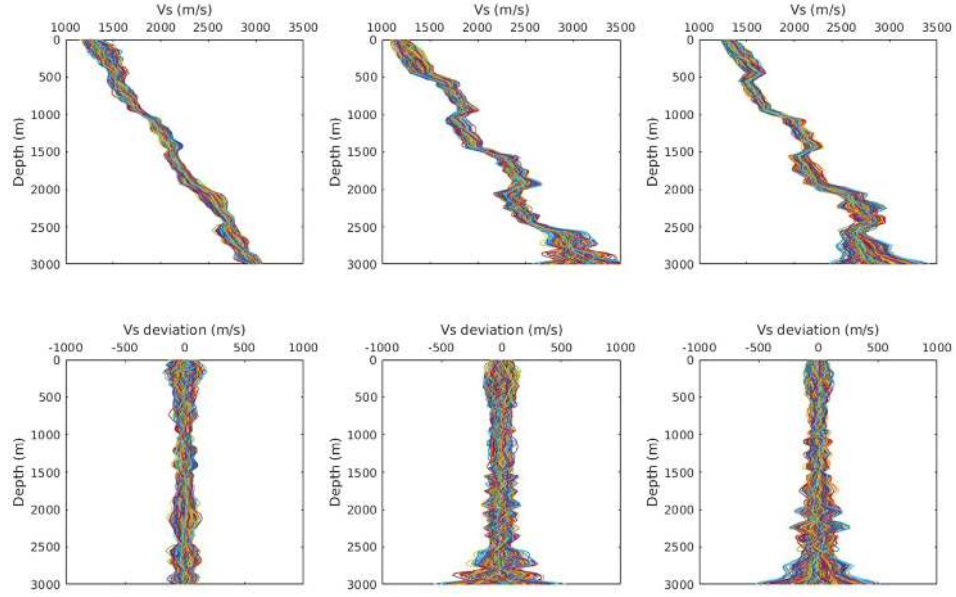

Figure B10. Posterior distributions of V_S in the elastic Checkerboard test. Top row (from left- to right-hand) panel: prior distributions against depth of the V_S model at $X = 3100$, 4600, 6900 m. Bottom row (from left- to right-hand) panel: posterior distributions against depth of the V_S deviation at $X = 3100$, 4600, 6900 m.

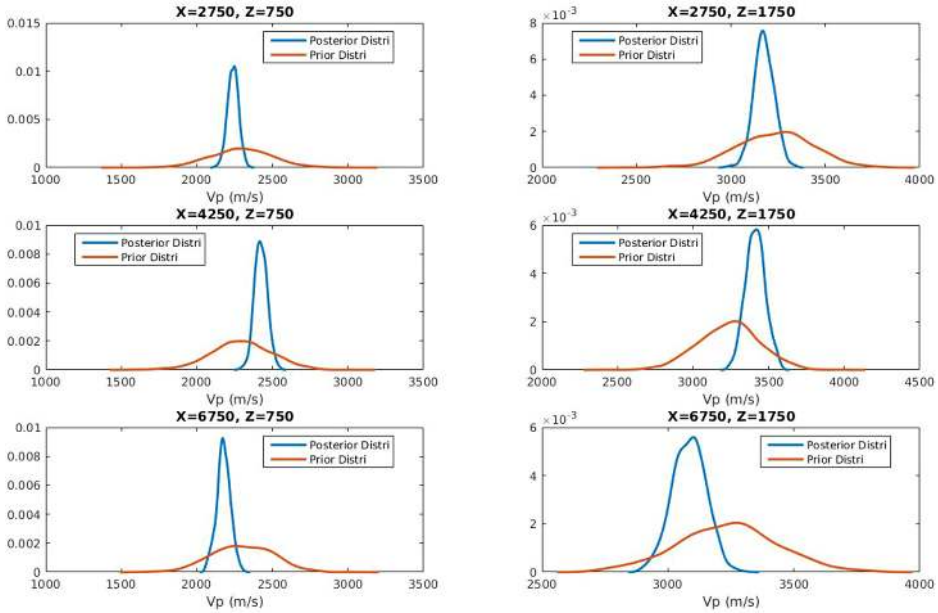

Figure B11. Samples for the comparisons of prior and posterior V_P distributions at different locations in the elastic Checkerboard test. Individual locations are titled on each subplot.

