


Random-Objective Waveform Inversion of 3D-9C Shallow-Seismic Field Data

Yudi Pan¹ , Lingli Gao², and Thomas Bohlen¹

¹Geophysical Institute, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, ²Department of Mathematics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Key Points:

- We present the methodology of random-objective waveform inversion (ROWI) and apply it to a 3D-9C shallow-seismic field data set
- We prove the high robustness of ROWI against random solution paths and poor initial models
- The Pareto solutions of multi-objective subproblems are analyzed in the model space for trade-off analysis

Correspondence to:

Y. Pan,
yudi.pan@kit.edu

Citation:

Pan, Y., Gao, L., & Bohlen, T. (2021). Random-objective waveform inversion of 3D-9C shallow-seismic field data. *Journal of Geophysical Research: Solid Earth*, 126, e2021JB022036. <https://doi.org/10.1029/2021JB022036>

Received 9 MAR 2021
Accepted 23 AUG 2021

Abstract Robustness and uncertainty estimation are two challenging topics in full-waveform inversion (FWI). To overcome these challenges, we present the methodology of random-objective waveform inversion (ROWI), which adopts a multi-objective framework and a preconditioned stochastic gradient descent optimization algorithm. The use of one shot per iteration avoids using redundant data and reduces the computational cost. The Pareto solutions represent a group of most likely solutions and their differences quantifies the model uncertainty associated with the trade-off between conflicting objective functions. Due to the high dimensionality in the data and model spaces, it is prohibitively expensive to check the Pareto optimality of all solutions explicitly. Thus, we decompose the original multi-objective function into shot-related subproblems and use the Pareto solutions of the subproblems for trade-off analysis. We apply ROWI to a field multi-component shallow-seismic data set acquired in Rheinstetten, Germany. The 3D near-surface model is successfully reconstructed by ROWI and the main target, a refilled trench, is delineated. We compare the results estimated by ROWI and a conventional least squares FWI to prove the high efficiency of ROWI. We run six ROWI tests on the field data with different solution paths to prove the robustness of ROWI against the random solution path. The validity of the reconstructed model is verified by multiple 2D ground-penetrating radar profiles. We estimate 246 Pareto solutions of multi-objective subproblems for trade-off analysis. Another four ROWI tests starting from different poor initial models are performed, whose results prove the relatively high robustness of ROWI against the initial model.

Plain Language Summary Seismic waveform contains abundant information about the physical properties of the Earth. The random-objective waveform inversion (ROWI) method provides an efficient way to estimate the Earth model by randomly probing a subset of seismic data and fitting one of its waveform characteristics. We can further evaluate the trade-off in the reconstructed subsurface model by comparing the differences among the admissible results that explain the data equally well. Here, we promote the ROWI method to 3D media and apply it to a 3D 9-component shallow-seismic field data set. The ROWI result nicely reveals the existence of a refilled ancient trench, and the validity of it is verified by multiple ground-penetrating radar profiles. Several ROWI tests with different setups show that ROWI is robust against its stochastic nature and poor starting models. Trade-off analysis shows that the boundaries of the refilled trench and another trench-like structure are less reliable compared with the other parts in the reconstructed model. Overall, the field example proves the validity and efficiency of ROWI in reconstructing the Earth's shallow subsurface.

1. Introduction

Full-waveform inversion (FWI) has gained increasing attention in the reconstruction of Earth's structures. One typical problem FWI faces is the ill-posedness of the inverse problem, which may make FWI converging to a local minimum. A common practical way to mitigate this problem is to use an appropriate initial model and adopt a multiscale strategy to progressively include the waveform of shorter wavelength (Bunks et al., 1995). Another approach is to use a more convex objective function that exhibits less local minima (Bozdağ et al., 2011; Métivier et al., 2018; Yang & Engquist, 2017).

Uncertainty estimation in FWI has become popular in recent years. Different approaches, such as local Hessian-based approach (Fang et al., 2018; Izzatullah et al., 2019; Liu & Peter, 2019; Zhu et al., 2016), global Bayesian approach (Gebraad et al., 2020; Zhang & Curtis, 2020), Bayesian learning via stochastic gradient

© 2021 The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Langevin dynamics (Brosse et al., 2018; Welling & Teh, 2011) have been made to approximate the posterior covariance matrix or to estimate the posterior distribution. Besides, the normalizing flows (Siahkoohi et al., 2021), data-assimilation approach (Thurin et al., 2019), and null-space approach (Keating & Innanen, 2021; Liu & Peter, 2020) have also been used for uncertainty characterization in FWI. A review of the uncertainty analysis in seismic tomography can be found in Rawlinson et al. (2014). One common problem for the uncertainty estimation in FWI is the prohibitively high computational cost due to the curse of high dimensionality and the cost of the forward problem (Curtis & Lomax, 2001). Shigapov (2019) proposed a probabilistic framework that formulates FWI as a deterministic multi-objective optimization problem. Pan et al. (2020) proposed a multi-objective waveform inversion on shallow-seismic data and showed that the distribution of Pareto solution in the model space represents the uncertainty (trade-off) information.

Generally, solving a multi-objective optimization problem is complicated and computationally expensive. Pan and Gao (2020) proposed a random-objective waveform inversion (ROWI) method to invert Rayleigh waves in 2D media by adopting a stochastic gradient descent algorithm and not checking the Pareto optimality in the multi-objective optimization. At every iteration of ROWI, only one shot assigned with one measurement (objective function) is used for model update. The inclusion of multiple objective functions increases the robustness of ROWI against local minima, and the stochastic gradient descent algorithm improves the computational efficiency since only one shot is used per iteration. Due to the huge computational cost in checking Pareto optimality, Pan and Gao (2020) used the final (last-iteration) models estimated by ROWI with different solution paths to approximate Pareto optimal solutions and showed their statistical distribution in the model space for trade-off analysis. This trade-off information shows the level of conflict among the subsets of data and using different objective functions. The Pareto front describes the trade-off among the models which are equally acceptable. It shares a similar idea with uncertainty information estimated by the nullspace (Backus & Gilbert, 1970; Fichtner & Zunino, 2019; Liu & Peter, 2020) that represents the ensemble of admissible solutions. In the nullspace approach, the solutions whose misfits are within a predefined tolerance from the 'best' solution (solution with the lowest misfit value) of a single-objective inverse problem are chosen as the admissible solutions, while in ROWI, the non-dominated Pareto solutions are chosen as the admissible solutions of the multi-objective inverse problem.

In this study, we promote ROWI to 3D media and apply it to a nine-component (9C) shallow-seismic field data acquired in Rheinstetten, Germany. This 3D 9C shallow seismic data has a dense acquisition system, relatively high signal-to-noise ratio, and high redundancy. Thus, it provides a convenient situation to study the performance of ROWI. The methodology of ROWI in 3D media is briefly introduced. We propose to decompose the original multi-objective function into multiple shot-related subproblems and use the Pareto solutions of the subproblems for trade-off analysis. We apply both conventional least squares FWI and ROWI on the field data to compare their efficiency. We run six ROWI tests with different random sequences and compare their final results to evaluate the robustness of ROWI against random solution paths. We also apply both conventional least squares FWI and ROWI on the data by using four different poor initial models to evaluate the robustness of ROWI against the poor initial model. The ROWI result is compared to several 2D GPR profiles for validation. We estimate the Pareto solutions of multi-objective subproblems and analyze their distribution in the model space for trade-off information.

2. Methodology

2.1. Multi-Objective Framework

FWI tries to find an optimal model (or a set of optimal models) \mathbf{m} that explains the observed data \mathbf{d}^{obs} under certain criteria

$$\min \Phi(\mathbf{m}) = \mathcal{L}[\mathbf{d}^{obs}, \mathbf{d}^{syn}(\mathbf{m})] \quad \text{s.t.} \quad \mathbf{d}^{syn}(\mathbf{m}) = P\mathcal{F}(\mathbf{m}) + \epsilon, \quad (1)$$

where \mathcal{L} represents a measurement that quantifies the misfit between the synthetic data \mathbf{d}^{syn} and observed data \mathbf{d}^{obs} ; \mathcal{F} and ϵ represent the forward solver and the errors in it, respectively; P represents the sampling of seismic wavefield at receivers' locations. Although the error ϵ always exists (i.e., numerical artifacts, imperfect theory, noise), it is usually assumed to be trivial and is ignored in FWI. In this study, we use the finite-difference modeling of 3D viscoelastic wave equation (Bohlen, 2002) as the forward solver $\mathcal{F}(\mathbf{m})$.

Possible reasons for the error ϵ in our forward solver include the undesired frequency dependency of the quality factor Q , imperfect source focal mechanism, and ignoring environmental noise.

Because the observed data typically contains multiple independent shots and we have multiple choices of measurements for misfit quantification, the FWI objective function can be discretized and written as

$$\Phi(\mathbf{m}) = \begin{bmatrix} \mathcal{L}_1(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}), & \dots, & \mathcal{L}_1(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}) \\ \vdots, & \ddots, & \vdots \\ \mathcal{L}_{N_L}(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}), & \dots, & \mathcal{L}_{N_L}(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}) \end{bmatrix}, \quad (2)$$

where the subscripts of \mathcal{L} and \mathbf{d} represent the indices of the measurements and shots, respectively, for example, $\mathcal{L}_k(\mathbf{d}_j^{obs}, \mathbf{d}_j^{syn})$ represents the misfit of the j -th shot with the k -th measurement; N_L and N_s are the numbers of measurements and shots, respectively, which define the size of the multi-objective function. Therefore, FWI is formulated as a multi-objective optimization problem.

There are different choices for the measurement \mathcal{L} in FWI. How to choose measurements is target- and data-oriented. Since we are dealing with shallow-seismic data which is dominated by surface waves, we choose three different measurements including the conventional least squares misfit that has an 'optimal' resolution (Pladys et al., 2017); the envelope misfit (Borisov et al., 2017; Yuan et al., 2015) that is sensitive to the group velocity and amplitude of surface waves; and the dispersion-spectra misfit (Masoni et al., 2013; Pérez Solano et al., 2014; Zhang et al., 2020) that is sensitive to the dispersion characteristics of surface waves

$$\mathcal{L}_1(\mathbf{d}^{obs}, \mathbf{d}^{syn}) = \frac{1}{2} \|\mathbf{d}^{syn} - \mathbf{d}^{obs}\|^2, \quad (3)$$

$$\mathcal{L}_2(\mathbf{d}^{obs}, \mathbf{d}^{syn}) = \frac{1}{2} \|\sqrt{(\mathbf{d}^{syn})^2 + \mathcal{H}^2(\mathbf{d}^{syn})} - \sqrt{(\mathbf{d}^{obs})^2 + \mathcal{H}^2(\mathbf{d}^{obs})}\|^2, \quad (4)$$

$$\mathcal{L}_3(\mathbf{d}^{obs}, \mathbf{d}^{syn}) = \frac{1}{2} \|\|\mathcal{F}_{3D}(\hat{\mathbf{d}}^{syn})\| - \|\mathcal{F}_{3D}(\hat{\mathbf{d}}^{obs})\|\|^2, \quad (5)$$

where \mathcal{H} and \mathcal{F}_{3D} represent the Hilber transform and 3D Fourier transform, respectively; $\|\cdot\|$ represents the absolute value in the frequency-wavenumber (f - k) domain; \mathbf{d} represents the data in the acquisition coordinates (typically the Cartesian coordinates that consist of the in-line horizontal, cross-line horizontal, and vertical coordinates); and $\hat{\mathbf{d}}$ represents the data in the ray-based coordinates (i.e., radial horizontal, transverse horizontal, and vertical directions). We use the ray-based coordinates in \mathcal{L}_3 so that Rayleigh and Love waves are separated in the f - k spectra of different horizontal components.

One common way to solve the multi-objective optimization problem is to scalarize the objective function (Yuan et al., 2020). Another way is to solve one of the objective functions progressively while using the other misfits as constraint (Pan et al., 2020). It is also possible to solve the multi-objective optimization problem by using an evolutionary optimization algorithm when the model dimension can be significantly reduced to a small number so that the computational cost is affordable. However, the mentioned approaches usually result in a relatively high computational cost that is more expensive than running a conventional FWI.

