



HAL
open science

Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern Nankai Trough revisited

Andrzej Górszczyk, Stéphane Operto, Michal Malinowski

► To cite this version:

Andrzej Górszczyk, Stéphane Operto, Michal Malinowski. Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern Nankai Trough revisited. *Journal of Geophysical Research : Solid Earth*, 2017, 122 (6), pp.4601 - 4630. 10.1002/2016JB013891 . hal-01737099

HAL Id: hal-01737099

<https://hal.science/hal-01737099>

Submitted on 5 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

RESEARCH ARTICLE

10.1002/2016JB013891

Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern Nankai Trough revisited

Key Points:

- New robust workflow minimizing nonlinearity of the crustal-scale FWI
- Extensive quality control for validation of the starting and final FWI model
- New geological insight into structure of the eastern Nankai Trough

Supporting Information:

- Supporting Information S1

Correspondence to:

A. Górszczyk,
agorszczyk@igf.edu.pl

Citation:

Górszczyk, A., S. Operto, and M. Malinowski (2017), Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern Nankai Trough revisited, *J. Geophys. Res. Solid Earth*, 122, 4601–4630, doi:10.1002/2016JB013891.

Received 27 DEC 2016

Accepted 8 MAY 2017

Accepted article online 11 MAY 2017

Published online 22 JUN 2017

Andrzej Górszczyk¹ , Stéphane Operto², and Michał Malinowski¹

¹Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland, ²Université Cote d'Azur, CNRS, OCA, Géoazur, Valbonne, France

Abstract Crustal-scale imaging by the full-waveform inversion (FWI) of long-offset seismic data is inherently difficult because the large number of wavelengths propagating through the crust makes the inversion prone to cycle skipping. Therefore, efficient crustal-scale FWI requires an accurate starting model and a stable workflow minimizing the nonlinearity of the inversion. Here we attempt to reprocess a challenging 2-D ocean-bottom seismometer (OBS) data set from the eastern Nankai Trough. The starting model is built by first-arrival traveltimes tomography (FAT), which is FWI assisted for tracking cycle skipping. We iteratively refine the picked traveltimes and then reiterate the FAT until the traveltimes residuals remain below the cycle-skipping limit. Subsequently, we apply Laplace-Fourier FWI, in which progressive relaxation of time damping is nested within frequency continuation to hierarchically inject more data into the inversion. These two multiscale levels are complemented by a layer-stripping approach implemented through offset continuation. The reliability of the FWI velocity model is assessed by means of source wavelet estimation, synthetic seismogram modeling, ray tracing modeling, dynamic warping, and checkerboard tests. Although the viscoacoustic approximation is used for wave modeling, the synthetic seismograms reproduce most of the complexity of the data with a high traveltimes accuracy. The revised FWI scheme produces a high-resolution velocity model of the entire crust that can be jointly interpreted with migrated images derived from multichannel seismic data. This study opens a new perspective on the design of OBS crustal-scale experiments amenable to FWI; however, a further assessment of the optimal OBS spacing is required for reliable FWI.

1. Introduction

The current paradigm on how crustal-scale velocity models based on long-offset stationary-receiver surveys are built relies on ray-based methods such as first-arrival traveltimes tomography (FAT) [e.g., Zelt and Barton, 1998]. This is mainly because such methods offer acceptable resolution for delineating the main crustal units while remaining affordable in terms of data acquisition. In practice, this means that during a typical marine wide-angle reflection/refraction (WARR) survey, ocean-bottom seismometers (OBSs) are nowadays placed 5–10 km apart [e.g., Lester et al., 2014]. However, when moving from the smooth velocity models inferred via FAT to the multilayer models constructed via joint refraction/reflection traveltimes inversion [e.g., Zelt and Smith, 1992], the interpreter's ability to distinguish different crustal phases in recorded seismograms and to associate them with particular layers becomes critical, causing the velocity-model-building process to become highly subjective [Malinowski, 2013].

As an alternative approach, since the pioneering work of Tarantola [1984], full-waveform inversion (FWI) has been undergoing active development in both academia and the oil and gas industry (see Virieux and Operto [2009] for a review). FWI allows one to develop subsurface models with a much better resolution than can be achieved using FAT (a half-wavelength versus the width of the first Fresnel zone [Williamson, 1991; Virieux and Operto, 2009]) and has the potential to be utilized for multiparameter reconstructions [Operto et al., 2013]. Moreover, it does not require the identification and picking of individual phases, since the ultimate goal of FWI is to automatically account for all types of arrivals. However, the cost of using more complex information in the data space is higher nonlinearity of the inversion, which can be mitigated by means of suitable data preprocessing and multiscale inversion strategies [e.g., Bleibinhaus et al., 2009; Kamei and Pratt, 2013]. This nonlinearity makes a careful appraisal of the FWI results mandatory. One of the most obvious quality control consists of computing full-waveform synthetic seismograms using the FWI-derived subsurface model

and assessing the agreement between the modeled and recorded data for all interpretable arrivals. In this way, FWI itself can be viewed as a tool for aiding in the interpretation of complex arrivals in the data and of geological features in the subsurface model.

In addition to the requirement for dense acquisition, FWI requires a good starting velocity model to converge toward the global minimum of the misfit function. "Good" here means that the model should predict the traveltimes of any arrival involved in the inversion to within half a period of the lowest inverted frequency when a classical misfit function based on the data difference is used [Virieux and Operto, 2009]. Otherwise, cycle skipping during data fitting will drive the FWI toward a local minimum. The cycle-skipping criterion becomes challenging to fulfill when the number of propagated wavelengths is large, since the relative traveltimes error is inversely proportional to this number [Pratt, 2008]. This is typically the case for WARR surveys that are dominated by diving waves and postcritical reflections that have propagated over several tens of wavelengths. Therefore, Sirgue [2006] and Plessix *et al.* [2012] advocate the necessity of recording as low frequencies as possible to mitigate the cycle-skipping problem. The recording of low frequencies is conventionally complemented by the design of multiscale approaches that rely mostly on frequency continuation from low values to higher ones [Bunks *et al.*, 1995; Pratt, 1999].

A feasibility study of crustal-scale imaging from 2-D WARR synthetic data via FWI was first presented by Pratt *et al.* [1996], while Operto *et al.* [2006] presented a pioneering application to real crustal-scale OBS data. In the latter study, the data set was collected during a 2-D dense OBS survey (with a 1 km instrument spacing) in the complex geological setting of the eastern Nankai Trough subduction zone (Tokai area), offshore of Japan. This was the very first attempt to use FWI to obtain a high-resolution crustal velocity model (105 × 25 km in size). Subsequent FWI applications at the crustal scale focused on the western Nankai Trough region [Kamei *et al.*, 2012, 2013, 2014], although at a smaller scale (65 × 15 km) and in a less tectonically complex environment.

The past decade has witnessed a number of significant developments in FWI. Therefore, 10 years after Operto *et al.* [2006], we attempt here to reprocess the same eastern Nankai Trough data set, with the aim of benefiting from state-of-the-art FWI technology. Our first key improvement is related to the careful quality control of the first-arrival traveltimes picking, which has prompted us to reiterate FAT until the resulting velocity model is sufficiently accurate to prevent cycle skipping at the starting FWI frequency. Second, we propose a new three-loop multiscale FWI workflow that combines Laplace-Fourier waveform inversion [Brossier *et al.*, 2009; Shin and Cha, 2009] with a layer-stripping approach implemented by means of offset continuation. The Laplace-Fourier inversion is helpful not only for designing a multiscale data-driven workflow by means of frequency continuation and progressive time-damping relaxation but also for generating low frequencies in the data. For this case study, the most aggressive time-damping factors allowed us to start the FWI with a frequency as low as 1.5 Hz. The misfit function and the optimization algorithm are two other key FWI ingredients. While Kamei *et al.* [2014] used phase-only inversion based on a logarithmic norm for deep crustal imaging, we rather favor a classical waveform-difference-based misfit function, in which the data residuals are properly scaled with offset, and a quasi-Newton optimization algorithm for efficient FWI gradient preconditioning. This optimization design allows us to build a P wave velocity (V_p) model of the entire crust with a significant resolution improvement compared to that achieved by FAT. This case study relies on viscoacoustic seismic modeling, and we only update V_p using density (ρ) and P wave quality factor (Q) as passive parameters. Although more realistic viscoelastic physics and multiparameter reconstruction are desired, our results strongly support the relevance of the viscoacoustic approximation for this case study. Third, we improve several other aspects of the FWI workflow. In particular, we improve the coherency and signal-to-noise ratio of the data during preprocessing, we account for multiples during seismic modeling and inversion, and we propose a complete workflow for appraising FWI models and quantifying the a posteriori misfits between the real and modeled data based on the use of Dynamic Image Warping (DIW) [Hale, 2013].

We begin our paper by reviewing the basic principles of frequency-domain FWI before presenting the steps of the Laplace-Fourier FWI workflow with offset continuation. Then, we proceed to apply the proposed workflow to the eastern Nankai Trough OBS data set. First, we present our revised data preprocessing sequence. Second, we review the steps of the initial velocity-model-building process performed using FAT. Then, we move on to the application of FWI and discuss in detail how the multiscale, layer-stripping FWI workflow is tuned. This is followed by the presentation of the FWI results and the interpretation of the main structural units. Subsequently, we address the quality control of the FWI results based on source wavelet estimation, synthetic seismogram, and ray tracing modeling and quantification of the misfits between the observed

and modeled data by means of dynamic seismogram warping. Resolution of our final velocity model is assessed based on checkerboard tests. We also present in the supporting information the influence of the OBS spacing on the FWI results.

2. Frequency-Domain FWI

FWI is an iterative least squares data-fitting procedure that is performed to reconstruct subsurface parameters such as wave speeds from recorded wavefields. Seismic modeling relies on solving the full wave equation to automatically account for all types of arrivals and finite frequency effects. In its original form [Tarantola, 1984; Lailly, 1984], FWI was implemented in the time domain. Frequency-domain FWI was subsequently introduced by G. Pratt in the nineties, with first application for cross-hole acquisition [Pratt and Worthington, 1990] and later on for surface wide-angle acquisition [Pratt et al., 1996]. However, the first convincing real data application was presented by Ravaut et al. [2004] based on land data collected from a thrust belt in the southern Apennines. The frequency domain provides the most natural framework for the implementation of a frequency continuation multiscale approach, while being computationally efficient for 2-D stationary-receiver wide-angle acquisitions [Sirgue and Pratt, 2004]. Moreover, the frequency-domain formulation allows one to easily implement attenuation during seismic modeling through the use of complex-valued velocities [Toksöz and Johnston, 1981; Song et al., 1995]. It also provides a feasible means of estimating source signatures, which is generally alternated with the estimation of the subsurface parameters during FWI iterations [Pratt, 1999].

The reader is referred to the following references about the pros and cons of the time-domain versus frequency-domain formulations of FWI according to several criteria as acquisition geometries, wave physics, and data preprocessing [Vigh and Starr, 2008; Virieux et al., 2009; Plessix, 2007; Operto et al., 2015].