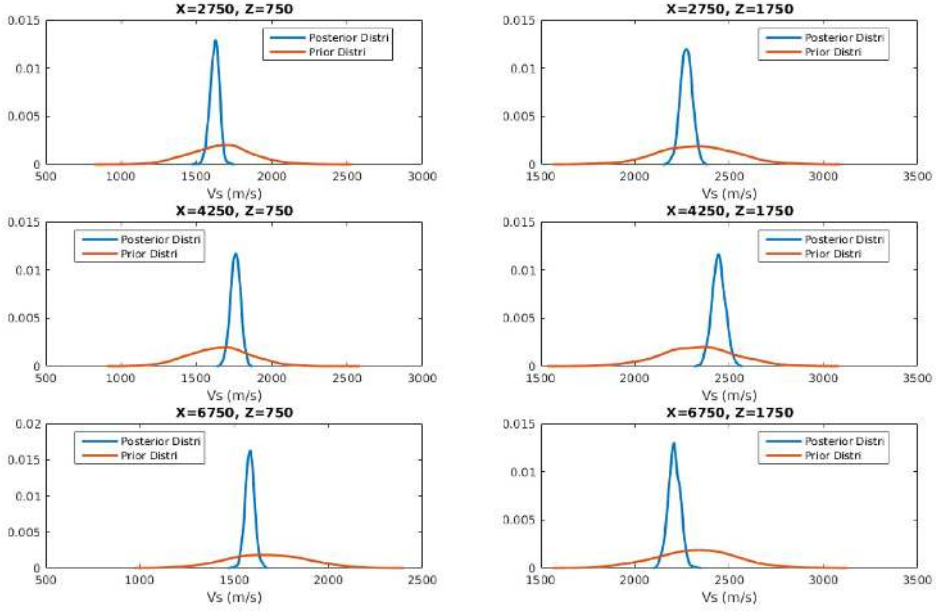

Figure B12. Samples for the comparisons of prior and posterior V_S distributions at different locations in the elastic Checkerboard test. Individual locations are titled on each subplot.

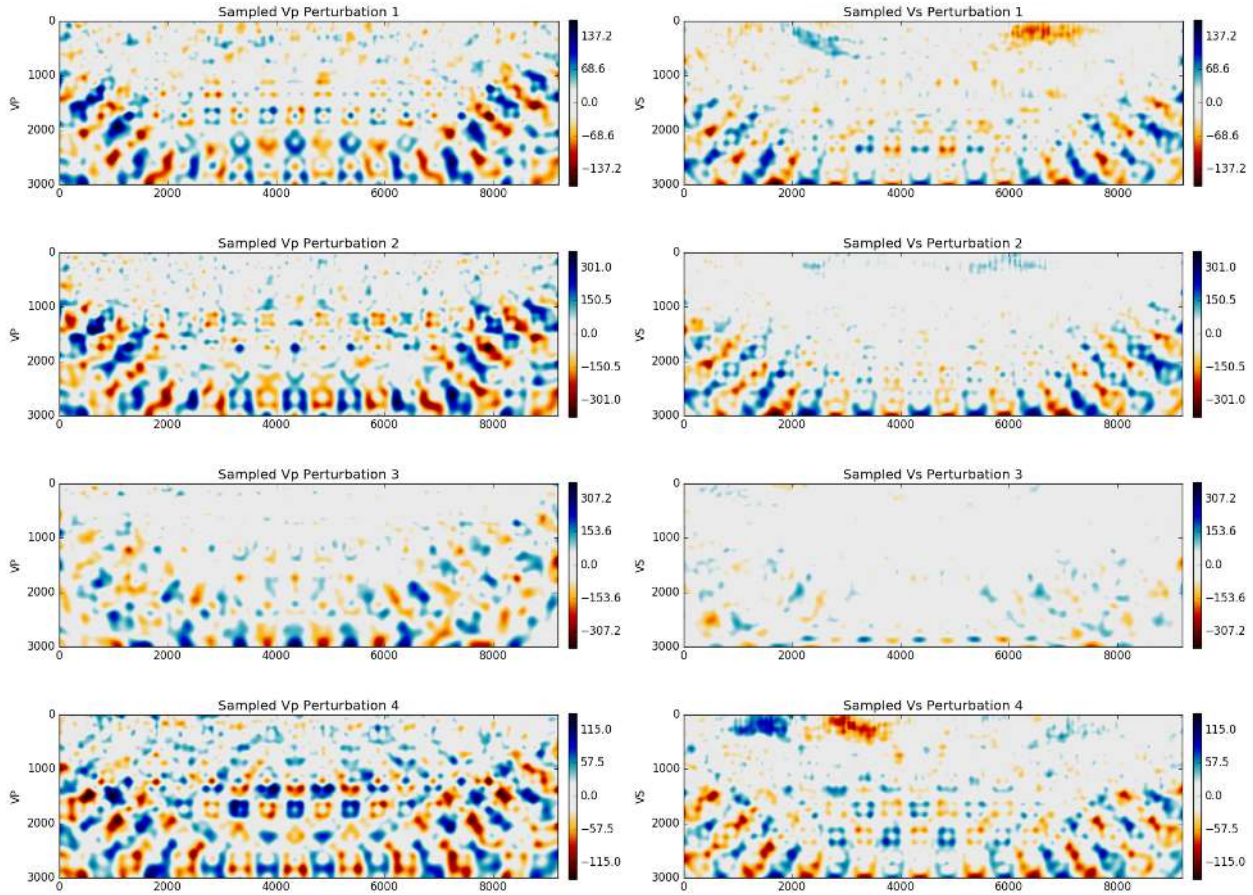

Figure B13. Random samples drawn from the posterior distributions about the V_P and V_S model perturbations in the elastic Checkerboard test.