2.2. Stochastic Gradient Descent Optimization

To improve computational efficiency, we adopt a stochastic gradient descent algorithm to optimize the multi-objective function. At every iteration, we randomly choose and treat one of the multi-objective misfits only, for example, the j -th shot with the k -th measurement is chosen at the i -th iteration. We calculate the gradient of the chosen single-objective function that

$$\delta \mathbf{m}_i = \nabla \mathcal{L}_k(\mathbf{d}_j^{obs}, \mathbf{d}_j^{syn}(\mathbf{m}_i)), \quad (6)$$

and update the model via a preconditioned steepest descent algorithm

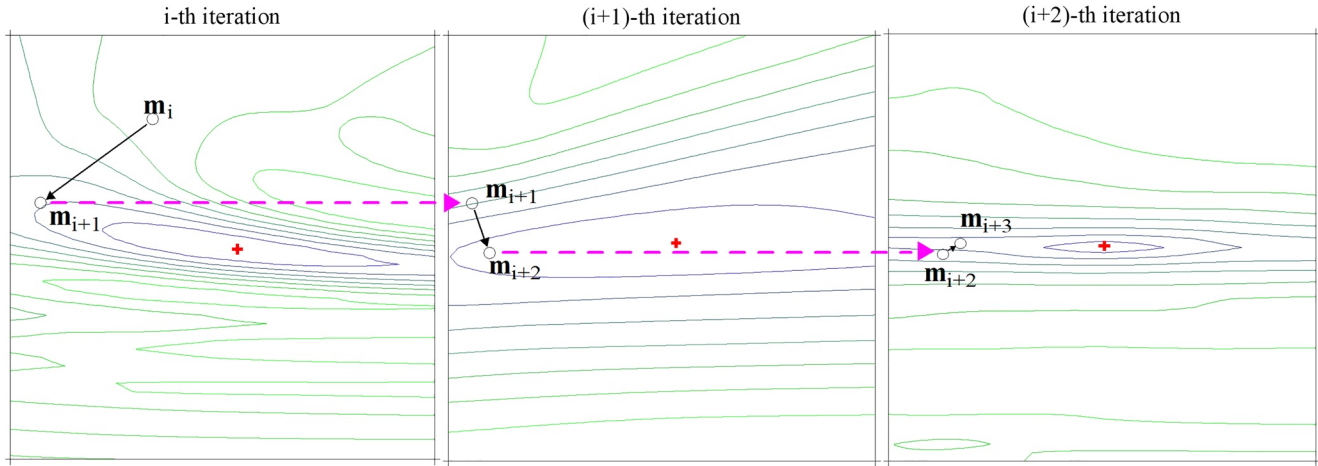


Figure 1. Schematic figure illustrating the solution path of random-objective waveform inversion. Colorful lines in the three images represent the contours of the objective functions assigned at the i -, $(i+1)$ -, and $(i+2)$ -th iteration, respectively. Circles represent the models on the solution path, which are lined up by arrows. The red cross represents the location of the global minimum in each assigned objective function.

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \alpha \mathbf{P}(\mathbf{m}_i) \delta \mathbf{m}_i, \quad (7)$$

where \mathbf{m}_i and \mathbf{m}_{i+1} are the models at the i -th and $(i+1)$ -th iteration, respectively; \mathbf{P} is a preconditioner using the inverse of an approximated diagonal of the Hessian with respect to the least-squares misfit function \mathcal{L}_i (Shin et al., 2001); α is a step length estimated by a line search algorithm with parabolic fitting (Nocedal & Wright, 2006). At every iteration, the model is updated using the single-objective function that is currently assigned to. This early stopping in the single-objective optimization problem (one iteration) helps to regularize the stochastic gradient descent algorithm. The model is updated iteratively by using randomly assigned single-objective functions until the stopping criterion (maximal iteration in this study) is satisfied. This strategy of choosing a single shot-gather and performing one iteration requires a high quality of the data and high redundancy in the data. We may need to choose a mini-batch (van Herwaarden et al., 2020; Yang et al., 2018) of objective functions per iteration if the data has relatively poor quality and low redundancy or when the use of single-shot data becomes inefficient in reducing the misfit. Mini-batches with growing sizes might further improve the quality of the final model (van Leeuwen & Herrmann, 2013).

Figure 1 shows schematically how ROWI proceeds iteratively. We randomly choose one shot with one measurement at every iteration and therefore, the objective function also changes iteratively (contours in Figure 1). We do not set any constraint on the random selection of objective function so that each objective function has a chance of $\frac{1}{(N_s \times N_L)^x}$ to be used x times in a row. ROWI probes and jumps between different objective functions and will converge to a common minimal region that is shared by all objective functions. Thus, similar to conventional FWI, ROWI might also become trapped in a local minimum region if it exists in all objective functions simultaneously.

In every iteration, we only need to solve a single-shot single-objective optimization problem to update the model, which is computationally inexpensive. The stochastic nature and the relatively low ill-posedness in the envelope misfit help ROWI avoid being trapped at local minimum regions. A multi-scale strategy (Bunks et al., 1995) can be combined with ROWI to further avoid cycle skipping (Virieux & Operto, 2009).

2.3. Pareto Solution and Trade-Off Information

In the ideal case, all single-objective functions share the same global minimum point in the model space that represents the optimal solution. However, due to the non-uniqueness of the inverse problem and the existence of errors in the data and the forward solver, the global minima may not locate identically among all objective functions (e.g., red crosses in Figure 1) in a realistic scenario. In this case, we are not able to find one optimal solution but a set of Pareto optimal solutions that are not dominated by any other solution.

Pareto solution: A solution \mathbf{m}_i dominates a solution \mathbf{m}_j if $\phi_k(\mathbf{m}_i) \leq \phi_k(\mathbf{m}_j)$ for all $k = 1, 2, \dots, n$ (n is the total number of objective functions) and with strict inequality for at least one k , where ϕ_k represents the k -th misfit value in the multi-objective function Φ . A Pareto solution is a solution that is not dominated by any other solution. In other words, the multi-objective misfits of a Pareto solution cannot be reduced simultaneously. The set of Pareto solutions is called the Pareto front.

All non-Pareto solutions are dominated by at least one Pareto solution and are, therefore, falsified and dropped (Tarantola, 2006). The Pareto solutions are the admissible solutions of the multi-objective inverse problem and their distribution in the model space represents the trade-off information.

The complete characterization of Pareto solutions $\mathbf{m}_{best} = \mathbf{argmin}(\Phi)$ is computationally intractable, mainly due to (a) the complete evaluation of Pareto optimality is prohibitively expensive (we need to simulate all shots for all candidate solutions); (b) the complete sampling of Pareto solutions in the model space is difficult; (c) the computational cost of solving the forward problem is high. One practical way to characterize the Pareto front is to use the final (last-iteration) solutions of ROWI with different solution paths or initial models as the samples of the Pareto front. It assumes that ROWI has successfully converged to a Pareto solution at the last iteration and, therefore, the number of samples on the Pareto front equals the number of independent ROWI runs.

Here, we propose an alternative way for the estimation of Pareto solutions. We decompose the multi-objective function into N_s shot-related subproblems in which the objective functions corresponding to the same shot are grouped as one subproblem

$$\Phi(\mathbf{m}) = \left\{ \underbrace{\begin{bmatrix} \mathcal{L}_1(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}(\mathbf{m})) \\ \mathcal{L}_2(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}(\mathbf{m})) \\ \mathcal{L}_3(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}(\mathbf{m})) \end{bmatrix}}_{\text{subproblem } \Phi_1}, \dots, \underbrace{\begin{bmatrix} \mathcal{L}_1(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}(\mathbf{m})) \\ \mathcal{L}_2(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}(\mathbf{m})) \\ \mathcal{L}_3(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}(\mathbf{m})) \end{bmatrix}}_{\text{subproblem } \Phi_{N_s}} \right\}. \quad (8)$$

Because every candidate model (models on the solution paths) has been assigned to one shot during the optimization, we only evaluate the Pareto optimality of the candidate model in the subproblem that corresponds to the shot it has been assigned to

$$\Phi(\mathbf{m}) = \left\{ \underbrace{\begin{bmatrix} \mathcal{L}_1(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}(\mathbf{m}^{\rightarrow 1})) \\ \mathcal{L}_2(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}(\mathbf{m}^{\rightarrow 1})) \\ \mathcal{L}_3(\mathbf{d}_1^{obs}, \mathbf{d}_1^{syn}(\mathbf{m}^{\rightarrow 1})) \end{bmatrix}}_{\text{subproblem } \Phi_1}, \dots, \underbrace{\begin{bmatrix} \mathcal{L}_1(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}(\mathbf{m}^{\rightarrow N_s})) \\ \mathcal{L}_2(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}(\mathbf{m}^{\rightarrow N_s})) \\ \mathcal{L}_3(\mathbf{d}_{N_s}^{obs}, \mathbf{d}_{N_s}^{syn}(\mathbf{m}^{\rightarrow N_s})) \end{bmatrix}}_{\text{subproblem } \Phi_{N_s}} \right\}. \quad (9)$$

where $\mathbf{m}^{\rightarrow j}$ represents the models that have been assigned to the j -th shot during the inversion. Therefore, the models which have been assigned to the same shot are now grouped into a common subproblem (Equation 9) and the Pareto optimality in a subproblem is independent of the other shots. Since the synthetic data $\mathbf{d}_j^{syn}(\mathbf{m}^{\rightarrow j})$ has already been simulated during the optimization, we just need to compare its single-shot misfits \mathcal{L}_1 , \mathcal{L}_2 , and \mathcal{L}_3 to the other candidate models in the same subproblem, and falsify the models which are dominated by at least one other model in a subproblem. The cost of evaluating Pareto optimality in the subproblems is negligible compared with the numerous forward simulations required in the optimization.

We estimate the Pareto solutions of all subproblems $\tilde{\mathbf{m}}_{best}$ to approximate the Pareto solutions \mathbf{m}_{best} and analyze their distribution in the model space for trade-off analysis

$$U(\tilde{\mathbf{m}}_{best}) = \frac{std(\tilde{\mathbf{m}}_{best})}{mean(\tilde{\mathbf{m}}_{best})}. \quad (10)$$

where *std* and *mean* represent the standard deviation and the mean value of the models, respectively. Here we use the relative standard deviation (RSD) U to represent trade-off information with respect to the

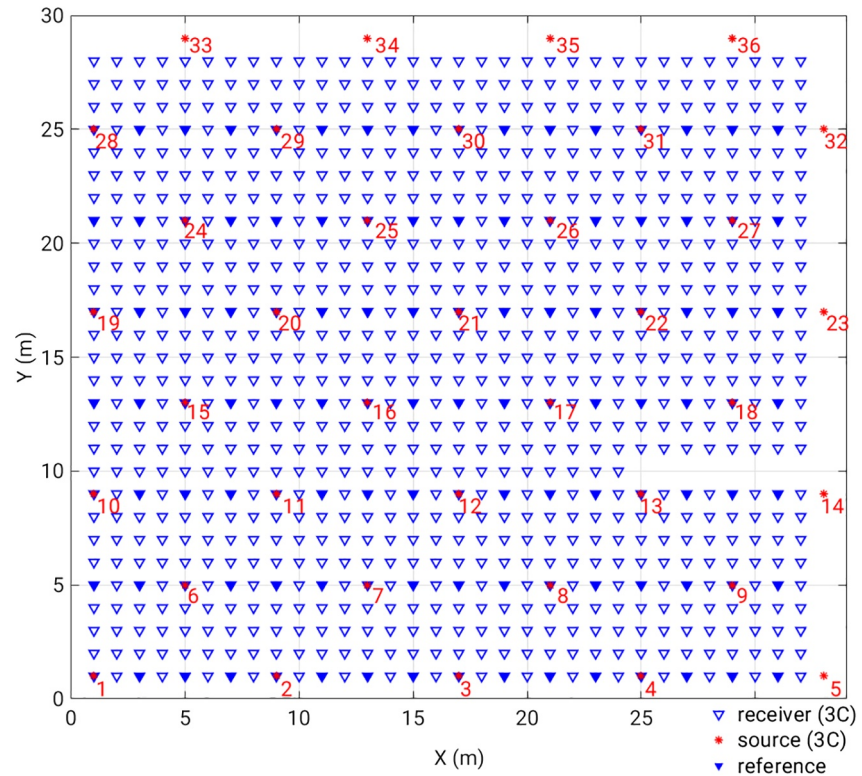


Figure 2. Acquisition system in the field example. Triangles and asterisks represent the locations of the 3C receiver and 3C source, respectively. The filled triangles represent the receivers' location for coarse-grid reference data acquired on the same day. The dense-grid data (triangles) is acquired during five days.

reference solution (averaged model). Low RSD indicates that different objective functions agree on what the optimal model is, while in parts where the RSD is high, using different objective functions may lead to different models. Similar analysis can also be applied to a preferred Pareto solution (Pan et al., 2020).