2.1. Forward Modeling

The 2-D viscoacoustic wave equation can be written in the frequency domain as follows:

$$\frac{\omega^2}{\kappa(x, z)} p(x, z, \omega) + \frac{\partial}{\partial x} \left(\frac{1}{\rho(x, z)} \frac{\partial p(x, z, \omega)}{\partial x} \right) + \frac{\partial}{\partial z} \left(\frac{1}{\rho(x, z)} \frac{\partial p(x, z, \omega)}{\partial z} \right) = -s(x, z, \omega), \quad (1)$$

where ω is the angular frequency, $\kappa(x, z)$ is the bulk modulus, $p(x, z, \omega)$ is the pressure field, $\rho(x, z)$ is the density, and $s(x, z, \omega)$ is the source term. After discretization, this equation can be recast in matrix form as

$$\mathbf{A}\mathbf{P} = \mathbf{S}, \quad (2)$$

where \mathbf{A} denotes the impedance matrix, the coefficients of which depend on the modeled frequency and the subsurface parameters, and the matrices $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n]$ and $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n]$ contain the monochromatic pressure wavefields and the sources, respectively, stored as column vectors. For areal stationary-receiver acquisitions such as WARR acquisitions, in which the number of receivers can be 1 order of magnitude smaller than the number of shots, the shots and receivers are processed in a reciprocal manner by virtue of the spatial reciprocity of Green's functions. The linear system described by equation (2) is typically solved using sparse direct solvers [Duff et al., 1986] to efficiently process the multiple right-hand sides of equation (2). In this framework, discretization techniques that minimize the numerical bandwidth of \mathbf{A} are useful to minimize the fill-in of the matrix during lower-upper factorization [e.g., Jo et al., 1996; Hustedt et al., 2004].

Conventionally, the data misfit vector is the differences between the recorded data \mathbf{d}_{obs} and the modeled data \mathbf{d}_{cal} for all n sources addressed in the survey:

$$\Delta \mathbf{d} = \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}, \quad (3)$$

where the vector \mathbf{d}_{cal} is built by sampling the modeled wavefields \mathbf{p}_i , $i = 1, \dots, n$, at the source positions, equation (1). Several frequency components can also be aggregated in the vectors \mathbf{d}_{obs} and \mathbf{d}_{cal} to perform simultaneous inversion of multiple frequencies.

2.2. Inversion

The subsurface model is updated by minimizing the weighted least squares norm of the data misfit, which is given by

$$C(\mathbf{m}) = \Delta \mathbf{d}^\dagger \mathbf{W}_d \Delta \mathbf{d}, \quad (4)$$

where \dagger denotes the transpose conjugate and \mathbf{W}_d is a data-preconditioning operator that weights each component of $\Delta \mathbf{d}$ in the misfit function C .

The misfit function C is minimized using local optimization techniques to iteratively and nonlinearly update a subsurface model by applying perturbations to a starting model \mathbf{m}_k :

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta \mathbf{m}_k, \quad (5)$$

where k denotes the iteration number. The model perturbation $\Delta \mathbf{m}_k$ is given by

$$\Delta \mathbf{m}_k = -\alpha_k \mathbf{H}_k^{-1} \nabla_{\mathbf{m}} C_k, \quad (6)$$

where the descent direction is the product of the steepest-descent direction $-\nabla_{\mathbf{m}} C$ with the inverse of the second derivative of C (the Hessian), whereas the distance to move along the descent direction is the step length α_k , which can be computed via line search [Nocedal and Wright, 2006]. The gradient $\nabla_{\mathbf{m}} C$ is typically computed using the adjoint-state method, which requires only two modelings per source [Plessix, 2006]. The explicit computation of the Hessian and its inverse is computationally expensive. To overcome this computational burden, diagonal approximations of the Gauss-Newton Hessian \mathbf{H}_g [e.g., Pratt et al., 1998] can be used to correct the gradient for geometrical spreading effects [Shin et al., 2001; Ravaut et al., 2004]. With this approach, the expression for the model perturbation becomes

$$\Delta \mathbf{m}_k = -\alpha_k (\mathcal{G} \text{diag} \mathbf{H}_{g_k} + \epsilon \mathbf{I})^{-1} \mathcal{G} \nabla_{\mathbf{m}} C_k, \quad (7)$$

where ϵ is a damping factor and \mathbf{I} is the identity matrix. In equation (7), we introduced a Gaussian smoothing operator \mathcal{G} as a regularization term, whose correlation length can be related to the local wavelength [Ravaut et al., 2004]. Alternatively, the action of the Hessian on the gradient can be more fully taken into account by using a limited-memory quasi-Newton algorithm such as l-BFGS [Byrd et al., 1995] or the truncated Newton method [Métivier et al., 2014]. In this context, the diagonal approximation of the Hessian in equation (7) can be used as an efficient preconditioner for these second-order optimization techniques [Nocedal and Wright, 2006; Métivier et al., 2014]. For a more thorough discussion of the benefits of second-order optimization methods in FWI, see Pratt et al. [1998].

2.3. Multiscale Layer-Stripping Workflow

Data-driven strategies are conventionally designed to mitigate the nonlinearity of FWI. The most common strategy relies on a multiscale approach based on frequency continuation. In the frequency domain, this strategy can be naturally implemented by means of successive inversions of monochromatic data sets of increasing frequency [e.g., Pratt, 1999]. Moreover, when a wide-angle acquisition generates a broad range of aperture angles, a computationally efficient FWI algorithm can be designed by limiting the inversion to a few discrete frequencies, which amounts to reducing the redundancy of the wave number coverage in the model space [Sirgue and Pratt, 2004]. Similar strategies can be implemented in the time domain by progressively augmenting the amplitude spectrum of the seismograms with higher frequencies over iterations [e.g., Bunks et al., 1995].

A second hierarchical level of multiscaling can be implemented by selecting specific arrivals in the inversion. Typically, in the framework of diffraction tomography, to which FWI is closely related [Wu and Toksöz, 1987], early arrivals such as diving waves and postcritical reflections are associated with large scattering angles, whereas short-offset reflections are associated with small scattering angles. Low frequencies and wide scattering angles control the long-wavelength updates during a tomography-like reconstruction, whereas high frequencies and narrow scattering angles control the reconstruction of the shorter wavelengths during a migration-like process [Pratt et al., 1996; Virieux and Operto, 2009, equation (30)]. In contrast to frequency continuation, scattering-angle continuation can be more naturally implemented in the time domain via windowing. The early stages of inversion are performed using narrow windows centered on the early arrivals, and the window length is then progressively increased to include late arrivals such as short-spread reflections. One issue with frequency-domain FWI is that the frequency-domain convolution that corresponds to the time-domain windowing cannot be implemented when only a limited number of discrete frequencies are processed simultaneously.

A second best solution is provided by the so-called Laplace-Fourier approach, in which exponential time damping is applied through the use of complex-valued frequencies to control the amount of information that is preserved after the first arrival [Brossier et al., 2009; Shin and Cha, 2009].

The Laplace-Fourier approach exploits the expression for the Fourier transform of a damped signal $p(t)e^{-(t-t_0)/\tau}$, which is given by

$$\text{TF} (p(t)e^{-(t-t_0)/\tau}) = \int_{-\infty}^{\infty} p(t)e^{-i(\omega+i/\tau)t} e^{t_0/\tau} dt, \quad (8)$$

where TF denotes the Fourier transform, t_0 the first-arrival traveltime, and τ the time-damping factor. This expression shows that a monochromatic component of a seismogram that is damped in time can be computed by solving the time-harmonic wave equation, equation (1), using a complex-valued frequency.

In a manner similar to the time-domain approach, multiscale strategies can be designed by progressively relaxing the damping (that is, increasing τ) over the successive steps of inversion. Such a damping continuation strategy can be nested as an inner loop within the frequency continuation implemented in the outer loop of the FWI algorithm, thereby yielding a two-level multiscale approach [Brossier *et al.*, 2009].

Although the frequency and scattering-angle continuation strategies are useful for mitigating the nonlinearity of FWI, they might not be sufficient to prevent cycle skipping when many wavelengths are propagated, as in WARR surveys. Therefore, a third data-driven strategy based on offset continuation can be implemented, such that the number of wavelengths that are propagated during seismic modeling is progressively increased as the accuracy of the subsurface model improves. The offset continuation and the time-damping relaxation should be judiciously combined in the data space to jointly perform a layer-stripping approach [e.g., Shipp and Singh, 2002; Wang and Rao, 2009; Takougang and Calvert, 2011] and the scale continuation in the model space. Such an approach implies that data recorded at short offsets and long recording times (i.e., deep near-offset reflections) preferably should not be inverted before longer-offset earlier arrivals (i.e., postcritical reflections) since it might violate the scale continuation condition. Therefore, the progression over the offset intervals should form the innermost loop of the FWI algorithm.

3. Application to OBS Data Collected in the Eastern Nankai Area

3.1. Geodynamical Context and Data Acquisition

The Nankai Trough is an oceanic trough that extends for a distance of approximately 900 km along the south-eastern coast of Japan. It delineates the subduction zone where the Philippine Sea plate subducts below the overriding Eurasian plate in the northwest direction at a rate of approximately 4.5 cm/yr [Seno *et al.*, 1993] (see Figure 1a). The entire fault region (530 km in length), from the Shikoku Island in the west to the Tokai district in the east, can be divided into four fault segments (Figure 1a), which have been cyclically active in the past with the recurrence time of large earthquakes estimated to be between 100 and 200 years [Ando, 1975a]. The easternmost Tokai segment has remained unruptured since 1854, and hence, it is expected to be the site of the next megathrust earthquake in the Nankai area [Ando, 1975b; Le Pichon *et al.*, 1996].

The proximity of the collision zone between the Izu-Bonin Arc and central Japan induces transitional subduction and collision behavior in the Tokai segment (Figure 1a). In particular, the Izu collision zone implies the formation of cyclical oceanic ridges, which affect the local geodynamic setting and govern the nucleation and propagation of seismic ruptures as well as the deformation of the surrounding structures, especially the accretionary prism [Mazzotti *et al.*, 1999, 2002; Kodaira *et al.*, 2003]. Within the Tokai area, three such topographical features parallel to the axis of the trench have been reported: (i) the Zenisu Ridge, which is currently located immediately before the trench; (ii) the Paleo-Zenisu Ridge, which is located beneath the accretionary prism; and (iii) the Deeper Paleo-Zenisu Ridge, which is subducting beneath the fore arc.