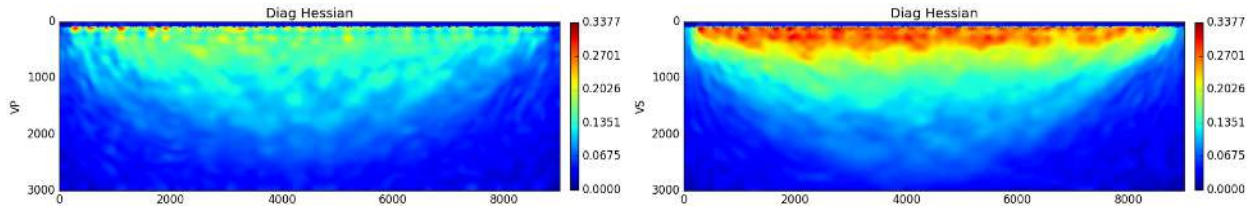

Figure B14. Square root values of Hessian diagonals in the elastic Checkerboard test. We plot the square-root form because the strict calculation of the data-misfit Hessian involves three cross-correlations between four Green's functions. In isotropic elastic media, each Green's function includes the effect of geometrical spreading. Thus the accumulated effect of fourfold geometrical spreading would become considerable.

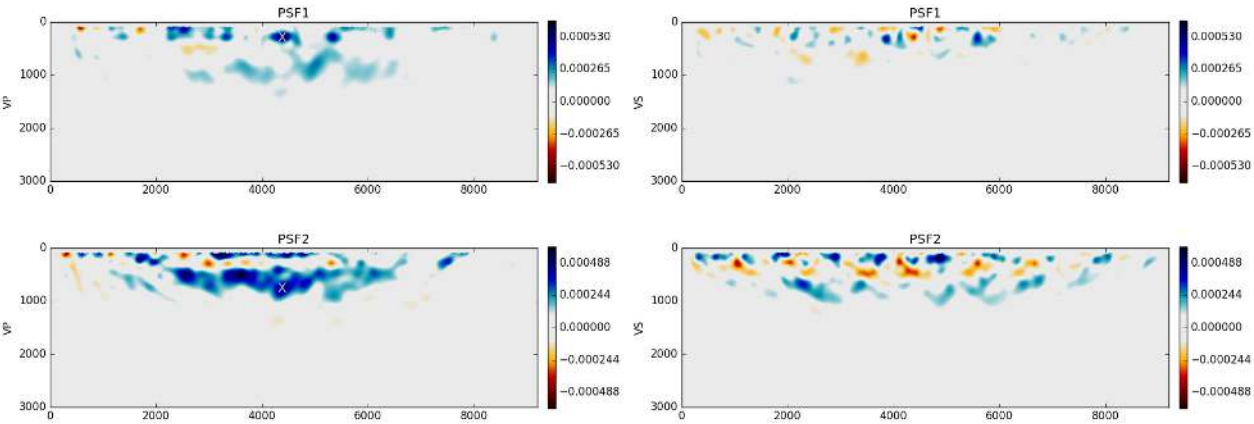

Figure B15. Two V_P PSFs from the estimated Hessian. The 'delta-like perturbations' are located at the V_P models, as indicated by the two 'crosses', respectively.

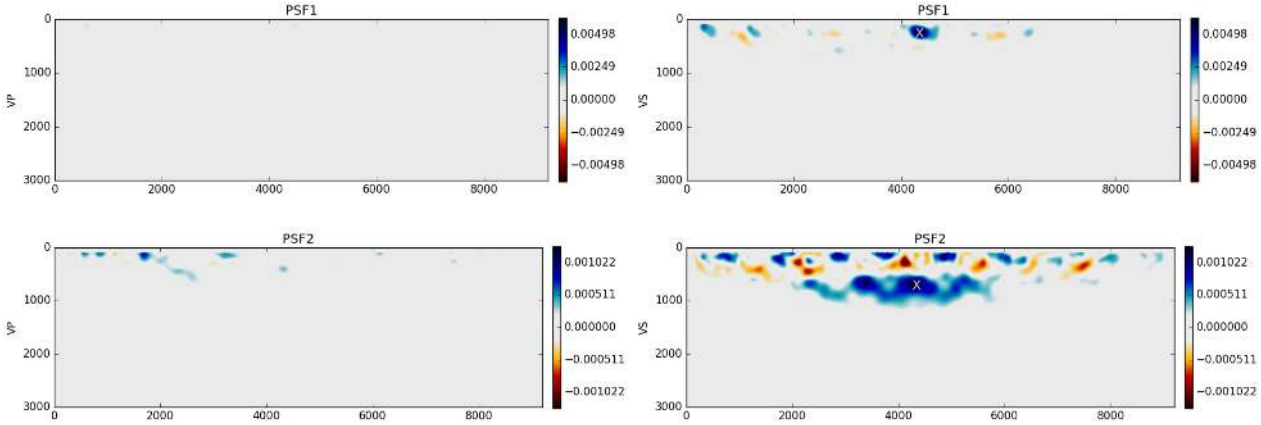

Figure B16. Two V_S PSFs from the estimated Hessian. The 'delta-like perturbations' are located at the V_S models, as indicated by the two 'crosses', respectively.