We use the Pareto solutions of subproblems $\tilde{\mathbf{m}}_{best}$ to approximate the Pareto solutions of the full problem \mathbf{m}_{best} by assuming that the Pareto fronts of all subproblems constitute the Pareto front of the full multi-objective problem (Equation 2). Because the candidate solutions are estimated by optimizing the full multi-objective function and are expected to be close to or located at the Pareto front, the error in the approximation of \mathbf{m}_{best} by $\tilde{\mathbf{m}}_{best}$ is expected to be acceptable.

The estimated trade-off information is mainly associated with the inherent inconsistency among different subsets of data and objective functions, the non-uniqueness of the inverse problem, and the influence of errors in the forward solver (ϵ in Equation 1). We perform several runs of ROWI with different solution paths to improve the diversity in the samples of the Pareto front. Overall, the decomposition of the original multi-objective function into shot-related subproblems provides a crude but computationally cheap (almost free) way for trade-off analysis.

3. Data Acquisition and Pre-Processing

A three-component (3C) \times 3C field data set is acquired in Rheinstetten, Germany. The main geological target is a refilled 'V'-shaped trench, namely 'Ettlingen Line' (EL), which locates approximately along the main diagonal of the survey area. A total number of 888 3C receivers' locations are placed on a 1 m \times 1 m grid (Figure 2). A total number of 36 source locations are placed on a 4 m \times 4 m staggered grid and fully cover the survey area. We use a 3C Galperin source (Häusler et al., 2018) to generate three orthogonal single-force sources, namely U , V , and W sources, on each source location, respectively. The U source is generated by a single force along X direction with a dipping angle of $\alpha = 35.3^\circ$, and the V and W sources are generated by

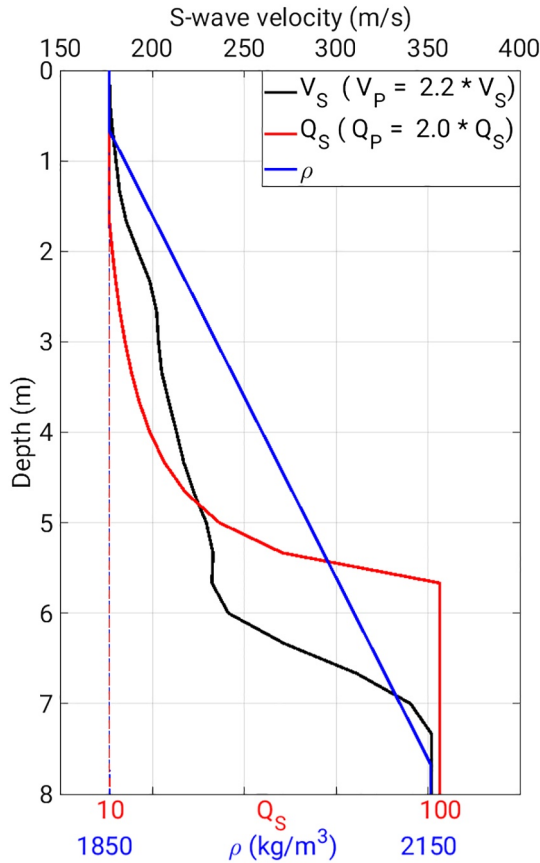


Figure 3. 1D initial model for the field example. The thin line highlights the Q_S and ρ values near the free surface.

rotating the U source with $2 \times \beta = 240^\circ$ and $\beta = 120^\circ$ in the horizontal (X - Y) plane (Figures 2 and 6 in Häusler et al., 2018), respectively. The Galperin sources can be projected on the Cartesian coordinates via

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ \cos \alpha \cos 2\beta & \cos \alpha \sin 2\beta & \sin \alpha \\ \cos \alpha \cos \beta & \cos \alpha \sin \beta & \sin \alpha \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \quad (11)$$

Because of the lack of sufficient 3C geophones and seismographs, we can not deploy all receivers' locations on the $1 \text{ m} \times 1 \text{ m}$ grid simultaneously. Thus, a reference data set on a coarse grid ($2 \text{ m} \times 4 \text{ m}$; filled triangle in Figure 2) with all source locations is acquired on the same day. The dense-grid ($1 \text{ m} \times 1 \text{ m}$) data is divided into six subsets, and each subset containing part of the receivers' locations and all sources is acquired on the same day. A detailed description of the data acquisition can be found in Schaneg (2017) and Irnaka et al. (2019).

A 4D effect, i.e., undesired phase and amplitude inconsistency between the data acquired on different days, is observed if we directly combine the data of all subsets. Possible reasons for the 4D effect include the repeatability of the source signal, different source's and receiver's coupling effect, different soil condition (saturation of the soil) and experimental environment (e.g., temperature and environmental noise). To overcome this 4D effect, a matching filter is estimated by minimizing the waveform difference between the dense-grid data and reference data. The matching filter is then applied to the dense-grid data to remove the 4D effects. The matching filter strategy nicely compensates for the 4D effects in the dense-grid data, and a detailed description is shown in Irnaka et al. (2019). We have a total of 108 shots ($3C \times 36$ source points) with 2,664 traces in each shot ($3C \times 888$ receivers). Low-quality traces are removed before the inversion.

We build a 1D initial V_S model (black line in Figure 3) by inverting Rayleigh-wave dispersion curve. The initial V_P is set as 2.2 times of V_S . The 1D Q_S and Q_P models are estimated by inverting Rayleigh-wave attenuation coefficient (Gao et al., 2018). A 1D initial density model is built by empiri-

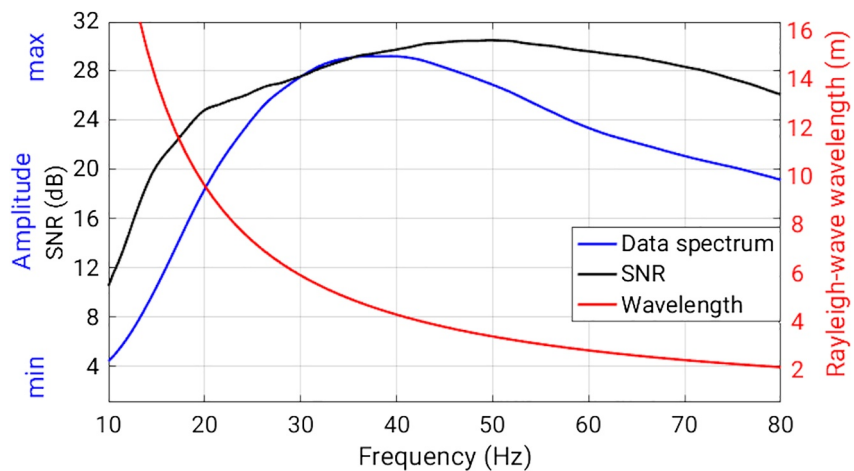


Figure 4. Spectrum, signal-to-noise ratio (SNR), and wavelength of the whole data set. The wavelength is calculated by using Rayleigh-wave phase velocities. The noise is calculated from the signal that arrives after 500 ms with a window length of 480 ms. Considering the quality (SNR) and resolution (half wavelength) of the data, we choose to progressively invert the data up to 25, 35, and 50 Hz, respectively.

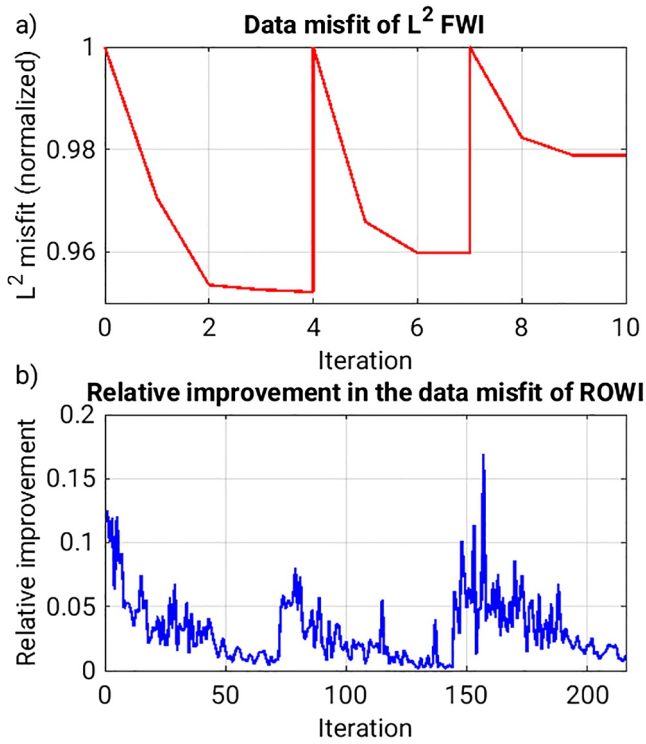


Figure 5. Evolution of data misfit in (a) conventional least squares full-waveform inversion and (b) random-objective waveform inversion (ROWI). The blue curve shows the relative improvement in the currently assigned single-objective function and represents the averaged performance of six ROWI tests.

cal estimation. We only update the elastic-parameter (V_s , V_p , and density) models in the inversion and use the attenuation models as passive information (Groos et al., 2017).

4. Conventional Least Squares FWI

We apply multiscale least squares FWI to the data. Considering the quality (signal to noise ratio) and resolution (half wavelength) of the data (Figure 4) as well as the availability of a good initial model (Figure 3), we choose to progressively invert the data up to 25, 35, and 50 Hz, respectively. All 108 shots are used simultaneously and the least squares misfits of all shots are summed as the objective function. A preconditioned steepest descent algorithm is used.

The least squares FWI converges after 10 iterations (Figure 5a) and its result is shown in Figure 6. Two low-velocity trenches are estimated, while the one going diagonally across the shallow part of the model corresponds to the EL. The EL is around 8 m wide and reaches 3 m in depth. The dimension of the second trench is smaller than the EL and was not discovered in the previous result estimated by the inversion of local Rayleigh-wave dispersion curves (Pan, Schaneng, Steinweg, & Bohlen, 2018). It is worth mentioning that we have also performed several FWI tests by using smaller frequency intervals and/or including stages with lower frequencies in the multiscale approach, whose final results are very similar to Figure 6. The inversion runs only a few iterations per stage (Figure 5a). One possible reason for the premature convergence of the conventional FWI is the relatively poor performance of the steepest descent algorithm, which sometimes generates small steps (e.g., Figure 5a in Métivier & Brossier, 2016), and as a result of which, prematurely triggers the stopping criterion (i.e., the inversion stops or moves to the next stage once the relative improvement in the misfit is less than 1%). Another possible reason

for the premature convergence is the unbalanced energy among different shots. In this example, we didn't perform shot normalization because the ratio between the L^2 -norm of the strongest and weakest shots is less than two. While this unbalanced energy among different shots is not a problem in ROWI because every single shot is used independently. The result of conventional FWI can be improved by adopting an appro-

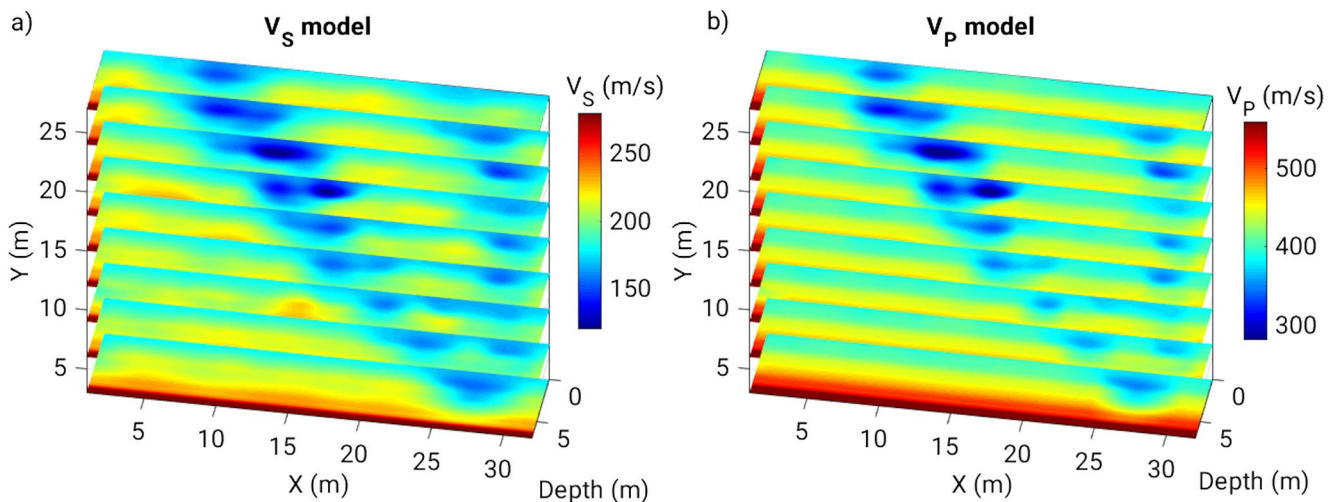


Figure 6. Conventional least squares full-waveform inversion result in the field example. (a and b) represent the final V_s and V_p models, respectively.