In order to gain new insight on the main structural factors controlling the seismological regime of the Tokai segment, a dense OBS data set was recorded in the eastern Nankai Trough during the 2001 KY0106 cruise (R/V *Kaiyo*) as part of the Seize France Japan (SFJ) project [Dessa *et al.*, 2004a, 2004b; Operto *et al.*, 2006]. For this purpose, 100 OBSs spaced 1 km apart were deployed along a 100 km long profile. A 140 km long profile of 1404 air gun sources spaced 100 m apart was acquired using an air gun array with a total volume of 196 L. The OBSs used during the survey were equipped with 4.5 Hz three-component geophones and hydrophones. Seismic shooting was performed during two staggered passes, with shots released every 200 m, to prevent the recording of water waves and noise from previous shots. The orientation of the profile was roughly perpendicular to the subducting structures, resulting in significant variations in bathymetry, from a water depth of approximately 500 m in the coastal region to 3750 m in the trench. A coincident regional multichannel

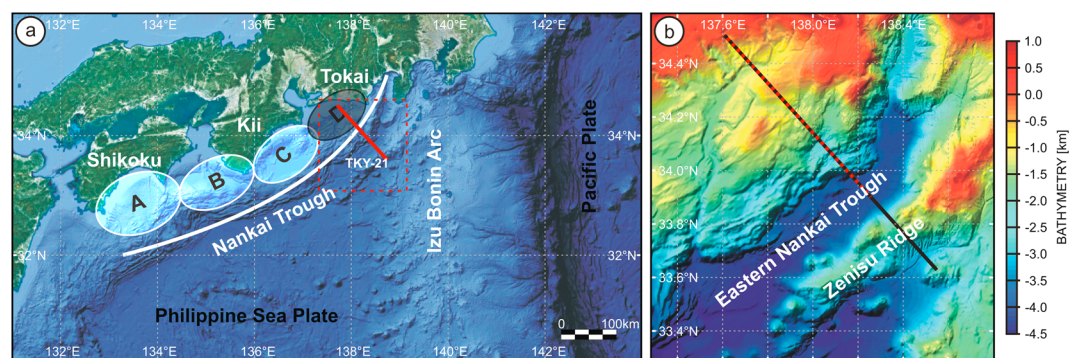


Figure 1. (a) Partitioning of the Nankai Trough into four fault planes as described by Ando [1975a]. Region D was left unruptured during the most recent sequence of two large earthquakes (1944 Tonankai and 1946 Nankaido). The solid red line represents the seismic profile of the TKY-21 experiment. (b) Zoomed view of the TKY-21 survey area, overlaid with the bathymetry variations. The black line and the dashed red line represent the shot profile and the receiver line, respectively.

seismic reflection (MCS) profile was also acquired using a 4.5 km long streamer during the SFJ-MCS survey. A prestack depth migrated (PSDM) section generated with this data set [Martin, 2003] is used in the following to verify the upper part of our FWI model.

3.2. Preprocessing of the Data

Ninety-two OBS gathers were available for inversion [Dessa *et al.*, 2004a; Operto *et al.*, 2006]. We use the vertical-component geophone data rather than hydrophone recordings because of their higher quality. The purpose of the FWI data preprocessing is to improve the quality of the data while preserving the true amplitudes to the greatest possible extent for use during the inversion. The first stage of the processing consists of removing the DC bias from the seismic traces, eliminating bad shots and applying an approximate 3-D to 2-D amplitude correction by multiplying each sample by the square root of time. An amplitude analysis reveals that the root-mean-square (RMS) values of the signals consistently differ from OBS to OBS, likely as a result of differences in the coupling conditions between the geophone and the sea bottom at each OBS position. Without corrections, these differences would cause OBS gathers with higher RMS amplitudes to have a stronger impact on the FWI gradient than those with lower RMS amplitudes. To mitigate this issue, we apply amplitude balancing among the OBS gathers while preserving the amplitude-versus-offset (AVO) variations in each gather.

Figures 2a and 2c show two OBS gathers after the first stage of preprocessing. The heterogeneity of the geological medium manifests itself in the complexity of the recorded wavefield. This is especially visible in the shadow zone indicated by the white shaded area.

In the second stage, we focus on trace editing and improving the signal-to-noise ratio (SNR). We apply despiking to attenuate high-amplitude outliers in the data. Additionally, we analyze the RMS amplitudes calculated before the first breaks (i.e., noise RMS, or NRMS) to identify traces that are affected by high-amplitude background noise. Such traces can potentially introduce artifacts and noise into the gradient during FWI. Under the assumption that the NRMS should be rather uniform among all channels (i.e., shot points), we search through all traces for abnormally high NRMS values and iteratively scale down their amplitudes until the NRMS matches the desired mean NRMS for the data set. We found this approach superior to simply skipping those noisy traces. When we remove them from the input data, we lose coverage in the deeper parts of the model, which is expressed as the overall worse data fit during the inversion.

The signal becomes weaker and more difficult to track as the offset increases (Figures 2a and 2c). In order to improve the coherency of the data, we use the Discrete Curvelet Transform (DCT) [Candès *et al.*, 2006] which outperforms more conventional coherency filterings (e.g., F-X deconvolution or F-K filtering) because of its ability to simultaneously track coherent events with variable dips. This minimizes the possibility of damaging the signal during filtering, especially in areas where different arrivals overlap with each other. Although it might be more efficient to directly filter monochromatic data in the shot-receiver plane for frequency-domain FWI [e.g., Adamczyk *et al.*, 2015], the 1 km spacing between OBSs in our case prevents the use of this approach. Instead, we apply the DCT in the time domain following the strategy presented by Górszczyk *et al.* [2014].

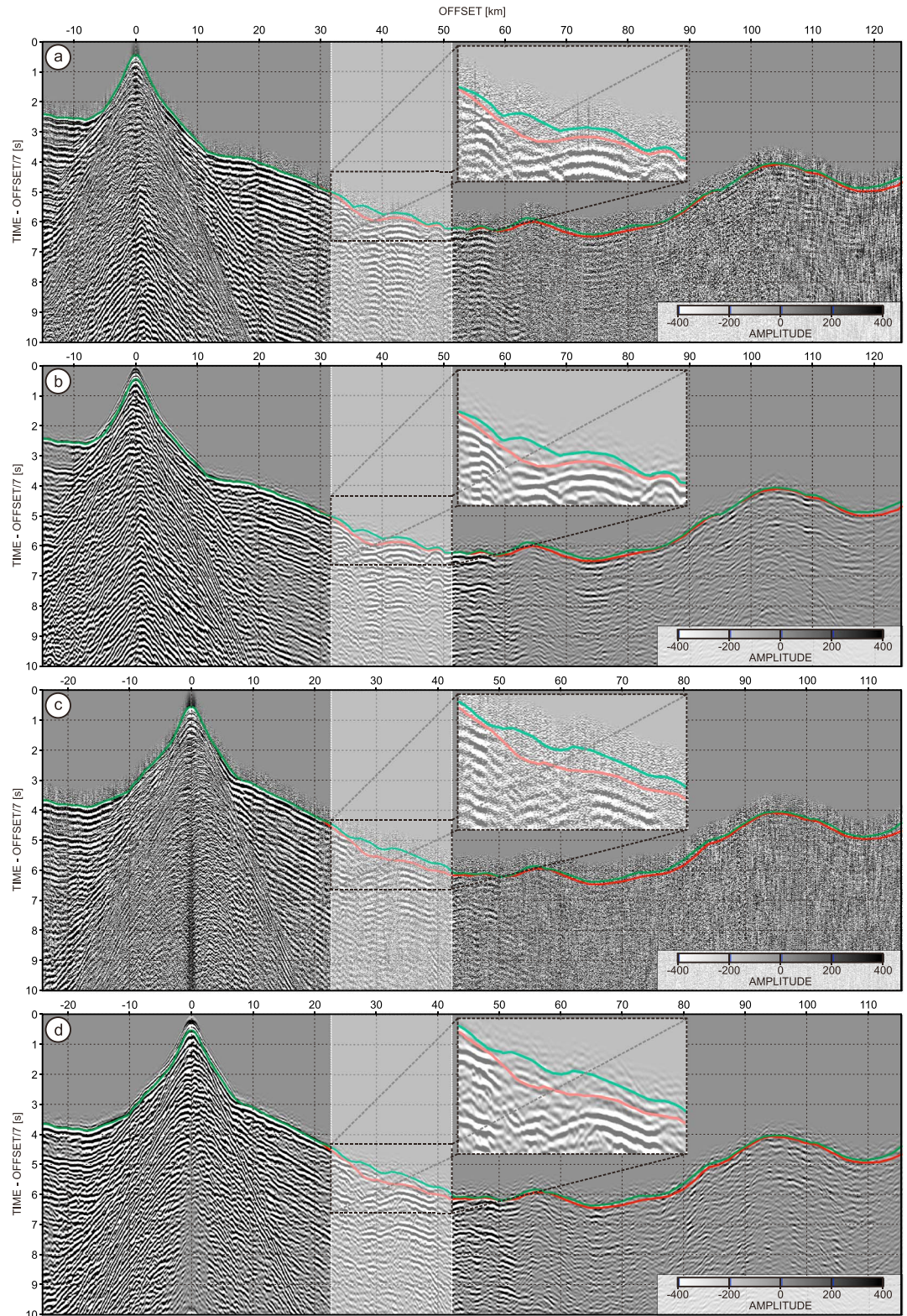


Figure 2. OBS gather 14 (a) before and (b) after coherency filtering and deconvolution. The red and green lines represent the initial and final first-arrival traveltime picks, respectively. The white shaded area highlights a complex portion of the data characterized by a shadow zone with a weak first arrival and a complex set of postcritical reflections and diffractions. (c, d) Same as Figures 2a and 2b but for OBS gather 23.

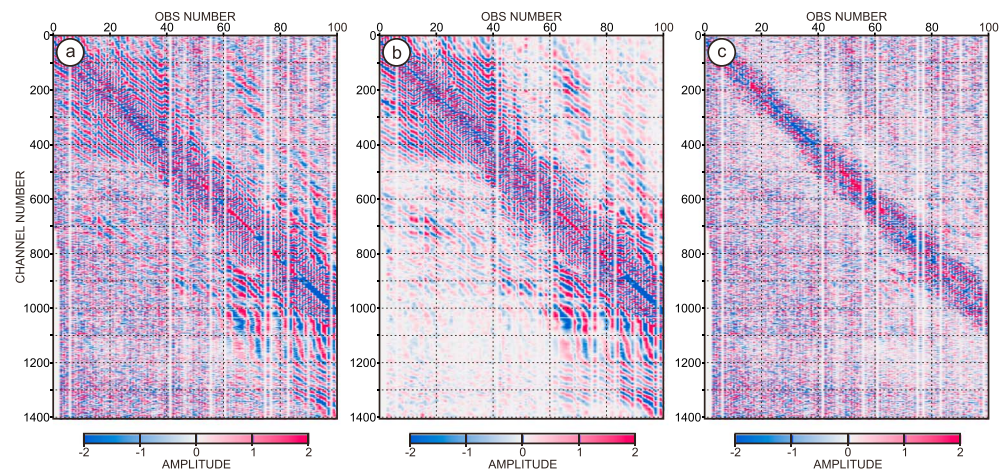


Figure 3. Damped monochromatic data visualized in the shot-receiver coordinate system. The presented frequency is 1.5 Hz, and the time-damping constant τ is 0.2 s. The data set is shown (a) before and (b) after coherency filtering. (c) The difference between Figures 3a and 3b, showing the removed noise and outliers.

In the final stage, we apply spiking deconvolution to whiten the amplitude spectrum and eliminate bubble effects. In the framework of FWI, the spectral whitening can be viewed as a data preconditioning, which balances the contribution of different frequencies in the misfit function. This preconditioning is particularly useful for our data set, whose amplitude spectrum exhibits significant notches [Operto *et al.*, 2006, Figure 5; Kamei *et al.*, 2014, Figure 1]. Finally, we apply a Butterworth band-pass filter that selects frequencies between 1 and 9 Hz to attenuate high-frequency noise. The OBS gathers presented in Figures 2a and 2c are shown after preprocessing in Figures 2b and 2d.