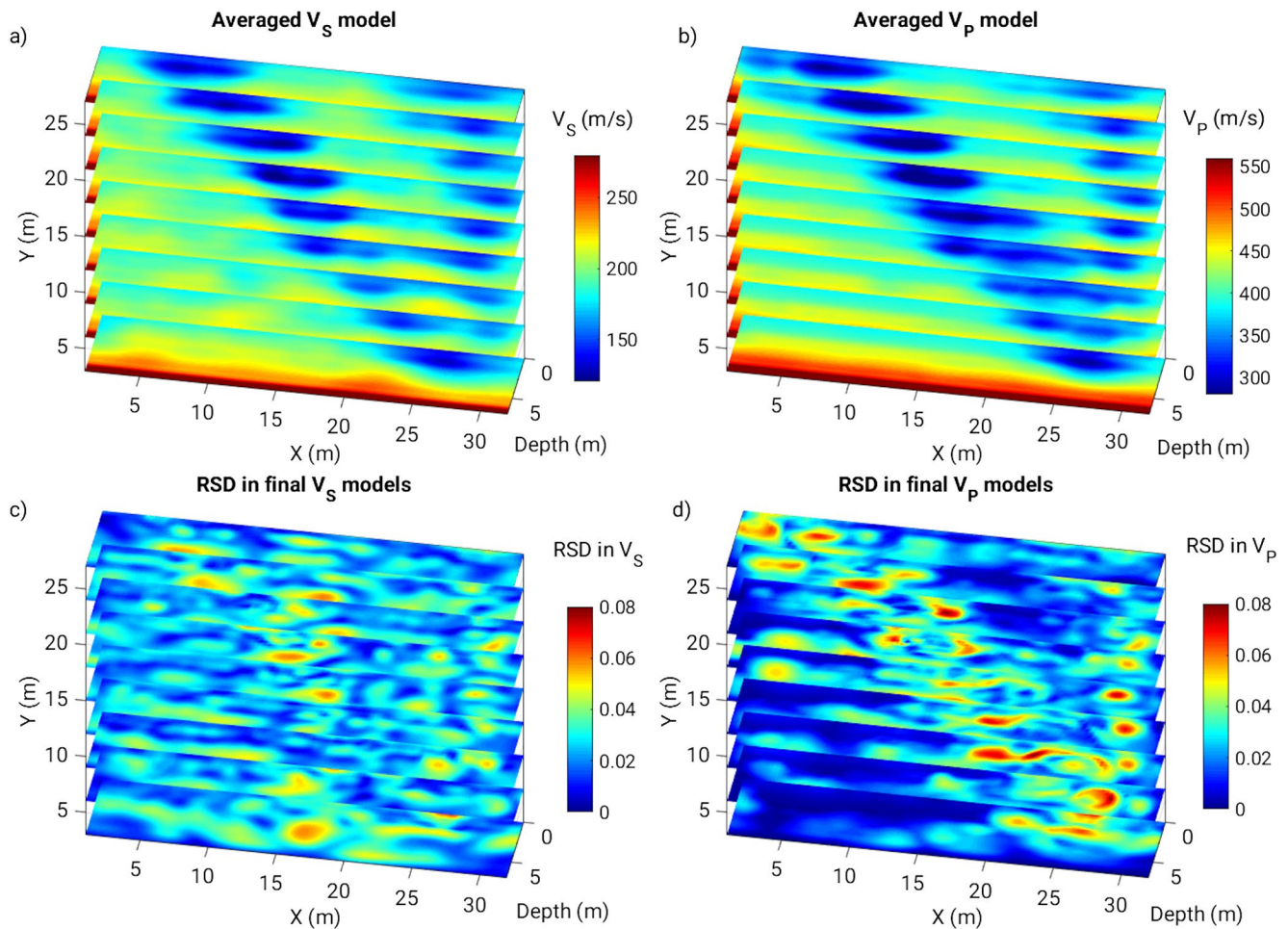


Figure 7. Random-objective waveform inversion (ROWI) result in the field example. The upper and lower rows represent the averaged inversion result of six ROWI tests and their relative standard deviation (RSD), respectively. The left and right columns represent the V_S and V_P models, respectively. The relatively low RSD among the six inversion results indicates the high robustness of ROWI against the random solution paths.

appropriate data normalization strategy (Louboutin et al., 2017) or using more advanced optimization strategies (e.g., Bessel gradient smoothing with l -BFGS method in Irnaka, 2021). It is worth mentioning that because the data is highly redundant, we can also improve the efficiency of the conventional FWI by only using part of the data (Irnaka, 2021).

5. ROWI Result and Its Robustness Against Random Solution Paths

We run six ROWI tests with different random solution paths on the field data. In each test, we use the same multiscale strategy and perform 72 iterations in each frequency stage. Thus, one ROWI test runs a total of 216 iterations, whose computational cost is approximately equivalent to running two iterations in a conventional FWI that uses all 108 shots. The objective value reduces rapidly in the first 30 iterations of each stage, and the speed of its reduction generally slows down with iteration (Figure 5b). On the averaged, the single-objective misfit is improved by 3% per iteration in the six ROWI tests.

Figure 7 shows the averaged inversion result estimated by the six ROWI tests and the RSD among the six final (last-iteration) models. Two low-velocity trenches are clearly identified in the reconstructed V_S and V_P models and are similar to the conventional FWI results (Figure 6). The two trenches are more continuous and more dominant in the ROWI result compared with the conventional FWI results. The resolution of the ROWI result might be further improved by performing conventional least squares FWI on the ROWI result.

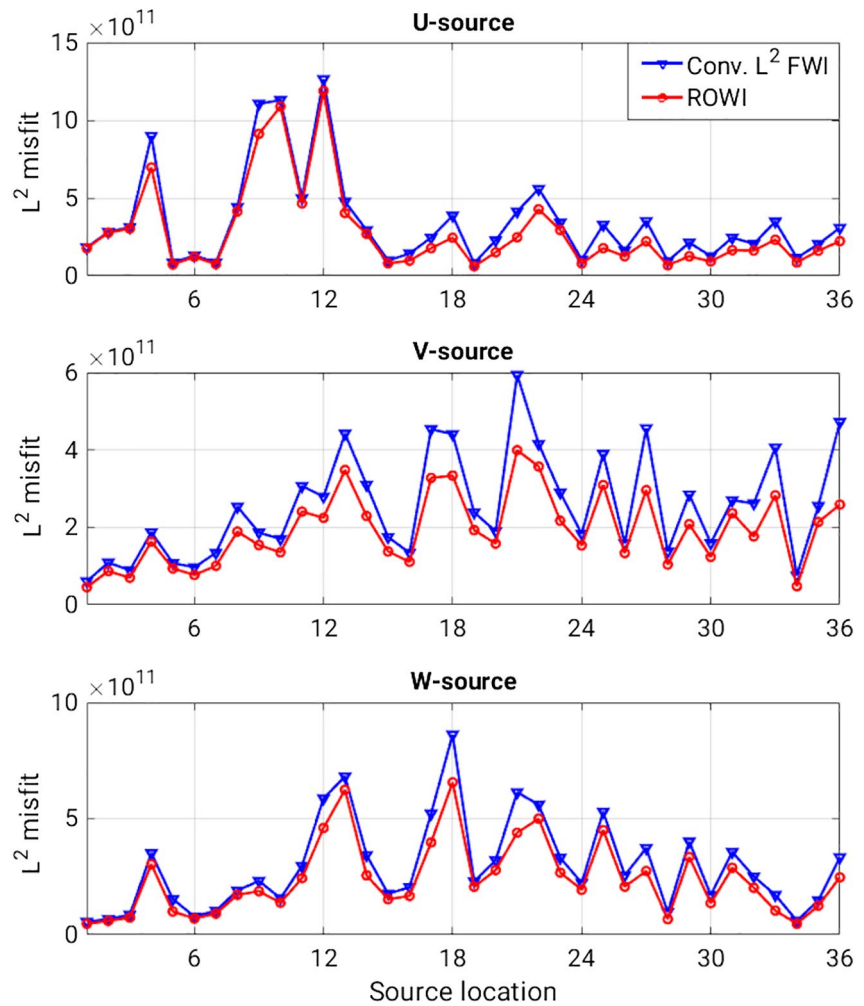


Figure 8. Comparison between the L^2 data misfits (up to 50 Hz) corresponding to the conventional least squares full-waveform inversion (FWI) result and the averaged random-objective waveform inversion (ROWI) result. Overall, the L^2 misfit of the ROWI result is on average 21% lower than that of the least squares FWI result.

The RSD values among the six ROWI solutions are small in most parts of the model (i.e., $RSD < 2\%$; the area in blue in Figures 7c and 7d). It indicates that the ROWI converges toward a similar result despite using different solution paths. The relatively high RSD values are mainly distributed sporadically at the boundaries of the trenches. The areas inside the EL do not have systematically high RSD values, indicating that the targeted trench-like structure is reconstructed consistently by the six ROWI tests with different solution paths.

6. Multi-Objective Data Fitting

We compare the least squares data misfits (up to 50 Hz) of all shots corresponding to the conventional FWI result (Figure 6) and the averaged ROWI result (Figure 7). We perform additional forward simulations on the averaged ROWI result (Figure 7) for the comparison of data misfit. Although the conventional FWI aims at minimizing the L^2 data misfit, its final misfit values of all shots are larger than the values corresponding to the ROWI result (Figure 8). The L^2 data misfit corresponding to the ROWI result is on average 21% lower than the conventional FWI result, showing that ROWI is more effective in reducing data misfit compared with the conventional FWI.

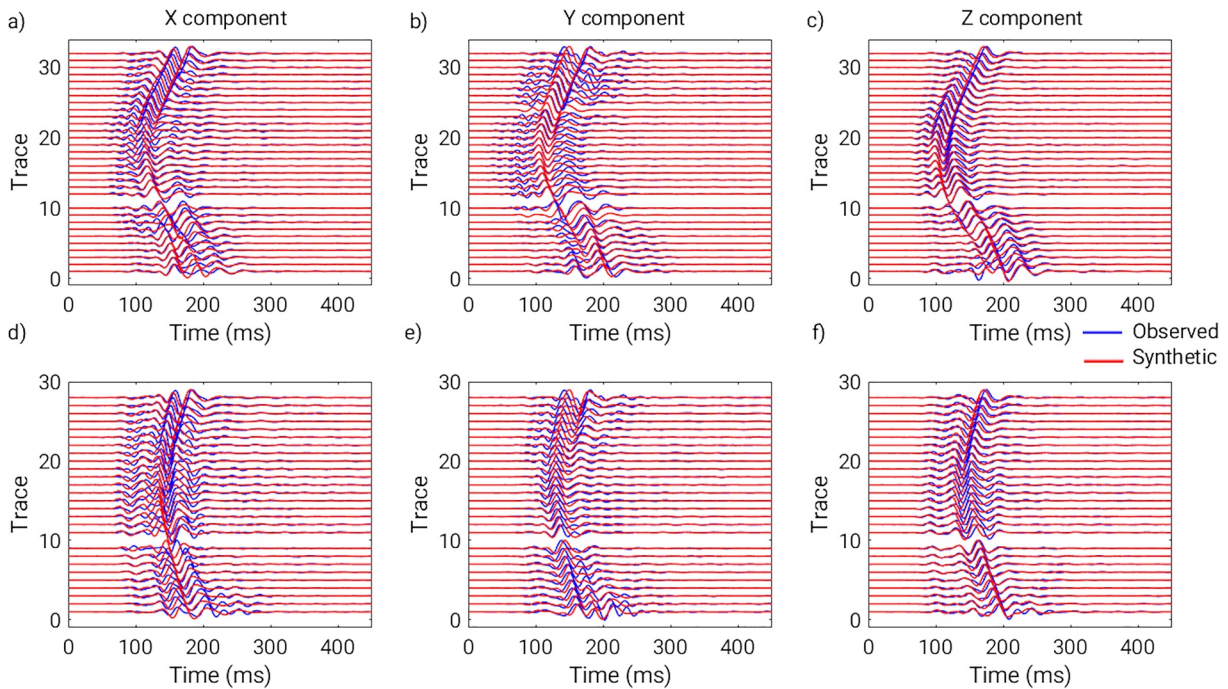


Figure 9. Comparison of the synthetic data corresponding to the averaged random-objective waveform inversion result and observed data generated by the V source at source location 21. The first and second rows represent the data acquired on the last line ($Y = 28$ m) and last column ($X = 32$ m) of receivers, respectively. Three columns represent the waveform acquired by the X-, Y-, and Z-component, respectively. The waveform is normalized in each trace for display.

We choose the data generated by V -source at source location 21 to show the waveform fitting of the ROWI result. This shot has the highest L^2 misfit among all shots generated by the V -source (Figure 8). The first comparison (Figure 9) is performed between the observed and synthetic waveforms acquired in the last row and last column of receivers (i.e., $Y = 28$ m and $X = 32$ m in Figure 2). Overall, the synthetic data fits the observed data fairly well. The waveform fitting in the vertical (Z-) component is generally better than in the horizontal components because the observed data has relatively higher amplitudes in the vertical component. We did not show the comparison between the waveform envelope whose nice data fitting can be inferred from the waveform comparison.

We perform another data comparison on the dispersion spectra (f - k spectra) of the same shot using all receivers. Figure 10 shows the wavenumber spectra of the observed (upper row) and modeled data (lower row) at a frequency of 35 Hz. The energy of the observed and synthetic data show a similar dispersion image in the wavenumber domain, which is mainly distributed along a circle. The circular energy represents the fundamental-mode Rayleigh (radial and vertical components) and Love waves (transverse component) at 35 Hz. Some weaker energy distributed along a smaller circle in the radial and vertical components belongs to high-mode Rayleigh waves. The azimuth distribution of energy in the Love-wave dispersion image is mainly caused by the focal mechanism of the V source, which can be decomposed as a vertical force and a horizontal force striking along 240° . Overall, the dispersion images of the observed data are nicely fitted, especially for the phase and the azimuth distribution of energy. The dispersion energy in the Rayleigh-wave components is stronger and better fitted compared with the Love-wave component.

7. Evaluation of ROWI Result

We evaluate the ROWI result by comparing its V_5 model to several 2D ground-penetrating radar (GPR) profiles acquired in the same survey area. Figures 11a and 11b show the comparison of 2D slices at $Y = 17$ m and $X = 21$ m, respectively, in which the strong contrast in V_5 nicely agrees with the boundaries delineated

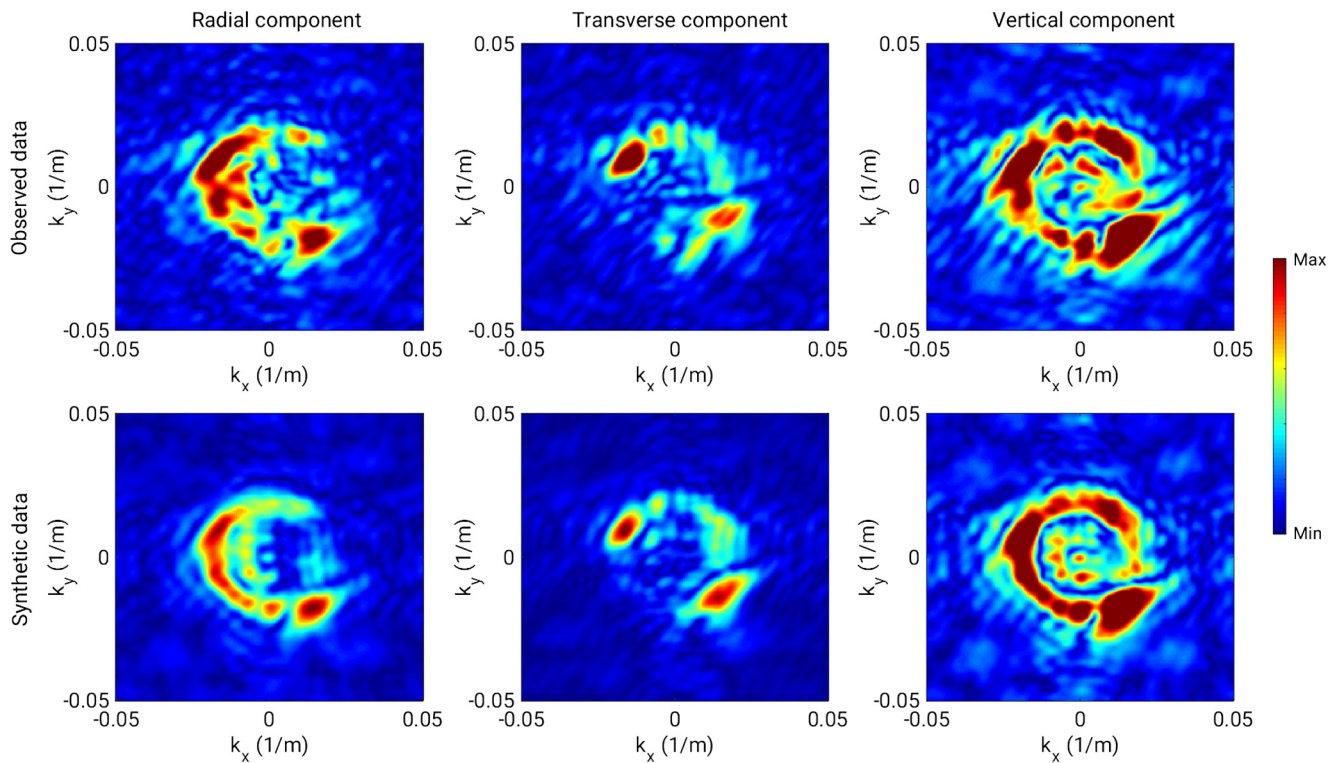


Figure 10. Comparison between the dispersion (f - k) spectra of the observed (first row) and synthetic (second row) data at a frequency of 35 Hz. Outer and inner circles of energy in the radial and vertical components represent the fundamental- and high-mode Rayleigh waves, respectively. The energy in the transverse component represents Love wave. The Rayleigh-wave energy is stronger than Love-wave in the data.

by the GPR profiles (wiggle plot; highlighted by the arrows). Figure 11c shows a 2D slice of V_S model at a depth of 0.8 m which is compared to the boundaries of the EL manually picked from multiple 2D GPR profiles (red lines). The location of the EL nicely matches its location estimated from GPR profiles in the 2D

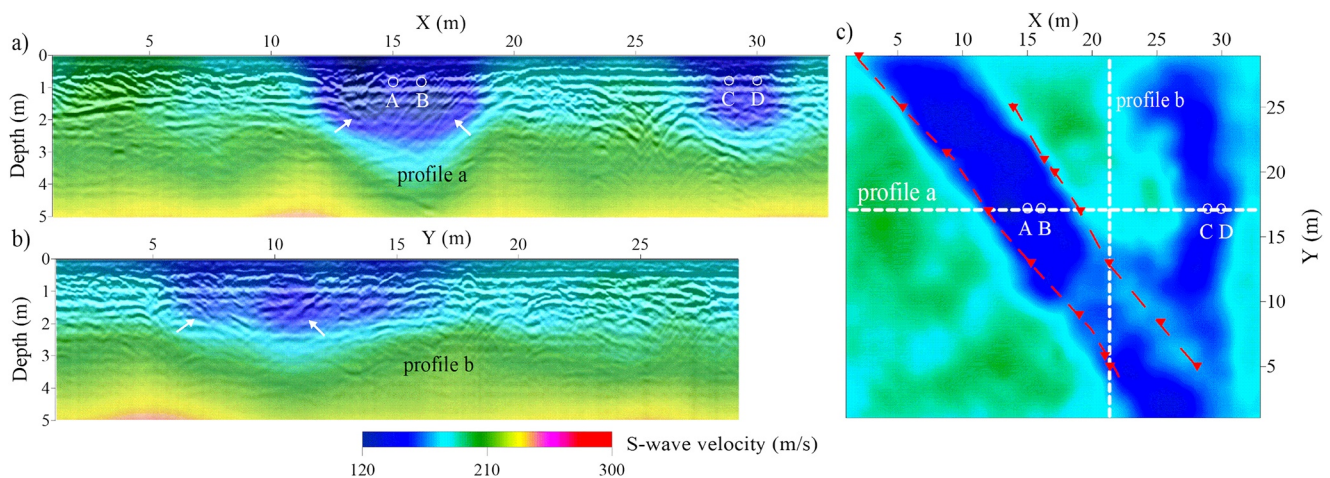


Figure 11. Comparison between the reconstructed V_S model and ground-penetrating radar (GPR) profiles. (a and b) represent the comparison in the 2D slices located at $Y = 17$ m and $X = 21$ m, respectively. (c) shows the comparison at a depth of 0.8 m. The triangles represent the boundary points of the EL picked from multiple 2D GPR profiles, which are lined up by two red lines. A, B, C, and D are the four chosen points for trade-off analysis. Arrows highlight the boundaries of the targeted trench, which are picked based on the prior information about the location and shape of the trench (Gao et al., 2020; Pan, Schaneng, Steinweg, & Bohlen, 2018).

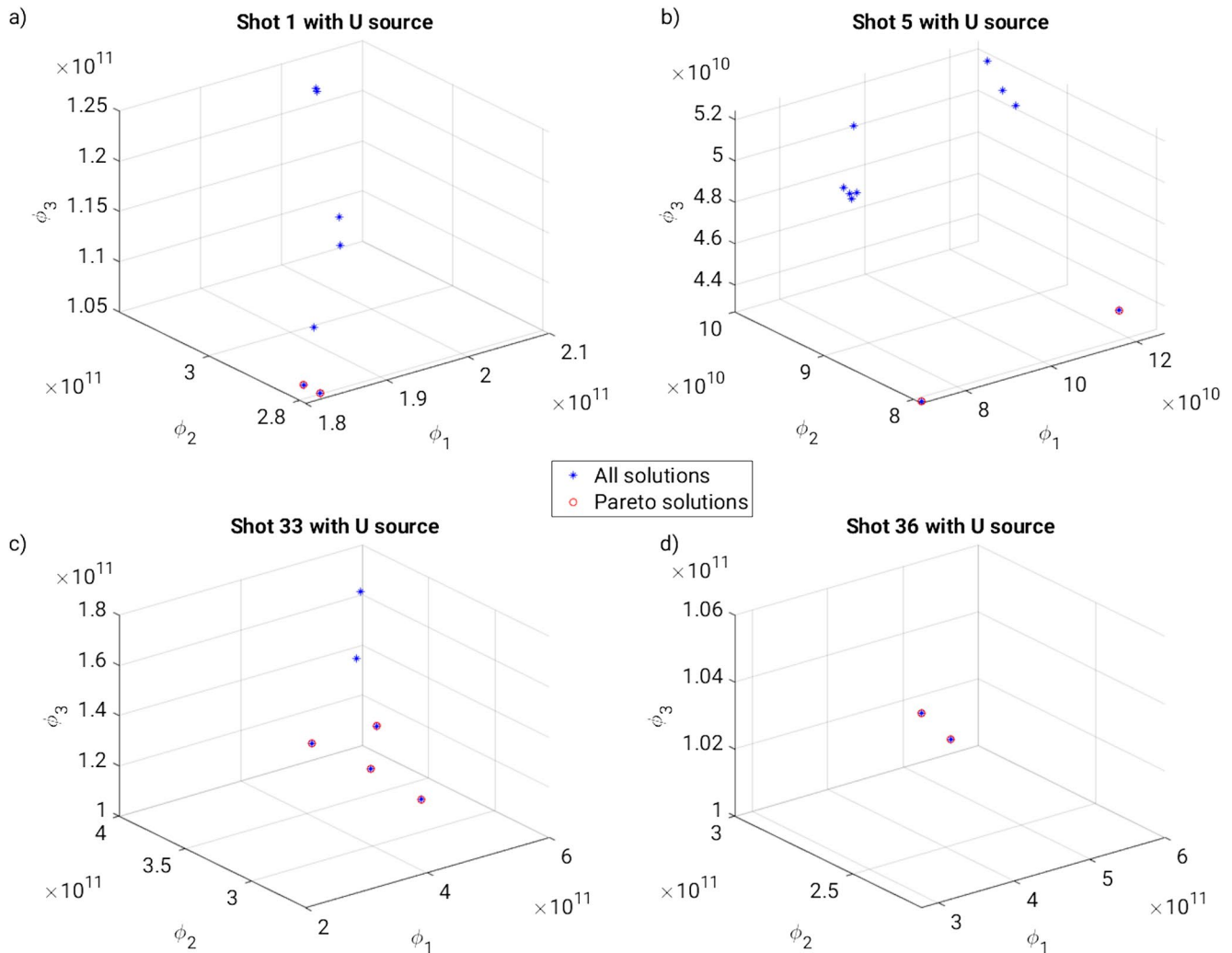


Figure 12. Distribution of models in the objective space of the subproblems corresponding to U -source shot 1, 5, 33, and 36. Those four sources are located at the corners of the survey area. Asterisks and circles represent the misfit values of all available samples and Pareto solutions in the subproblems, respectively.

depth slice. Overall, the good match between the GPR profiles and our model, especially for the location and shape of the EL, proves the validity of the model estimated by ROWI.