As we perform FWI in the frequency domain, we assess our data preprocessing in this domain. The comparison between a damped monochromatic data set before and after trace editing and coherency filtering shows the significant reduction of the incoherent energy and the improved continuity of the data after processing (Figures 3a and 3b). Even for a frequency as low as 1.5 Hz, we still observe a considerable coherent signal (Figure 3b). Areas of weaker amplitudes are also clearly evident, which are probably related to geological structures characterized by high attenuation. The residuals between the data before and after coherency filtering reveal that a significant number of random energy and amplitude outliers are removed during processing (Figure 3c). Moreover, there is no evidence of coherent patterns, meaning that we have fully preserved the useful signal for FWI.

3.3. First-Arrival Traveltime Tomography

In this study, we perform FAT using the FAST code [Zelt and Barton, 1998] to build the initial velocity model for FWI. This model must satisfy the cycle-skipping criterion; namely, it must predict the recorded traveltimes with an error that does not exceed half the period of the lowest-frequency data used in the FWI. Pratt [2008] reformulates this criterion as the following inequality:

$$\frac{\Delta t}{T} < \frac{1}{2N_\lambda}, \quad (9)$$

where $\Delta t/T$ is the relative time error and N_λ is the number of propagated wavelengths.

In this study, this condition can be quite challenging to fulfill with long-offset data even if we manage to begin the inversion at a frequency as low as 1.5 Hz. For this frequency, more than 40 wavelengths are propagated for a maximum offset of 140 km and a mean velocity of 5 km/s, whereas this number reaches 238 for a frequency of 8.5 Hz (the maximum frequency used in this study). Therefore, we must carefully design the workflow for the first-arrival traveltime picking and the quality control of the FAT results as presented hereafter.

3.3.1. First-Arrival Traveltime Picking

First-break picking was a challenging task for our data set. Despite the significant improvement in the quality of the data achieved during preprocessing, the interpretation of the first breaks remains ambiguous in some areas. In particular, low-velocity zones (LVZs), described in detail by Park *et al.* [2010] and references therein, generate shadow zones in the OBS gathers (Figure 2, shaded area), in which the first breaks are difficult to pick

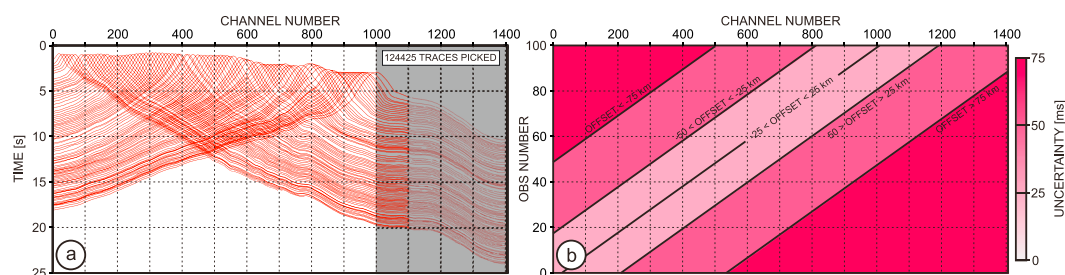


Figure 4. (a) First-arrival traveltimes plotted in the shot-time coordinate system. The gray-shaded area represents the region not covered by OBSs. (b) First-arrival traveltimes uncertainties used during FAT, plotted in the shot-receiver coordinate system.

because of vanishing amplitudes (Figure 2, red curves in insets). On the one hand, omitting these complicated portions of the data during picking degrades illumination of important structures, and hence the accuracy of the velocity model. During subsequent FWI, this degraded accuracy can generate cycle skipping of energetic wide-angle reflections arriving shortly after the ambiguous first arrivals (Figure 2, insets). Likewise, assigning large traveltimes uncertainties to weak arrivals during FAT will not solve this issue as they will also impact the accuracy of the portions of the velocity model illuminated by these arrivals. On the other hand, the consistent picking of incorrect phases across multiple OBS gathers will automatically drive the tomography to incorrect results.

Because of these two difficulties, we choose to make an extra effort to involve the problematic parts of the first-arrival wavefield during FAT and pick as many traces as possible (124425 in total, Figure 4a) with model-driven approach in speculative areas and at far offsets. However, as we are aware of the possibility of introducing inconsistent data, we perform our FAT in three steps, during which we repeat the picking process in complex areas based on the assessment of the traveltimes and phase residuals and velocity model achieved during previous steps.

3.3.2. FAT Results

In this study, we use the same starting model for FAT as that used by *Dessa et al.* [2004a] (Figure 5a). The grid interval for the first-arrival traveltimes computations is 25 m, and the inversion cell size is defined as 1 km in the vertical direction and 2 km in the horizontal direction to keep the resulting model smooth. The shaded area in Figure 5 corresponds to the region covered by shots but not by receivers. Velocity model in this region is constrained by rays coming only from one side; hence, its validity is approximate. However, it is still reasonable to utilize these shots for the sake of better illumination in the remaining parts of the model.

Consistently with the 2-D nature of our tomography, we first project the OBS positions onto a line such that the offsets between each OBS and the first shot are preserved. Then, we project the shot positions on the line to restore the initial shot-receiver offsets. Most of the real OBSs positions deviate by less than 150 m from the projected ones, especially on the landward side of the profile. Higher deviations are observed on the deep-water side: Eight OBSs near the trench are offset by more than 150 m from the profile and three of them by about 350 m.

We perform four nonlinear iterations of traveltimes inversion. In each iteration, we iteratively search for the best trade-off parameter that minimizes the data misfit while guaranteeing the desired smoothness of the velocity model [Zelt and Barton, 1998]. We use the bathymetry as a boundary above which the velocity model is not updated and is equal to 1.5 km/s.

As a result of testing a few methods of assigning the optimal picking uncertainty, an approach that consists of progressively increasing the uncertainties from 25 ms to 75 ms with increasing offset is found to provide a more uniform distribution of the traveltimes residuals and a more reliable velocity model compared with an alternative strategy based on a constant uncertainty (see Figure 4b). We also found that the offset-class method yields similar results to the one obtained when uncertainties are directly inferred from the data, yet being simpler to implement.

FAT performed using the initial set of picked traveltimes produces the velocity model shown in Figure 5b. Although the shallow velocities are geologically plausible, there are several deep, large-scale anomalies at distances between 65 and 80 km whose validity is questionable, as suggested by significant positive

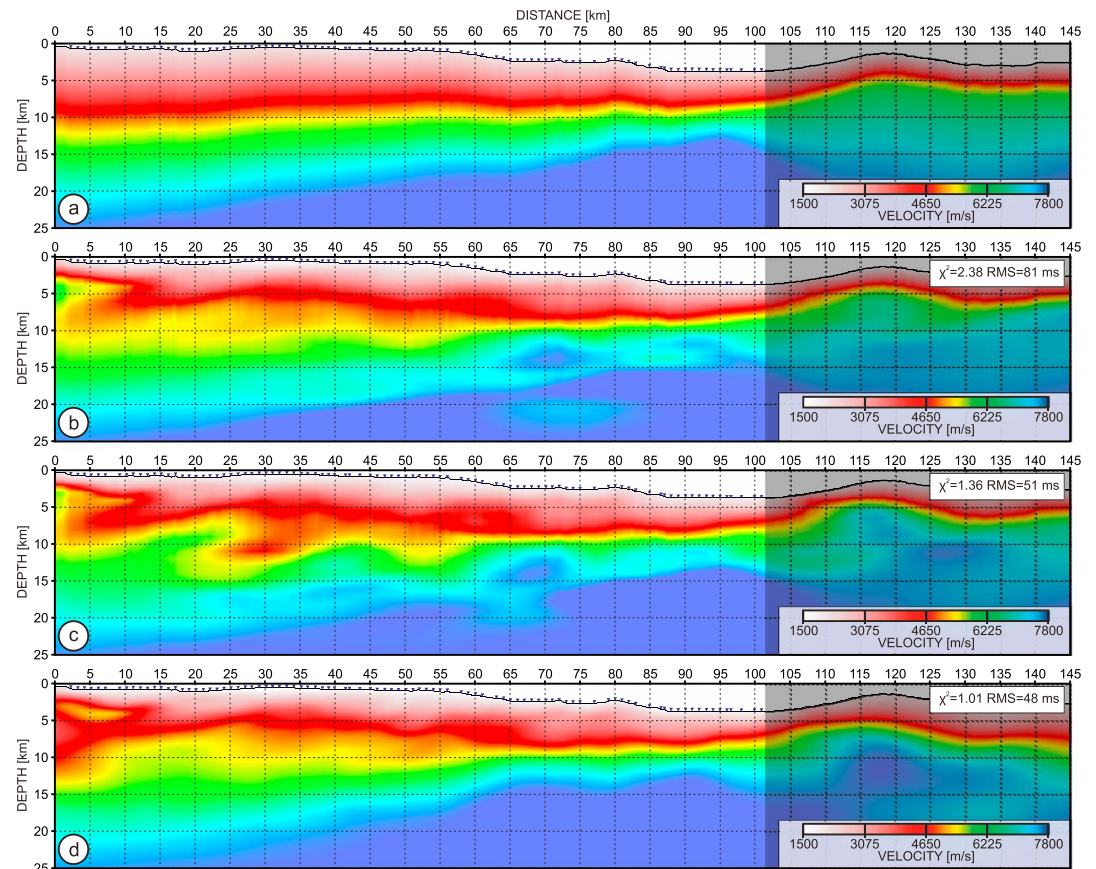


Figure 5. Velocity models obtained at different stages of the FAT workflow. The solid black line and triangles represent the bathymetry and the OBS positions, respectively. The gray-shaded area represents the region not covered by OBSs. (a) Initial model. Velocity models after the (b) first, (c) second, and (d) third stages of FAT (see the text for details).

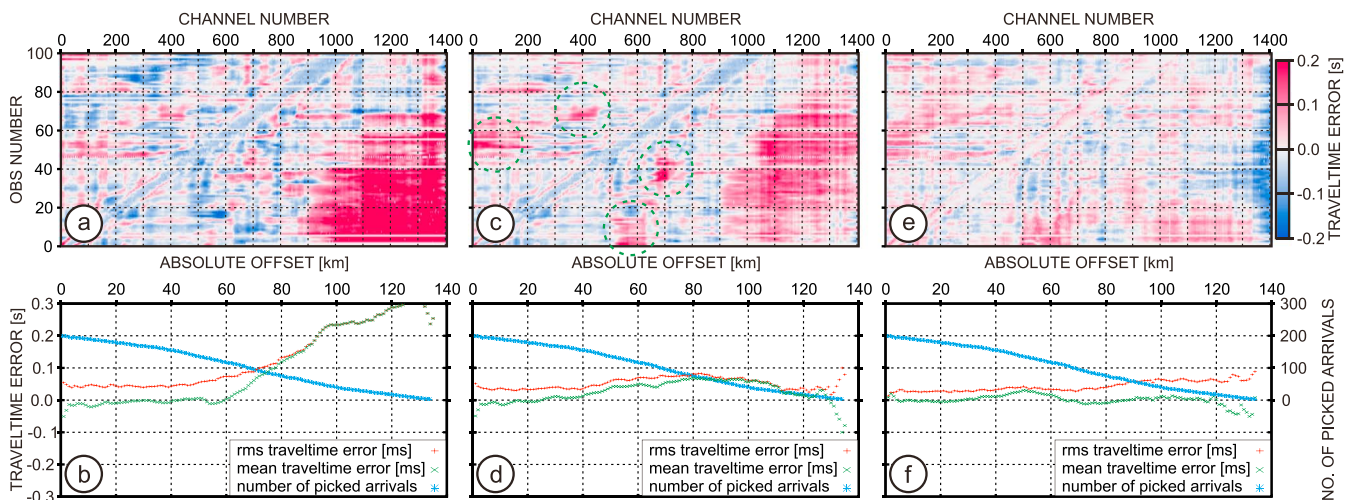


Figure 6. Traveltime misfit analysis. (a, c, and e) Residuals between the picked and computed traveltimes (plotted in the shot-receiver coordinate system) after the first (Figure 6a), second (Figure 6b), and third (Figure 6c) stages of FAT. The dashed circles in Figure 6c delineate the complex parts of the data, which contain the shadow zones indicated in Figure 2. (b, d, and f) Root-mean-square and mean traveltimes errors computed at 1 km offset intervals. Each blue star denotes the number of picks for a given absolute offset.

traveltime residuals at offsets beyond 60 km, which indicate overestimated velocities (Figures 6a and 6b). After this first stage of FAT, the final RMS and the normalized χ^2 are 81 ms and 2.26, respectively. Note that the channels with values higher than 1000 in Figure 6a, 6c, and 6d (see also Figure 4a, shaded part) correspond to area not covered by OBSs.

To improve the FAT results, we refine our picking in the ambiguous areas using the traveltimes computed in the previous FAT model as a guide for potential picking revision. This is helpful for identifying potentially improperly picked traveltimes, especially at far offsets or in the shadow zones and refining them such that they show better agreement with the computed ones. Indeed, this local model-driven picking requires careful quality control of the resulting FAT model as discussed in the following.

After rerunning the FAT using the updated picks, we obtain the velocity model shown in Figure 5c. Compared with the model from the previous stage (Figure 5b), we see subtle differences at intermediate and greater depths. However, the traveltime residuals are significantly decreased (compare Figures 6a and 6c), and the error distribution with offset is now more uniform than before (compare Figures 6b and 6d). We notice, however, that the better data fits at far offsets are achieved at the expense of slightly higher traveltime mismatches at intermediate offsets. The areas marked with dashed circles in Figure 6c correspond to shadow zones in the OBS gathers where the continuity of the first arrivals is lost, leading to higher residuals. After this second stage of FAT, the global RMS and normalized χ^2 have decreased (51 ms and 1.36, respectively) compared with the previous iteration. Further error analysis reveals that 95.4% of the traveltime residuals fall below the cycle-skipping criterion under the assumption that the FWI starts at a frequency of 3 Hz. Therefore, the FAT model obtained in this iteration is considered to be sufficiently accurate for initial FWI tests conducted for validation of the starting model.

The last step of our FAT workflow first aims to identify with the approach of *Shah et al.* [2012] whether the FAT model generates cycle skipping. Even if local cycle skipping may be corrected during subsequent FWI, it would clearly indicate that the FAT model needs further improvements. Analysis of the phase residuals generated by the FAT model suggested occurrence of cycle skipping in the small area indicated in Figure 6c by the dashed circle near OBS 40 and channel 700. This prompts us to perform a few iterations of FWI to check whether they improve the kinematic accuracy of the velocity model. Finally, in a manner similar to the second FAT stage, we use the traveltimes computed in this improved velocity model to assist a second iteration of the picking in the shadow zones (Figure 2, green curves). This new set of traveltime picks was then used to perform the final FAT iteration. The resulting velocity model (Figure 5d) is clearly smoother than the previous one (Figure 5c), while providing an improved traveltime fit, with an RMS and a normalized χ^2 of 48 ms and 1.01, respectively (Figures 6e–6f).

In summary, we conclude that our iterative approach to velocity model building using FAT, albeit heuristic, should provide us with a suitable starting model for FWI in the sense that it satisfies three important specifications. First, its smoothness suggests that erroneous data have not been overfitted via artificial short-scale structures. It is important as it guarantees that the FAT model will not leave any footprint of artificial short-scale structures in the FWI updates. Second, it fits those first-arrival traveltimes that have been unambiguously picked. Third, it does not generate obvious cycle skipping of the early wavefield at the starting FWI frequency. This last requirement has been fulfilled during the last stage of our workflow by assisting the first-arrival picking procedure in complex areas with preliminary FWI results.

3.4. Full-Waveform Inversion

We use the 2-D frequency-domain FWI code TOY2DAC developed within the framework of the SEISCOPE project. Forward modeling is performed with the finite-difference method of *Hustedt et al.* [2004] and the multifrontal solver MUMPS [*MUMPS team*, 2015]. We process sources and the vertical geophone sensors in a reciprocal manner, which means that the vertical geophones are processed as vertical forces and the explosive sources as hydrophones [*Operto et al.*, 2006]. The reciprocal hydrophones and vertical forces are positioned in the finite-difference grid with windowed sinc functions and their vertical derivatives, respectively [*Hicks*, 2002].

In contrast to *Operto et al.* [2006, Figures 5a and 5b], who muted the observed data to remove the footprint of free-surface multiples and imposed an absorbing boundary condition atop the velocity model during the modeling step, we account for free-surface multiples by imposing a free-surface boundary condition. This is one of the key points in our case study because hydrophone data are not sufficiently reliable for the

elimination of free-surface multiples using the method of *Amundsen et al.* [2001]. Moreover, because the water depth is quite shallow above the Shimanto wedge (500 m), the arrival time of a free-surface multiple is quite close to that of the first arrival (delayed by 0.67 s). Applying a mute just above the free-surface multiple would dramatically limit the information content of the seismograms. Moreover, the rigorous equivalent of such a time mute could not be applied to the modeled data during frequency-domain FWI. Therefore, in addition to the improved subsurface illumination that might be provided by the wave paths of the free-surface multiples, considering these multiples during FWI provides us with the necessary versatility to tune the time damping in Laplace-Fourier FWI independently of the water depth.

The crustal velocity model used for FWI is 35 km deep and 145 km long. We discretize it with a 50 m grid interval in order to satisfy the criterion of four grid points per wavelength at a frequency of 8.5 Hz in most part of the model. This discretization criterion provides accurate solutions with the finite-difference stencil of *Hustedt et al.* [2004], while being optimal for the half-wavelength resolution of FWI.

We use viscoacoustic modeling with $Q = 10000$ in the water column and $Q = 200$ below the seabed. The latter value was empirically selected by comparing the AVO trends of the real and synthetic data [*Malinowski et al.*, 2011]. Although a uniform Q model might sound very approximate, *Kurzmann et al.* [2013] have shown that even such simple Q models are enough to improve the FWI results when Q is used as a passive parameter during the inversion.

The density (ρ) model is derived from the FAT velocity model using the Gardner relation [*Gardner et al.*, 1974] and remains unchanged during inversion. We attempted to update ρ using the Gardner relation after the inversion of each frequency group; however, in our case, it caused the data fit to partially degrade. Note that updating ρ from an explicit relationship with V_p during FWI would require to modify the gradient of the misfit function with respect to V_p accordingly since ρ should be formally substituted by its expression as a function of V_p in the wave equation for gradient derivation (R. Brossier, personal communication, 2015).

We perform the inversion using the L-BFGS algorithm and a line search satisfying the Wolfe conditions, as implemented in the SEISCOPE optimization toolbox [*Métivier and Brossier*, 2016]. We use the inverse of the diagonal pseudo-Hessian approximation as a preconditioner in the L-BFGS method, and regularization is performed via gradient smoothing, as shown in equation (7). For data preconditioning, we apply a linear gain with offset to the data misfit within the misfit function, operator \mathbf{W}_d in equation (4). This data weighting helps to balance deeper perturbations relative to shallower ones and complements the gradient preconditioning applied by the diagonal Hessian [*Operto et al.*, 2006]. One source wavelet is estimated for all OBS gathers and is updated every five iterations by linear inversion from the data weighted by the linear gain with offset [*Pratt*, 1999]. We have also tested the estimation of a single signature per OBS gather; however, we concluded that this approach to source estimation might favor the possible leakage of missing velocity perturbations into the estimation of the OBS-ground couplings.

To mitigate the nonlinearity of the inversion, we apply the scheme described in section 2.3. Below, we review the specific design of each element of the workflow for the eastern Nankai case study.

3.4.1. Frequency Management

We perform FWI in the 1.5–8.5 Hz frequency band. Unlike the efficient approach of *Sirgue and Pratt* [2004] based on a sparse subset of frequencies, we densely sample the frequencies with intervals of 0.25 Hz and 0.5 Hz below and above 4.5 Hz, respectively. This fine frequency sampling is motivated by the characteristics of the OBS data, which are dominated by wide-angle arrivals at the expense of clear short-spread reflections from deep reflectors. This leads to illumination in a quite narrow range of scattering angles, which should be balanced by a fine frequency sampling to prevent aliasing artifacts.

When a sparse set of frequencies is used in efficient frequency-domain FWI, the multiscale imaging is generally recast as sequential inversions of monochromatic datasets. By contrast, we exploit our fine frequency sampling to cluster few close frequencies during each step of the multiscale inversion to reduce the influence of noise by stacking more redundant information [*Brossier et al.*, 2009].

In this study, we tested three types of frequency grouping. In the first approach, each group of three frequencies is inverted sequentially, with an overlap of one frequency between each pair of consecutive groups (see Table 1, Sequential column). A drawback of this approach is that the data fit at lower frequencies can degrade as the inversion proceeds toward higher-frequency groups [*Kamei et al.*, 2015]. However, this effect can be mitigated by introducing the higher frequencies progressively. In the second approach, referred to

Table 1. Frequency-Domain Sampling

Group	Frequency (Hz)	
	Sequential	Quasi-Progressive
1	1.50 1.75 2.00	1.50 1.75 2.00
2	2.00 2.25 2.50	1.75 2.00 2.25 2.50
3	2.50 2.75 3.00	1.75 2.25 2.50 2.75 3.00
4	3.00 3.25 3.50	1.75 2.25 2.75 3.00 3.25 3.50
5	3.50 3.75 4.00	1.75 2.25 2.75 3.25 3.50 3.75 4.00
6	4.00 4.25 4.50	1.75 2.25 2.75 3.25 3.75 4.00 4.25 4.50
7	4.50 5.00 5.50	1.75 2.25 2.75 3.25 3.75 4.25 4.50 5.00 5.50
8	5.50 6.00 6.50	1.75 2.25 2.75 3.25 3.75 4.25 5.00 5.50 6.00 6.50
9	6.50 7.00 7.50	1.75 2.25 2.75 3.25 3.75 4.25 5.00 6.00 6.50 7.00 7.50
10	7.50 8.00 8.50	1.75 2.25 2.75 3.25 3.75 4.25 5.00 6.00 7.00 7.50 8.00 8.50

as the progressive approach, each new group is created by adding higher frequencies to the previous group to prevent the degradation of the low-frequency data fit. This scheme, which includes all previous frequencies each time a new group is generated, mimics the time-domain approach of *Bunks et al.* [1995]. As a consequence, the processing of the last group requires the simultaneous inversion of 21 frequencies in the 1.5–8.5 Hz band. Obviously, the progressive approach, although it improves the data fit, is far more computationally expensive than the sequential scheme. This results from the fact that the computational complexity of the frequency-domain modeling scales to the number of modeled frequencies. It is a clear disadvantage of the frequency-domain formulation relative to the time-domain counterpart when the inversion of broadband data is performed.