8. Pareto Solutions and Trade-Off Analysis

We evaluate all solutions along the six search paths on the final stage (72 iterations \times 6 solution paths = 432 models) and sort them into shot-related subproblems (Equation 9). Each shot has an average chance to be used four times, in other words, we have an average of four samples in each shot-related subproblem. In each subproblem, we compare the three misfit values of every sample to determine the Pareto solutions (see Section 2.3). Figure 12 shows the distributions and selections of Pareto solutions in the subproblems related to shot 1, 5, 33, and 36 generated by U source, respectively. In the subproblem which only contains one sampled solution, this solution automatically becomes a Pareto solution. Overall, a total of 246 local Pareto solutions are chosen from the 432 models (red and blue lines in Figure 13). Six ROWI tests contribute to a similar number (i.e., 16%, 19%, 18%, 15%, 15%, and 17%) of samples to the estimated Pareto solutions.

Figure 14 shows the mean and RSD of all 246 Pareto optimal solutions in the model space. Overall, the trade-off (RSD) is low throughout the whole model, while two trenches, especially the one locates near the boundary ($X = 31$ m) of the survey area, have relatively higher trade-offs compared to the surrounding

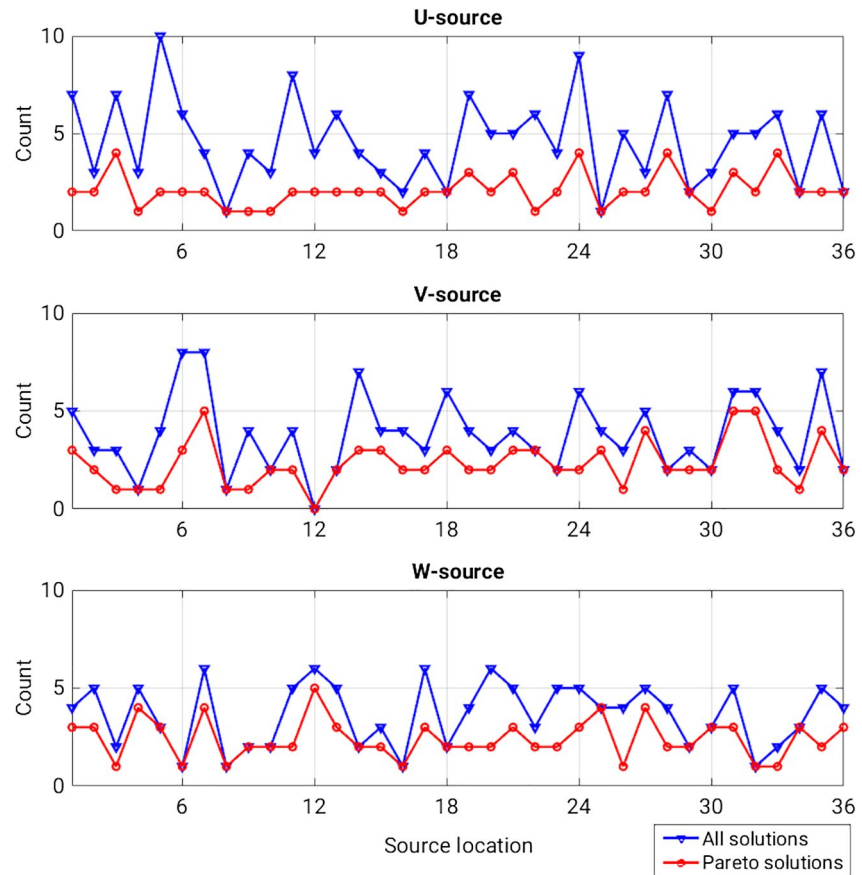


Figure 13. Numbers of samples (blue) and Pareto solutions (red) in each subproblem.

areas. It implies that the second trench is less reliable compared to the EL. Both the averaged model and trade-off information estimated by the Pareto solutions show a relatively high similarity to the results estimated by the six final (last-iteration) solutions of ROWI with different solution paths (Figure 15 compared to Figure 7). It indicates that the final ROWI results estimated by different solution paths nicely sample the Pareto front for trade-off analysis (Pan & Gao, 2020; Shigapov, 2019).

Besides analyzing the overall trade-off throughout the whole model, ROWI also allows analyzing the 1D or 2D joint distributions at some points or between pairs of points. Figure 15 shows the distribution of Pareto optimal solutions of V_S at four specific points located inside two trenches in the 2D slice $Y = 17$ m and at a depth of 0.8 m (A, B, C, and D in Figure 11). The point pairs A-B and C-D are located inside the EL and the second trench, respectively, with an inner distance of 1 m.

The 1D distributions of samples at all four points are non-Gaussian and the distributions are more centralized at points A and B compared to points C and D (marginals in Figure 15). The velocities at points A and B are relatively lower than the velocities at points C and D. The high probability values mainly distribute along the diagonals of the 2D distributions (Figure 15). It indicates that the reconstructed model is fairly homogeneous inside the two trenches. The samples are distributed fairly centralized in both the pairs A-B and C-D, indicating that both two trenches contain relatively low trade-offs at those points.

9. Robustness of ROWI Against Poor Initial Models

We rerun the field example starting from several poorer initial models to evaluate the robustness of ROWI against the poor initial models. We replace the 1D V_S structure above the bedrock with a constant value of 150, 160, 170, and 180 m/s (V_S above 7 m in Figure 3), respectively, to generate four different initial models. The initial V_P is set as 2.2 times of the modified V_S . The same 1D attenuation model, density model, and

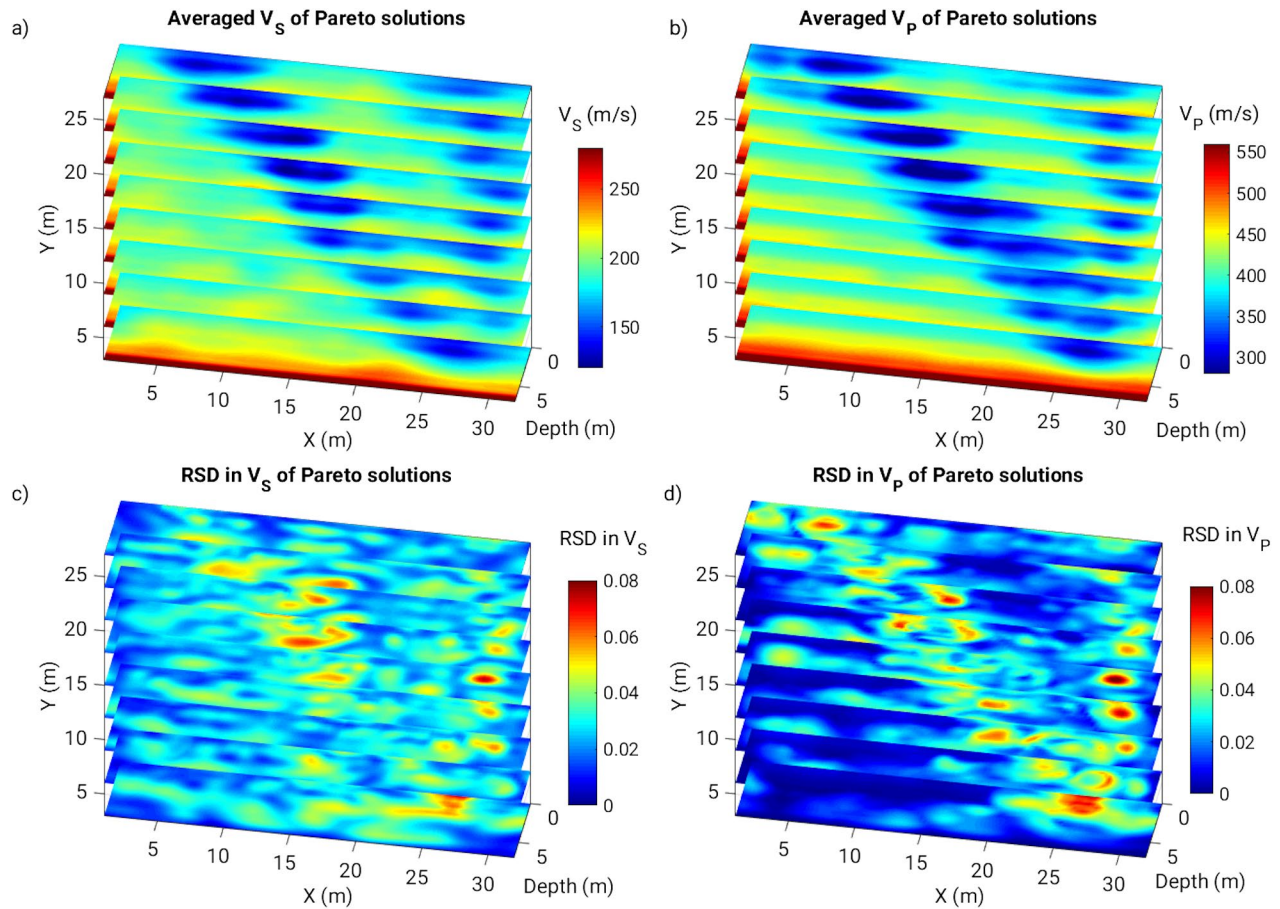


Figure 14. Mean and relative standard deviation of the 246 Pareto solutions. Overall, they show a high similarity to Figure 7.

inversion setup are used. We only perform one ROWI test starting from each of the four poor initial models. For comparison, we also perform conventional FWI with different objective functions on the data by using the same poor initial models.

Figure 16 shows the inversion results estimated by the conventional FWI with different objective functions and ROWI starting from the four poor initial models. Although we can roughly identify the existence and the locations of both two trenches in the conventional least squares FWI results (first row of images in Figure 16), the shapes and boundaries of the two trenches can not be reconstructed reliably. The four least squares FWI results differ significantly from each other, which highlights a high dependence of least squares FWI on the initial model. The results of envelope-based FWI also reconstruct the existence of the two trenches when using the four poor initial models (second row of images in Figure 16). Besides, the velocity in the deep part of the model is also updated. However, the results of envelope-based FWI suffer from relatively strong artifacts in the near-surface region. This is because the envelope misfit function mainly uses (fits) the amplitude information of the waveform, while the quality of amplitude information in our observed data is not ideal due to the 4D effect (amplitude inconsistency among traces). Similarly, because the f - k -spectra misfit is sensitive to the amplitude information of the data, the f - k -spectra based FWI underperforms in the reconstruction of the model compared to the least squares FWI, especially when a very poor initial model is used (third row of images in Figure 16). It is worth noting that the comparison between the performances of conventional FWI with different objective functions is not fair because the 4D effect in the data is corrected via a matching filter that minimizes the least squares waveform difference between the dense-grid data and reference data. The performances of the envelope-based and f - k -spectra based FWI can be improved when a more appropriate misfit function is used in the matching filter or by applying an appropriate trace-normalization to the data (Louboutin et al., 2017).