As a compromise, we therefore propose a quasi-progressive approach in which each new frequency group is generated by merging a new finely sampled high-frequency group with a sparse selection of frequencies from the previous groups (Table 1, Quasi-Progressive column). This scheme, which is the final selected strategy, yields a data fit that is comparable to that obtained using the progressive approach while being less computationally expensive.

3.4.2. Laplace-Constant Management

As mentioned in section 2.3, frequency group continuation is implemented in the outer loop of our FWI algorithm as the top hierarchical level of our multiscale imaging, whereas the second loop runs over the time-damping constants τ (see equation (8)). We use four time-damping constants τ (0.2, 0.5, 1, and 2 s), which are associated with four increasing traveltime intervals (Figure 7). Starting with a τ value as small as 0.2 also results in the generation of low frequencies through the circular convolution applied to the data in the frequency domain [*Ha and Shin, 2012*]. This artificial broadening of the frequency bandwidth helps us to begin FWI at a frequency as low as 1.5 Hz, which makes a key contribution to preventing cycle skipping (Table 1).

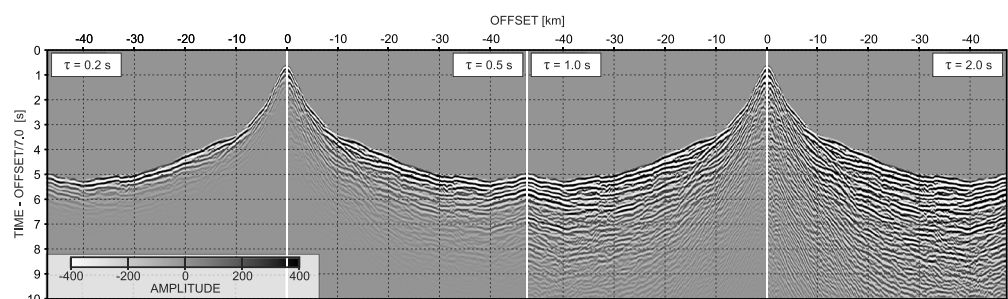


Figure 7. Negative offsets of OBS gather 46 after time damping. The time damping here is applied beginning with the first-arrival traveltime t_0 (equation (8)), and the data are muted above this traveltime. The four panels correspond to the four time-damping constants used during FWI and illustrate the temporal information that is preserved during inversion. From left to right: $\tau = 0.2$ s, $\tau = 0.5$ s, $\tau = 1$ s, and $\tau = 2$ s.

Table 2. Number of Iterations Within a Single Frequency Group for Each Scheme, Depending on the τ Value and Offset Interval

τ value (s)	Scheme I					Scheme II				
	Offset Interval (km)					Offset Interval (km)				
	25	50	75	100	140	25	50	75	100	140
0.2	5	10	15	20	25	5	–	–	–	–
0.5	30	35	40	45	50	10	15	–	–	–
1.0	55	60	65	70	75	20	25	30	–	–
2.0	80	85	90	95	100	35	40	45	50	55

3.4.3. Offset Management

The innermost loop of the FWI algorithm proceeds from short offsets to longer ones to implement our layer-stripping approach. We divide the data into five offset intervals (0–25, 0–50, 0–75, 0–100, and 0–140 km), which are processed hierarchically. By means of the weighting operator \mathbf{W}_d in equation (4), we apply inner and outer cosine tapers that penalize the data with offsets corresponding to the first and last 5 km of each interval to prevent truncation artifacts in the perturbation models.

Inside the three nested loops over frequency groups, time damping and offsets, we perform five FWI iterations using the preconditioned I-BFGS algorithm. During our initial investigations, we considered all five offset intervals for each frequency group and time-damping value. This approach led to a total number of 100 iterations per frequency group, as outlined in Table 2, Scheme I. However, we observed that the FWI of long-offset data performed with strong damping can lead to narrow artificial perturbations smeared along deep diving-wave paths. These artifacts probably result from back-projected noisy signal recorded at far offsets, whose influence on the misfit function gradient is boosted by the offset-dependent gain applied to the data residuals by the operator \mathbf{W}_d in equation (4). To overcome this issue, we restrict the inversion with strong damping ($\tau = 0.2$ s) to data with a maximum offset of 25 km. The high SNR and accurate picking of the first breaks within this offset range prevents the exponential damping from boosting the residual noise before first breaks (which is likely to occur at far offsets). Moreover, this short-offset interval contains the vast majority of the complex wavefield originating from the thrusts in the upper part of the model. Hence, strong damping efficiently minimizes the nonlinearity of the FWI in this offset range. For the subsequent damping constants, we start iterating again from offsets of up to 25 km; however, we progressively add longer-offset intervals as τ increases, until the last value of $\tau = 2.0$ s is reached, for which we perform the inversion for all offset intervals. We apply this new workflow (Table 2, Scheme II), which consists of 55 iterations for each frequency group, leading to a vast reduction in computational cost while providing consistent updates of the subsurface model.

3.4.4. Role of Gradient Preconditioning

When the inversion proceeds to longer-offset data, we notice that in addition to updating deeper parts of the model, it can generate shallow perturbations that are inconsistent with those already introduced in the previous inversion of the short-offset data. To overcome this issue, we complement the layer-stripping procedure by adapting the damping factor ϵ in the gradient preconditioner (equation (7)) to the offset interval. We assign one value of ϵ to each offset interval ($\epsilon = 1e-2, 1e-3, 1e-4, 1e-5, \text{ and } 1e-6$) such that ϵ decreases as the maximum offset increases. This implies that we favor the updating of the shallow subsurface when short offsets are used, consistently with our layer-stripping strategy. Moreover, we strongly damp the update of the shallow subsurface during the inversion of the long offsets to focus on the updating of the deep structures. This is illustrated in Figure 8, which shows two gradients preconditioned using the diagonal Hessian for the first frequency group and $\tau = 2$ s. In Figure 8a, the gradient has been computed for the first offset interval (maximum offset of 25 km) using $\epsilon = 1e-2$, leading to a damping of the gradient in the deep part of the model. By contrast, Figure 8b, the gradient has been computed for the final offset interval using $\epsilon = 1e-6$, leading to a damping of the gradient in the shallow part.

We find this approach to be especially useful during the early stages of the FWI when the velocity model is not yet accurate. At this stage, decreasing the damping factor of the diagonal Hessian helps to minimize the risk of introducing shallow artifacts while reliably updating the deeper parts of the model.

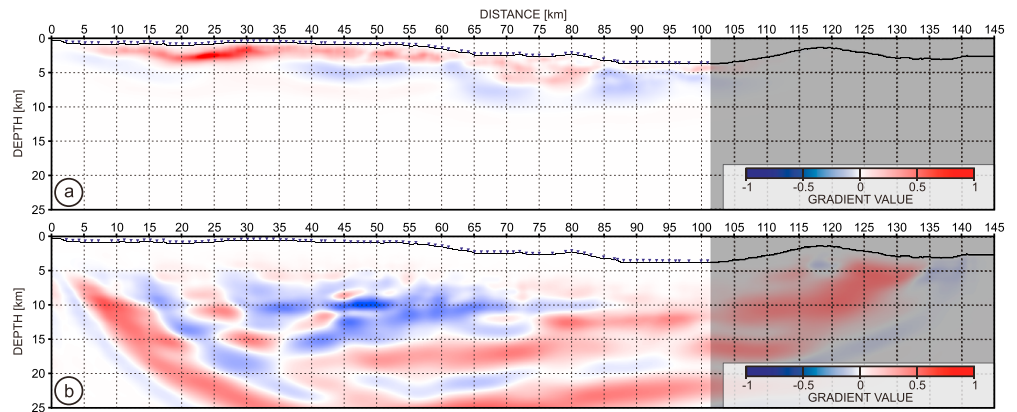


Figure 8. Preconditioned gradient of the FWI misfit function for different damping factors ϵ (see equation (7)). (a) $\epsilon = 1e-2$, used to process the first offset interval. (b) $\epsilon = 1e-6$, used to process the last offset interval. See Table 2.

3.5. FWI Velocity Models

The final FWI velocity model is shown in Figure 9a. The resolution improvement achieved by the FWI relative to the FAT result is significant, starting from shallow sediments and extending down to the deep crust (compare with Figure 5d). However, the fact that the velocities span a wide range of values in a crustal-scale velocity model can make the interpretation of geological structures challenging. For more convenient interpretation,

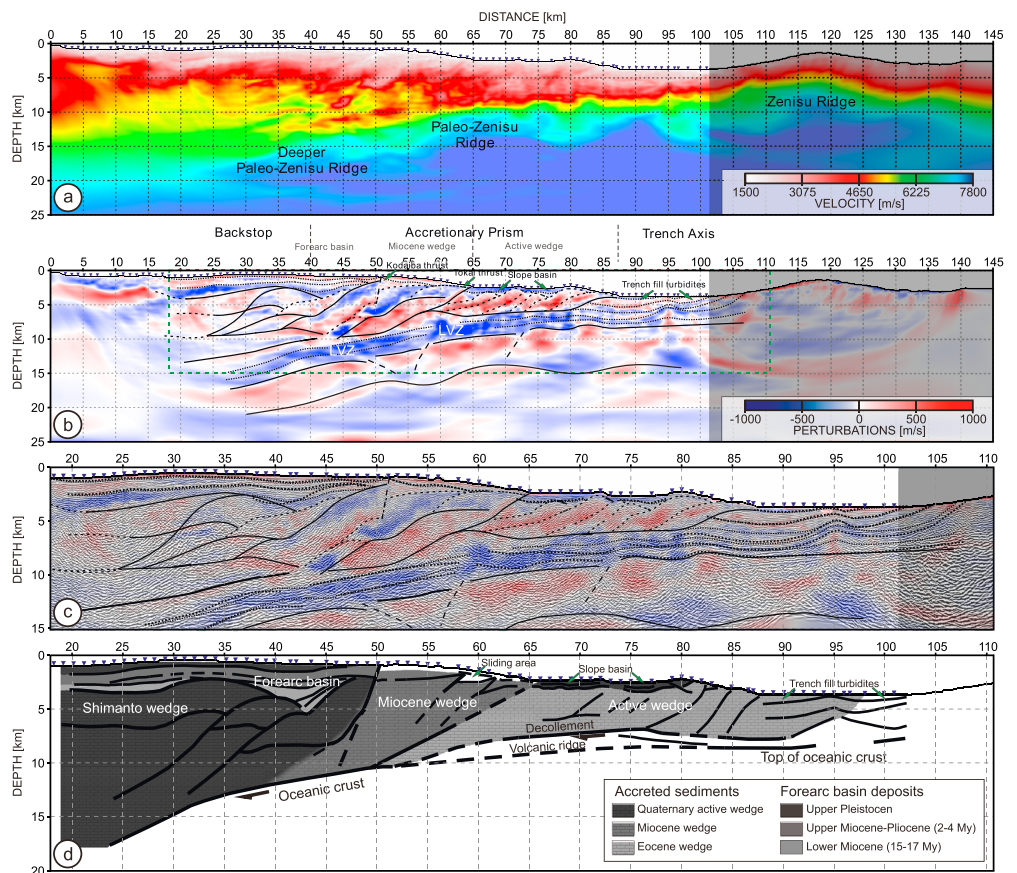


Figure 9. FWI results. (a) Final velocity model. (b) Perturbation model after detrending, with the interpretation overlaid. The main structural domains constituting the Nankai Trough, as interpreted by Henry *et al.* [2014], are labeled at the top of the panel. (c) Depth-migrated image from Martin [2003] overlaid on the detrended model. (d) Main structural units of the eastern Nankai Trough as interpreted by Henry *et al.* [2014]. Figure 9c is also shown in details in Figure S1 in the supporting information.