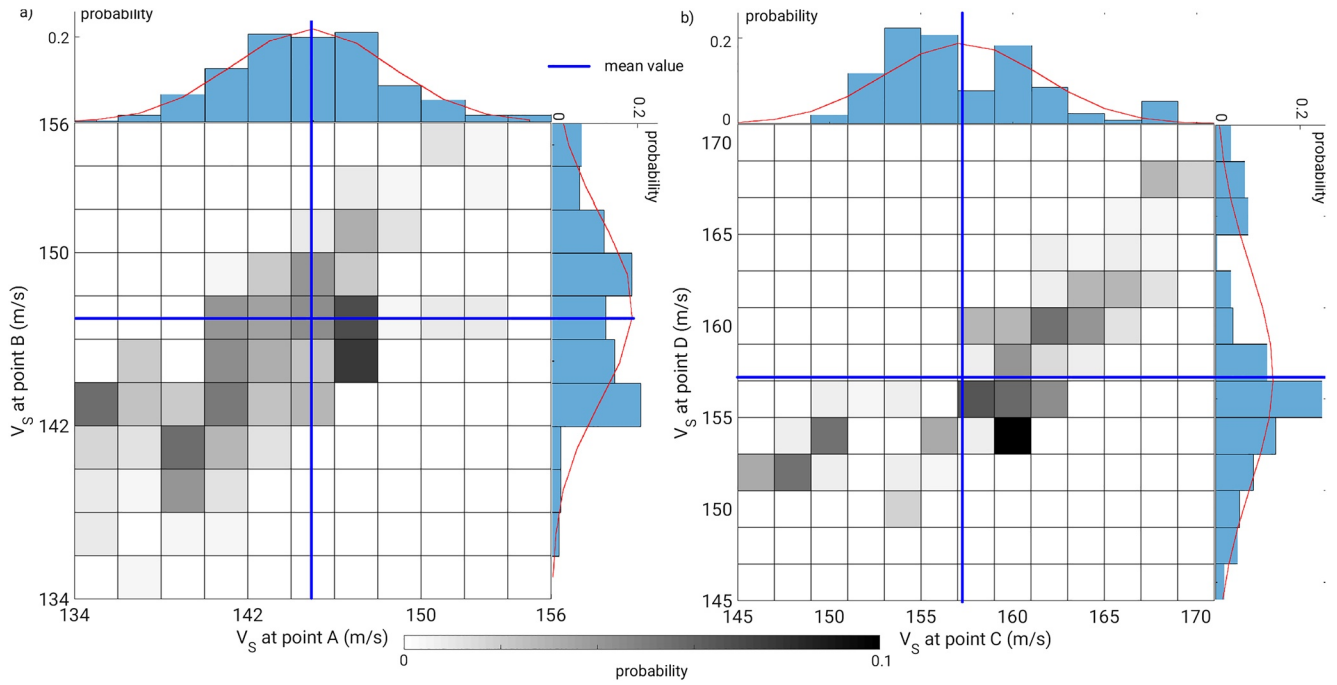


Figure 15. Trade-off information of the V_S values at four points located inside the two trenches. (a and b) show the 2D joint distributions of V_S at correlated points A-B and C-D, respectively. The marginals (blue bars on top and right side of the images) show the 1D distribution of V_S at every single point and the red curve represents the corresponding normal distribution fit. The blue lines represent the mean values.

The ROWI results are much more consistent regardless of the different poor initial models as well as the different solution paths (last row of images in Figure 16), especially for the reconstruction of the two trenches. The ROWI results better delineate the shapes and boundaries of the two trenches compared with the results estimated by conventional FWI, especially when a very poor initial is used (e.g., $V_S = 150$ m/s and $V_S = 160$ m/s in Figure 16). The difference among the four ROWI results is relatively small in most parts of the model, indicating that the V_S model is reconstructed consistently by ROWI even though different poor initial models are used. Relatively bigger differences mainly exist around the boundary of the model and below 5 m depth, which are caused by the low illumination of the wavefield in this part. The ROWI results also show a fairly high similarity to the result starting from a good initial model (Figure 7), especially for the part above 5 m depth due to a limited penetrating depth of surface wave. It indicates that both the inversion result and the trade-off analysis in the areas with low sensitivity (illumination) are not reliable. Overall, these examples prove the relatively higher robustness of ROWI against poor initial models compared with the conventional FWI approaches.

10. Discussion

10.1. Comparison Between Multi-Objective Trade-Off and Bayesian Uncertainty

We discuss the relationship between the trade-off information estimated by ROWI and the uncertainty based on the Bayesian approach (Tarantola, 2005). To do so, we scalarize our multi-objective function Φ (Equation 2) to a single-objective function ψ via

$$\psi(\mathbf{m}, \mathbf{w}) = \sum_{i=1}^{N_s} \sum_{j=1}^{N_L} w_{i,j} \mathcal{L}_j(\mathbf{d}_i^{obs}, \mathbf{d}_i^{sym}(\mathbf{m})), \quad (12)$$

where $\mathbf{w} = [w_{1,1}, w_{1,2}, \dots, w_{N_s, N_L}]$ represents arbitrary weights of positive values. Minimizing Equation 12 provides a sufficient condition for Pareto optimality in the multi-objective function Φ . Once an arbitrary weight $\hat{\mathbf{w}}$ is given, the posterior probability density function (PDF) is expressed as

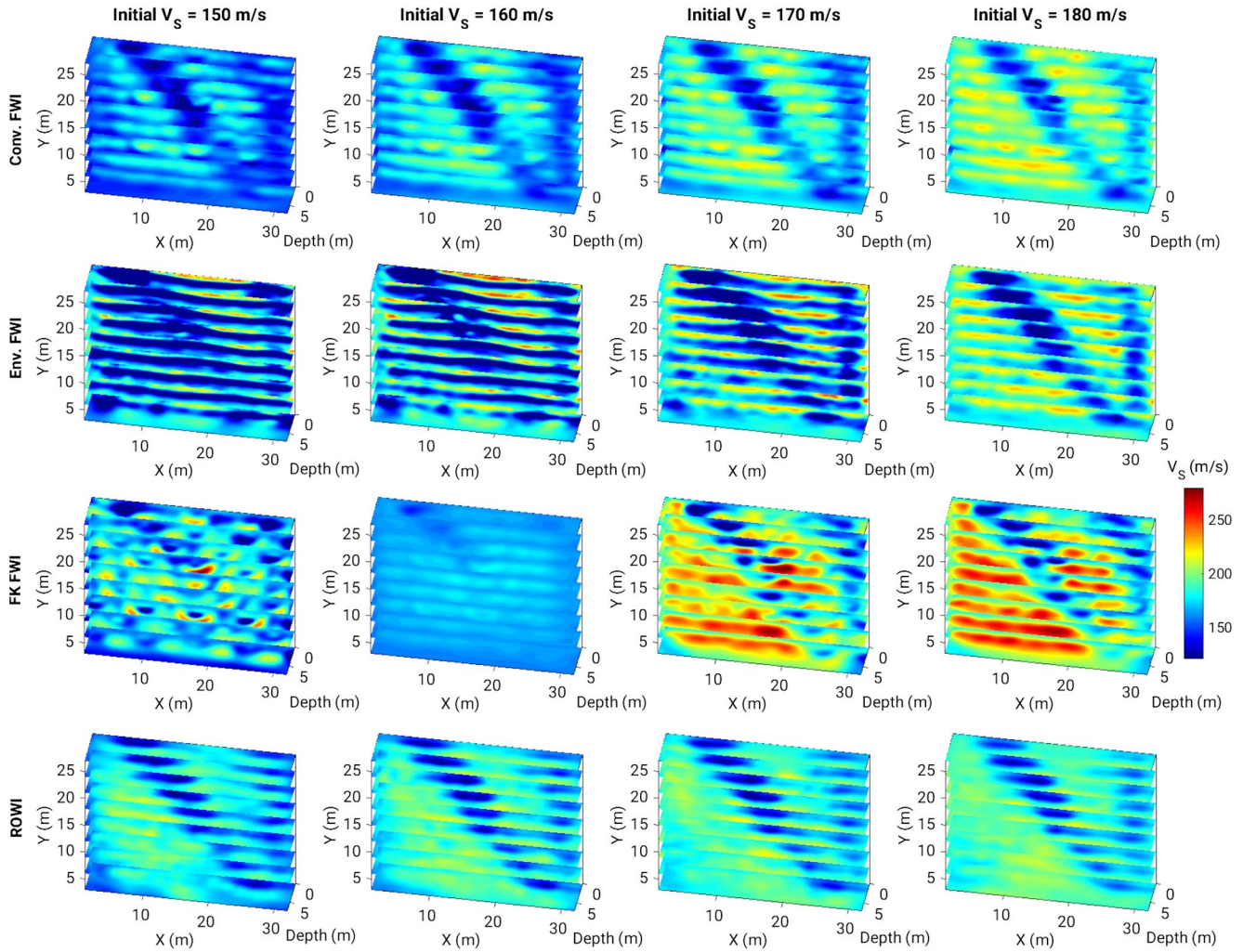


Figure 16. Inversion results in the field example by using the four poor initial models (four columns). Four rows represent the inversion results estimated by the conventional least squares full-waveform inversion (FWI), envelope-based FWI, f - k -spectra based FWI, and random-objective waveform inversion (ROWI), respectively.

$$\sigma(\mathbf{m}) = k \rho_M(\mathbf{m}) \exp[-\psi(\mathbf{m}, \hat{\mathbf{w}})], \quad (13)$$

where k is a normalization constant and $\rho_M(\mathbf{m})$ is the PDF of a priori information (Tarantola, 2005). The Pareto solution corresponding to the weight $\hat{\mathbf{w}}$ represents the maximum a posteriori (MAP) model.

Similarly, the posterior PDF corresponding to the multi-objective function Φ (we follow Shigapov, 2019 and call it parametric probability) is expressed as

$$\sigma(\mathbf{m}, \mathbf{w}) = k \rho_M(\mathbf{m}) \exp[-\psi(\mathbf{m}, \mathbf{w})]. \quad (14)$$

In a realistic case, multiple MAP models are estimated with different \mathbf{w} , and these models correspond to the Pareto solutions of the multi-objective function Φ . The collection of Pareto solutions represents the samples that maximize the parametric probability (MAP models that maximize the posterior PDFs with varying \mathbf{w} in Equation 14) and we estimate the trade-off information by analyzing their distribution in the model space. This trade-off information differs from the uncertainty based on the Bayesian approach that represents posterior PDF centered at the MAP model.

The trade-off analysis performed in this study (Figure 14) uses six ROWI tests, which overall costs around 9,700 times of forward/adjoint simulations (6 tests \times 216 iterations per test \times around 6 times forward/adjoint simulation per iteration; the remaining forward simulations are used for the estimation of source time functions). The computational cost is approximately equivalent to that of running 16 iterations in a conventional FWI. Although the local Hessian-based approach for uncertainty estimation is computationally inexpensive by the low-rank approximation of the inverse Hessian at the MAP model and may only need to solve several forward/adjoint equations per shot, the estimation of the MAP model is time-consuming by running a conventional FWI. Therefore, the overall computational cost for the trade-off analysis in ROWI and the uncertainty estimation using the local Hessian-based approach is comparable. The joint use of ROWI (to estimate MAP model) and local Hessian-based approach (to estimate the inverse Hessian at the MAP model) might provide a computationally efficient way for uncertainty estimation, which deserves to be studied in the future.

10.2. Pareto-Solution and Convergence Diagnostics

The trade-off analysis relies on the accurate estimation of Pareto solutions. The multiple runs of ROWI better explore the solution space and are essential to a comprehensive estimation of the Pareto front. It provides more candidate solutions for a fair evaluation of Pareto optimality. Although a single run of ROWI might be able to recover the subsurface model, it might not be sufficient for a reliable estimation of the trade-off information.

We use a maximal iteration as our stopping criterion in ROWI because the information from single-misfit values could not assure the convergence of the multi-objective function. The maximal iteration number needs to be large enough to guarantee the convergence of ROWI toward a Pareto solution. Once ROWI arrives at the Pareto front, it starts to explore the Pareto front. Therefore, a large number of iteration in ROWI benefits the reliability of the inversion results and their trade-off information on the price of an increased computational cost.

10.3. Limitations of ROWI

ROWI benefits the advantages and, at the same time, suffers the drawbacks of the chosen objective functions (e.g., non-convexity of the objective function). Because ROWI uses a local optimization algorithm and the objective functions are mostly non-convex, the inversion might converge toward a local Pareto solution or become inefficient when trapped at a saddle point (Ge et al., 2015), especially when an inappropriate initial model is chosen.

One technical problem ROWI faces is its computational performance on the parallel supercomputer. Because only one shot is used per iteration, we can only parallelize the code based on model-domain decomposition (Bohlen, 2002), which might become computationally inefficient when many (e.g., $> 10^3$; depending on the size of the model) cores are available. However, it will not be a problem when multiple ROWI tests are needed because they can be run independently in parallel.

Shallow-seismic data with dense acquisition provides a preferred situation for the application of ROWI. This is because surface waves, which dominate the observed data, have a similar and homogeneous illumination in the shallow subsurface among different shots. Thus, it results in relatively high redundancy in the data. How to adapt ROWI to other data sets deserves further study.

Although ROWI estimates the trade-off information of multiparameter models (i.e., V_S and V_P in Figure 14), it does not account for the interparameter trade-off (crosstalk). The interparameter trade-off is difficult to be assessed in ROWI because multiple measurements are used simultaneously and independently. It is more preferable to use a local Hessian-based approach (Fichtner & van Leeuwen, 2015; Pan, Geng, & Innanen, 2018) for the estimation of interparameter trade-off.

11. Conclusions

We promoted the ROWI method for 3D multi-component shallow-seismic data and applied it to a field data set. In every iteration, we randomly choose one shot and randomly assign one measurement to it to compute the model update. We applied ROWI to a nine-component shallow-seismic field data set and successfully delineated the targeted refilled trench. Several 2D GPR profiles acquired in the same test site also proved the validity of the estimated model. Trade-off analysis was performed statistically by using 246 approximated samples of the Pareto front, which suggested relatively lower reliability at the boundaries of the targeted trench and in another trench near the boundary of the survey area. We also ran four additional ROWI tests on the field data starting from poor initial models. The inversion results reconstructed the main structures reasonably well, which highlighted the high robustness of ROWI against poor initial models. Because the Pareto solutions of the subproblems do not accurately represent the Pareto solutions of the full multi-objective problem, efficient approaches for Pareto optimality evaluation and the exploration of the Pareto front require further studies.

Data Availability Statement

The raw field data is freely available from the KITopen repository (<https://doi.org/10.5445/IR/1000125628>). We thank students from KIT and Université Grenoble Alpes for their help in the field-data acquisition.

Acknowledgments

The authors thank Marwan Theodosios Irnaka for preprocessing the field data and Claudia Payne for proofreading. The authors also thank Dirk-Philip van Herwaarden and an anonymous reviewer for their constructive comments and suggestions. The authors gratefully acknowledge the financial support by the Deutsche Forschungsgemeinschaft (DFG) with Project-ID 258734477 - SFB 1173. Open access funding enabled and organized by Projekt DEAL.

References

- Backus, G., & Gilbert, F. (1970). Uniqueness in the inversion of inaccurate gross Earth data. *Philosophical Transactions of the Royal Society of London - Series A: Mathematical and Physical Sciences*, 266(1173), 123–192. <https://doi.org/10.1098/rsta.1970.0005>
- Bohlen, T. (2002). Parallel 3-D viscoelastic finite difference seismic modelling. *Computers & Geosciences*, 28(8), 887–899. [https://doi.org/10.1016/S0098-3004\(02\)00006-7](https://doi.org/10.1016/S0098-3004(02)00006-7)
- Borisov, D., Modrak, R., Gao, F., & Tromp, J. (2017). 3D elastic full-waveform inversion of surface waves in the presence of irregular topography using an envelope-based misfit function. *Geophysics*, 83(1), 1–45.
- Bozdag, E., Trampert, J., & Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, 185(2), 845–870.
- Brosse, N., Durmus, A., & Moulines, E. (2018). *The promises and pitfalls of stochastic gradient Langevin dynamics*. arXiv preprint arXiv:1811.10072.
- Bunks, C., Saleck, F. M., Zaleski, S., & Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5), 1457–1473. <https://doi.org/10.1190/1.1443880>
- Curtis, A., & Lomax, A. (2001). Prior information, sampling distributions, and the curse of dimensionality. *Geophysics*, 66(2), 372–378. <https://doi.org/10.1190/1.1444928>
- Fang, Z., Da Silva, C., Kuske, R., & Herrmann, F. J. (2018). Uncertainty quantification for inverse problems with weak partial-differential-equation constraints. *Geophysics*, 83(6), R629–R647. <https://doi.org/10.1190/geo2017-0824.1>
- Fichtner, A., & van Leeuwen, T. (2015). Resolution analysis by random probing. *Journal of Geophysical Research: Solid Earth*, 120, 5549–5573. <https://doi.org/10.1002/2015jb012106>
- Fichtner, A., & Zunino, A. (2019). Hamiltonian nullspace shuttles. *Geophysical Research Letters*, 46, 644–651. <https://doi.org/10.1029/2018gl080931>
- Gao, L., Pan, Y., & Bohlen, T. (2020). 2-D multiparameter viscoelastic shallow-seismic full-waveform inversion: Reconstruction tests and first field-data application. *Geophysical Journal International*, 222(1), 560–571. <https://doi.org/10.1093/gji/ggaa198>
- Gao, L., Pan, Y., Tian, G., & Xia, J. (2018). Estimating Q factor from multi-mode shallow-seismic surface waves. *Pure and Applied Geophysics*, 175(8), 2609–2622. <https://doi.org/10.1007/s00024-018-1828-7>
- Ge, R., Huang, F., Jin, C., & Yuan, Y. (2015). Escaping from saddle points-online stochastic gradient for tensor decomposition. In *Conference on learning theory* (pp. 797–842).
- Gebräad, L., Boehm, C., & Fichtner, A. (2020). Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo. *Journal of Geophysical Research: Solid Earth*, 125, e2019JB018428. <https://doi.org/10.1029/2019jb018428>
- Groos, L., Schäfer, M., Forbriger, T., & Bohlen, T. (2017). Application of a complete workflow for 2D elastic full-waveform inversion to recorded shallow-seismic Rayleigh waves. *Geophysics*, 82(2), R109–R117. <https://doi.org/10.1190/geo2016-0284.1>
- Häusler, M., Schmelzbach, C., & Sollberger, D. (2018). The Galperin source: A novel efficient multicomponent seismic source. *Geophysics*, 83(6), P19–P27. <https://doi.org/10.1190/geo2018-0020.1>
- Irnaka, T. M. (2021). *3D elastic full waveform inversion for subsurface characterization. study of a shallow seismic multicomponent field data (Doctoral dissertation, Université Grenoble Alpes)*. Retrieved from <https://tel.archives-ouvertes.fr/tel-03202027>
- Irnaka, T. M., Brossier, R., Métivier, L., Bohlen, T., & Pan, Y. (2019). Towards 3D 9C elastic full waveform inversion of shallow seismic wavefields-case study Ettligen Line. In *81st EAGE conference and exhibition 2019* (pp. 1–5). <https://doi.org/10.3997/2214-4609.201900994>
- Izzatullah, M., van Leeuwen, T., & Peter, D. (2019). Bayesian uncertainty estimation for full waveform inversion: A numerical study. In *SEG technical program expanded abstracts 2019* (pp. 1685–1689). Society of Exploration Geophysicists. <https://doi.org/10.1190/segam2019-3216008.1>
- Keating, S. D., & Innanen, K. A. (2021). Null-space shuttles for targeted uncertainty analysis in full-waveform inversion. *Geophysics*, 86(1), R63–R76. <https://doi.org/10.1190/geo2020-0192.1>

- Liu, Q., & Peter, D. (2019). Square-root variable metric based elastic full-waveform inversion—Part 2: Uncertainty estimation. *Geophysical Journal International*, 218(2), 1100–1120. <https://doi.org/10.1093/gji/ggz137>
- Liu, Q., & Peter, D. (2020). Square-root variable metric-based nullspace shuttle: A characterization of the nonuniqueness in elastic full-waveform inversion. *Journal of Geophysical Research: Solid Earth*, 125, e2019JB018687. <https://doi.org/10.1029/2019jb018687>
- Louboutin, M., Guasch, L., & Herrmann, F. J. (2017). Data normalization strategies for full-waveform inversion. In *79th EAGE conference and exhibition 2017* (pp. 1–5). <https://doi.org/10.3997/2214-4609.201700720>
- Masoni, I., Brossier, R., Virieux, J., & Boelle, J. (2013). Alternative misfit functions for FWI applied to surface waves. In *75th EAGE conference & exhibition incorporating spe europec 2013*. <https://doi.org/10.3997/2214-4609.20130297>
- Métivier, L., Allain, A., Brossier, R., Mérogot, Q., Oudet, E., & Virieux, J. (2018). Optimal transport for mitigating cycle skipping in full waveform inversion: A graph space transform approach. *Geophysics*, 83(5), 1–84. <https://doi.org/10.1190/segam2018-2997001.1>
- Métivier, L., & Brossier, R. (2016). The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. *Geophysics*, 81(2), F1–F15. <https://doi.org/10.1190/geo2015-0031.1>
- Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media.
- Pan, W., Geng, Y., & Innanen, K. A. (2018). Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: Synthetic experiments and Hussar land data set application. *Geophysical Journal International*, 213(2), 1305–1333. <https://doi.org/10.1093/gji/ggy037>
- Pan, Y., & Gao, L. (2020). Random objective waveform inversion of surface waves. *Geophysics*, 85(4), EN49–EN61. <https://doi.org/10.1190/geo2019-0613.1>
- Pan, Y., Gao, L., & Shigapov, R. (2020). Multi-objective waveform inversion of shallow seismic wavefields. *Geophysical Journal International*, 220(3), 1619–1631. <https://doi.org/10.1093/gji/ggz539>
- Pan, Y., Schaneng, S., Steinweg, T., & Bohlen, T. (2018). Estimating S-wave velocities from 3D 9-component shallow seismic data using local Rayleigh-wave dispersion curves—A field study. *Journal of Applied Geophysics*, 159, 532–539. <https://doi.org/10.1016/j.jappgeo.2018.09.037>
- Pérez Solano, C., Donno, D., & Chauris, H. (2014). Alternative waveform inversion for surface wave analysis in 2-D media. *Geophysical Journal International*, 198(3), 1359–1372. <https://doi.org/10.1093/gji/ggu211>
- Pladys, A., Brossier, R., & Métivier, L. (2017). FWI alternative misfit functions—what properties should they satisfy? In *79th EAGE conference and exhibition*, (pp. 1–5).
- Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M. K. (2014). Seismic tomography and the assessment of uncertainty. In *Advances in Geophysics* (Vol. 55, pp. 1–76). Elsevier. <https://doi.org/10.1016/bs.agph.2014.08.001>
- Schaneng, S. P. (2017). *Erstellung eines 3D Modells der Scherwellengeschwindigkeit im Bereich der Ettlinger Linie (Rheinstetten) aus der 1D Inversion der lokalen Dispersion von Rayleigh-Wellen (Master's thesis, Karlsruhe Institute of Technology)*. <https://doi.org/10.5445/IR/1000080199>
- Shigapov, R. (2019). *Probabilistic waveform inversion: Quest for the law (Doctoral dissertation, Karlsruhe Institute of Technology, (KIT))*. <https://doi.org/10.5445/IR/1000091433>
- Shin, C., Jang, S., & Min, D.-I. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49(5), 592–606. <https://doi.org/10.1046/j.1365-2478.2001.00279.x>
- Siahkoobi, A., Rizzuti, G., Louboutin, M., Witte, P. A., & Herrmann, F. J. (2021). *Preconditioned training of normalizing flows for variational inference in inverse problems*. arXiv preprint arXiv:2101.03709.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.
- Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature Physics*, 2(8), 492–494. <https://doi.org/10.1038/nphys375>
- Thurin, J., Brossier, R., & Métivier, L. (2019). Ensemble-based uncertainty estimation in full waveform inversion. *Geophysical Journal International*, 219(3), 1613–1635. <https://doi.org/10.1093/gji/ggz384>
- van Herwaarden, D. P., Boehm, C., Afanasiev, M., Thrastarson, S., Krischer, L., Trampert, J., & Fichtner, A. (2020). Accelerated full-waveform inversion using dynamic mini-batches. *Geophysical Journal International*, 221(2), 1427–1438.
- van Leeuwen, T., & Herrmann, F. J. (2013). Fast waveform inversion without source-encoding. *Geophysical Prospecting*, 61, 10–19. <https://doi.org/10.1111/j.1365-2478.2012.01096.x>
- Virieux, J., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6), WCC1–WCC26. <https://doi.org/10.1190/1.3238367>
- Welling, M., & Teh, Y. W. (2011). Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (icml-11)* (pp. 681–688).
- Yang, H., Jia, J., Wu, B., & Gao, J. (2018). Mini-batch optimized full waveform inversion with geological constrained gradient filtering. *Journal of Applied Geophysics*, 152, 9–16. <https://doi.org/10.1016/j.jappgeo.2018.02.011>
- Yang, Y., & Engquist, B. (2017). Analysis of optimal transport and related misfit functions in full-waveform inversion. *Geophysics*, 83(1), A7–A12. <https://doi.org/10.1190/segam2017-17754089.1>
- Yuan, Y. O., Bozdağ, E., Ciardelli, C., Gao, F., & Simons, F. J. (2020). The exponentiated phase measurement, and objective-function hybridization for adjoint waveform tomography. *Geophysical Journal International*, 221(2), 1145–1164. <https://doi.org/10.1093/gji/ggaa063>
- Yuan, Y. O., Simons, F. J., & Bozdağ, E. (2015). Multiscale adjoint waveform tomography for surface and body waves. *Geophysics*, 80(5), R281–R302. <https://doi.org/10.1190/geo2014-0461.1>
- Zhang, X., & Curtis, A. (2020). Variational full-waveform inversion. *Geophysical Journal International*, 222(1), 406–411. <https://doi.org/10.1093/gji/ggaa170>
- Zhang, Z., Alajami, M., & Alkhalifah, T. (2020). Wave-equation dispersion spectrum inversion for near-surface characterization using fibre-optics acquisition. *Geophysical Journal International*, 222(2), 907–918. <https://doi.org/10.1093/gji/ggaa211>
- 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. <https://doi.org/10.1190/geo2015-0641.1>