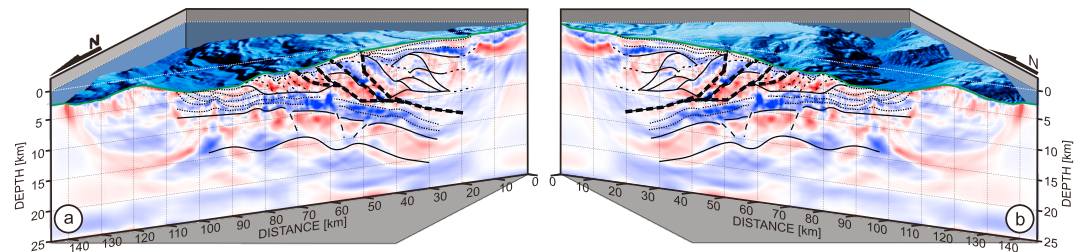


Figure 10. Cross section through the Nankai Trough along the profile overlaid with the bathymetry, the detrended model, the structural interpretation, and a megasplay fault branching toward the seabed, as indicated by the thick dashed line. Perspective view toward the (a) south and (b) north.

one can subtract either the starting FAT model or a background trend model from the final FWI model [Kamei *et al.*, 2012]. Here the background model is estimated as follows. First, each vertical column from the FWI model is replaced with the corresponding lateral average within a ± 1 km interval. Second, a fifth-order polynomial is fitted to each resulting 1-D section.

To support the interpretation of this detrended model, we show in Figure 9d the main structural units of the eastern Nankai Trough as interpreted by Henry *et al.* [2014]. Combining this information with a PSDM section [Martin, 2003] and other geophysical studies [Dessa *et al.*, 2004a, 2004b; Operto *et al.*, 2006] leads to the preliminary interpretation of the main structural units presented in Figures 9b and 9c.

Indeed, the detrended model (Figure 9b) reveals complex structures consistent with the overall geological characteristics of a subduction zone. In particular, we observe large-scale stacked thrust sheets of the Shimanto wedge (acting as a backstop) covered by the sediments of the fore-arc basin (Figure 9d). These structures spatially extend to the area of the Miocene wedge in the accretionary prism forming the Kodaiba and Tokai thrusts (Figure 9b at 50 km and 65 km distances). Going seaward, we can identify a sequence of smaller thrusts delineating an active wedge covered by slope basins. On the deep water side, we image the thick layers of sediments deposited within the trench between the active wedge and the Zenisu Ridge.

The upper part of the detrended model matches well with several horizons of the PSDM section (Figures 9c and S1 for figure magnification). Although the limited quality of the migrated image, good agreement is apparent in the shallow basins (fore-arc and slope basins) and in the Miocene wedge as well as in the trench-fill turbidites and in the active wedge affected by significant thrusting.

Figure 10 shows a perspective view of the detrended model along with the bathymetry of the Nankai area. We observe a good correlation between all of the topographic details and the underlying shallow structures, starting from the backstop and extending through the thrusts in the Miocene and active wedges up to the trench. One can track the extension of these structures down to their roots in the deeper parts of the model, in particular the faults and thrusts related to the branching of the splay fault (marked with dashed lines).

In supporting information (Text S1) we briefly discuss some potential geodynamic implications coming from our model. However, detailed geological interpretation of the obtained results is beyond the scope of the current paper.

3.6. FWI Model Appraisal

Although our final model reveals a lot of geologically consistent information, we must still validate its correctness in terms of the data fit. For this purpose, we estimate time-domain source wavelets and compare the recorded OBS gathers with the time-domain seismograms computed using the FAT and FWI models. Furthermore, we perform ray tracing modeling for interpretation of dominant phases followed by the checkerboard tests assessing resolution of our final model.

3.6.1. Source Wavelet Estimation

The estimation of source wavelets by linear waveform inversion relies on the assumption that the velocity model is known [Pratt, 1999]. This implies that any inaccuracy in the velocity model translates into a degradation in the amplitude and waveform repeatability of the source wavelets from one OBS gather to the next. Therefore, source wavelet estimation can be used as a tool for the FWI model assessment [Prioux *et al.*, 2011; Malinowski *et al.*, 2011]. Moreover, the sensitivity of the source wavelet estimation to the velocity model can be further increased by enhancing the contribution of the long offsets at the expense of the short ones in the

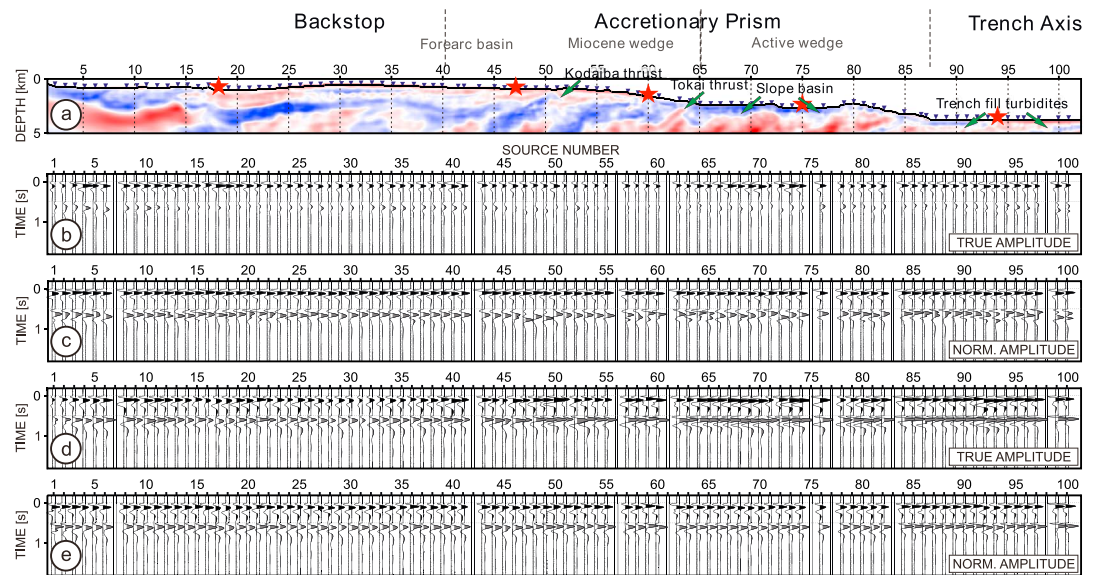


Figure 11. Source wavelet estimation. (a) Shallow part of the model covered by OBSs (triangles). The red stars indicate the positions of the OBSs presented in Figures 12 and S2–S7. (b, c) Source wavelets estimated from the weighted data using the FAT (gray wiggles) and FWI (black wiggles) velocity models. The wavelets are plotted without (Figure 11b) and with (Figure 11c) trace-by-trace amplitude normalization. The wavelets inferred from the FAT model are shifted by 0.5 s. (d, e) Same as Figures 11b and 11c for source wavelets estimated from unweighted data. The same amplitude scales are used in Figures 11b and 11d, and 11c and 11e.

source-estimation misfit function [Operto *et al.*, 2015]. Here we use the same linear gain with offset used for the FWI to weigh the data during the source estimation. A comparison of the source signatures inferred from the FAT and FWI models reveals that the amplitude and waveform repeatability of the wavelets is significantly improved after FWI, thereby validating the relevance of the FWI results (Figure 11b).

We also estimate the source signatures to enable a quantitative comparison of the recorded and synthetic seismograms after convolution with the estimated signatures (see the next section). For this purpose, we infer the source signatures from unweighted data to decrease the sensitivity of this estimation to the velocity model. Indeed, when unweighted data are used, the source estimation is predominantly driven by the high-amplitude direct water wave, which can be accurately modeled. This claim is supported by Figure 11d, which shows that the source wavelets computed from unweighted data using the FAT and FWI models are very similar. However, we still see increasing wavelet amplitudes with increasing water depth, which reflects that the source signatures become increasingly accurate in terms of amplitude as the water wave is recorded as a first arrival over a wider offset range. To remove this trend, we apply a trace-by-trace amplitude normalization to the wavelets in Figures 11c and 11e. After normalization, the waveforms are quite consistent from one wavelet to the next when they are inferred from unweighted data, irrespective of the velocity model used (Figure 11e). In contrast, only the amplitude-normalized wavelets computed with the FWI velocity model show a good waveform repeatability when weighted data are used (Figure 11c).

3.6.2. Synthetic Seismograms and Ray Tracing Modeling

We use the source wavelets shown in Figure 11d to generate synthetic seismograms that can be readily compared with the recorded ones in a quantitative way. We also investigate the AVO variations of early arrivals (in a 400 ms window after the first breaks) to assess the agreement of both the absolute and relative amplitudes. To facilitate data fit assessment and phase interpretation, we complement the synthetic seismogram modeling with first-arrival and reflection ray tracing modeling following the approach of Operto *et al.* [2006]. For better readability, the recorded and modeled seismograms are shown after band-pass filtering between 3 and 7 Hz and are plotted without any trace balancing or gain control, which means that the true amplitudes are preserved.

In what follows, we limit our analysis to the most salient features of the data collected by OBS 46 (Figures 12 and 13). Other illustrative examples are shown in the supporting information for OBSs 17, 59, 74, and 93

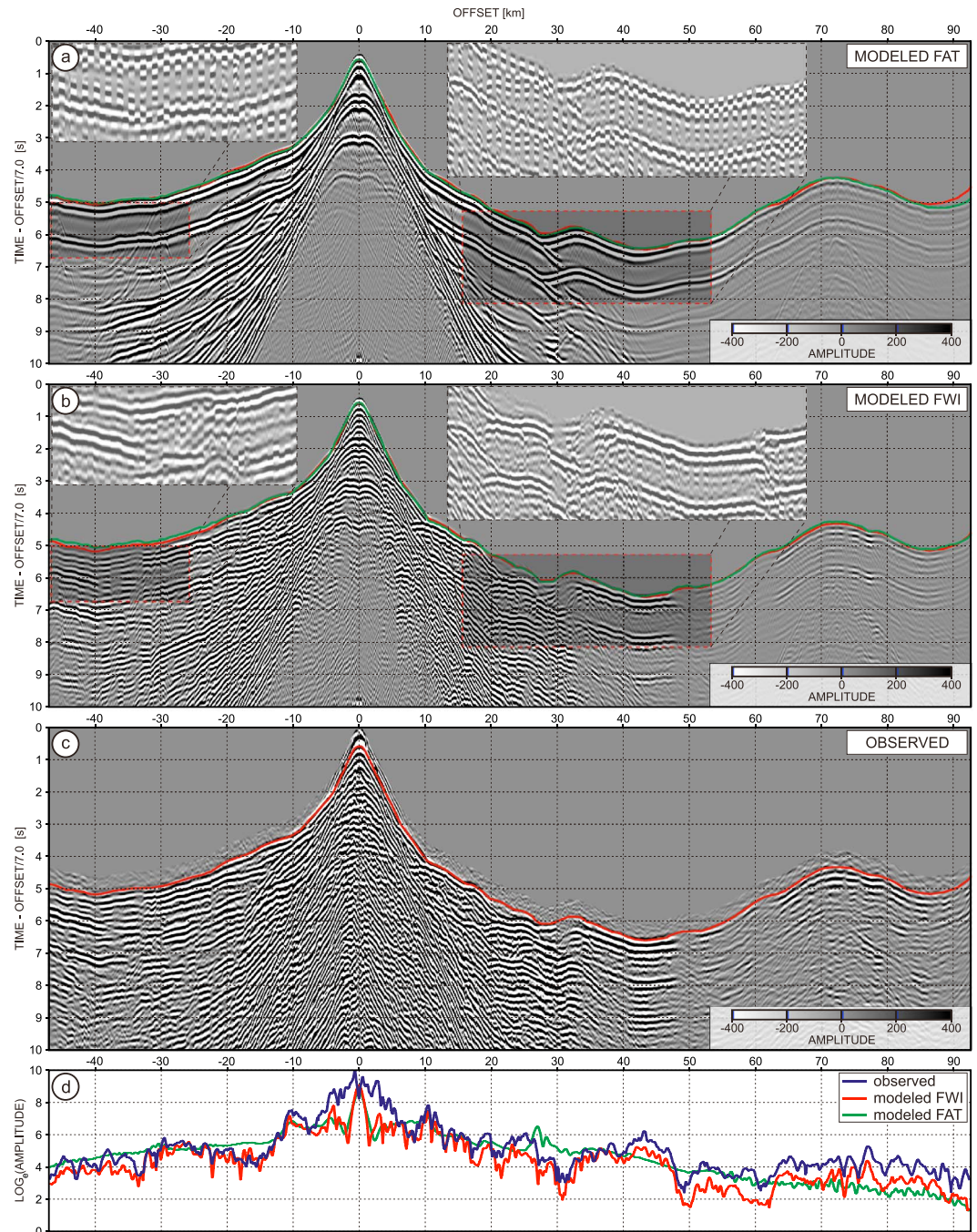


Figure 12. Comparison between the observed and modeled seismicograms for OBS 46. Synthetic seismicograms computed using the (a) FAT and (b) FWI models. The source wavelets, which have been convolved with the synthetic seismicograms, are shown in Figure 11d and 11e. (c) Observed seismicogram. In Figures 11a and 11b, the insets show zoom of the recorded and modeled wavefields where five consecutive synthetic traces are interleaved with the recorded counterparts. The red curves (Figures 11a–11c) represent the picked first-arrival traveltimes. The green curves represent the modeled first-arrival traveltimes computed using the FAT (Figure 11a) and FWI (Figure 11b) models. The true-amplitude seismicograms in Figures 11a–11c are plotted using the same amplitude scale. (d) Comparison of the AVO curves extracted from the seismicograms shown in Figures 11a–11c. The amplitudes are extracted from a 400 ms window after the first arrivals. The AVO curves are plotted on a logarithmic scale.

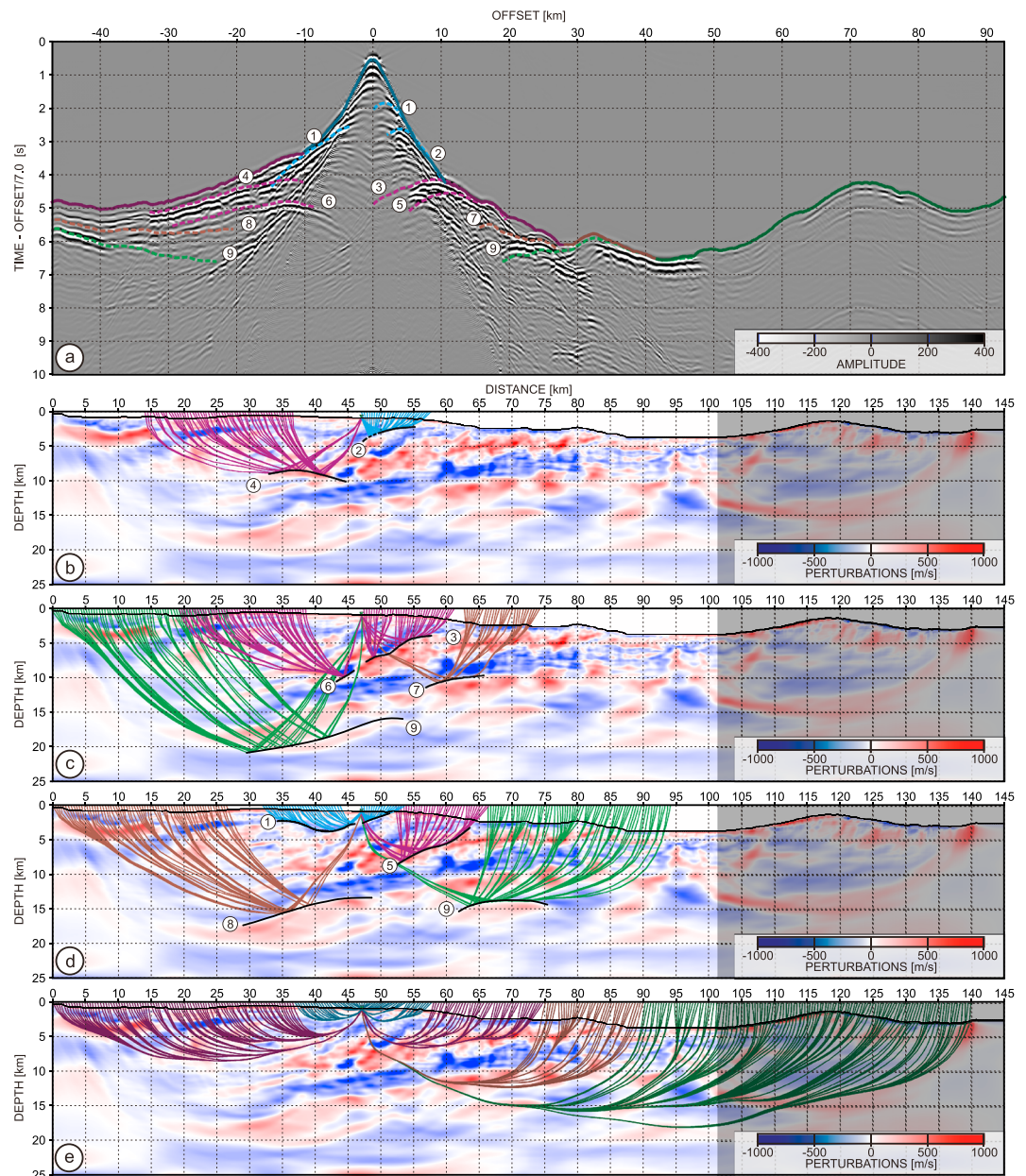


Figure 13. Ray tracing for OBS 46. (a) OBS gather with the reflection and first-arrival traveltimes overlaid. The numbers in the circles label the reflectors associated with each traveltimes curve. These reflectors have been interpreted in the detrended model as shown in Figure 13b–13d. The seismograms have been computed with an absorbing boundary conditions at the top of the model, and the direct water wave has been eliminated via F-K filtering for easier visualization of the arrivals that have been modeled via ray tracing. (b–d) Interpreted reflectors and corresponding reflection raypaths: (1) fore-arc basin reflector, (2) top of the Kodaiba thrust, (3 and 5) branches of the megasplay fault, (4) reflector in the backstop, (6) extension of interface 3 into the root of the Kodaiba thrust, (7) lower subducting oceanic crust, (8) Deeper Paleo-Zenisu Ridge, and (9) Moho of the subducting oceanic crust. (e) First-arrival raypaths.

(Figures S2, S4, S6, and S7 and Texts S2–S4 for synthetic seismogram modeling and Figures S3 and S5 for ray tracing). The positions of these OBSs are denoted in Figure 11a by the red stars.

OBS 46 is located above steeply dipping large thrusts of the Miocene wedge between the backstop and the Kodaiba thrust (Figure 9b). Despite the complexity of the wavefield and the acoustic approximation used for seismic modeling, the match between the recorded and modeled seismograms is quite good (Figures 12b and 12c). It is worth noting that the FWI successfully corrects small traveltimes inaccuracies in complex areas where several arrivals can interfere. This is highlighted by the mismatches between the picked and modeled

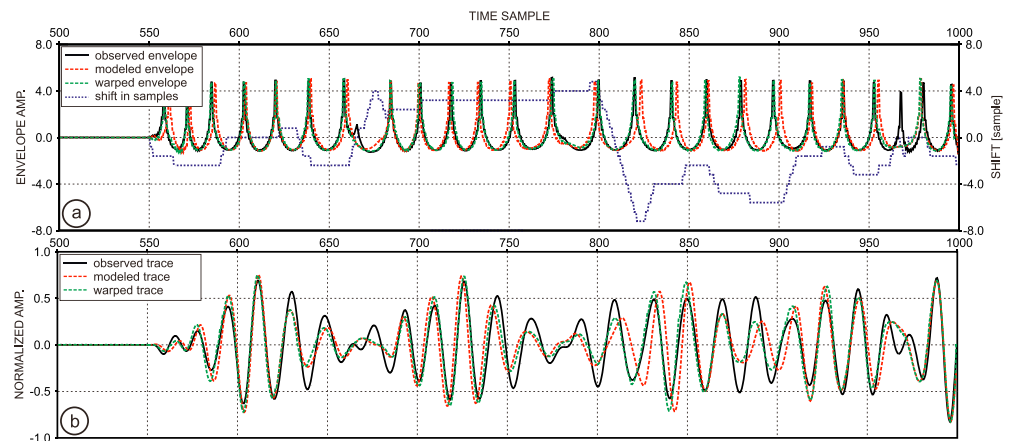


Figure 14. Illustration of DIW for a single trace. (a) Instantaneous phase envelopes. Solid black line: observed data. Dashed red line: modeled data. Dashed green line: modeled and warped data. Dotted blue line: approximated local time shifts. (b) Seismogram (AGC with a 2 s window applied) associated with Figure 14a. Solid black line: observed seismogram. Dashed red line: modeled seismogram. Dashed green line: modeled seismogram after warping with the shifts estimated in Figure 14a.

traveltimes computed using the FWI model for offsets between -47 and -25 km. Clearly, the first-arrival traveltimes computed using the FWI model show better agreement with the first breaks in the recorded seismograms than do the picked traveltimes. Such small traveltime adjustments performed during the FWI procedure were observed in most of the OBS gathers.

Between offsets of 20 and 50 km, there is a quite complex group of energetic arrivals arriving shortly after a quite weak first arrivals (Figures 12b and 12c), which probably corresponds to a mixture of wide-angle reflections from the reflective lower crust and the Moho of the subducting oceanic crust (Figures 13c and 13d, events 7 and 9). The right-hand side inset in Figures 12a and 12b is centered on these arrivals and shows a direct comparison between recorded seismograms and synthetic seismograms computed in the FAT and FWI models with an interleaved representation. It illustrates how the overestimated amplitudes of the first arrivals computed in the smooth FAT velocity model have progressively vanished without significant modification of their traveltimes at the same time as energetic wide-angle reflections built up as FWI was injecting short-scaled structures in the velocity model. This highlights the need of picking ambiguous first breaks with a model-driven approach to drive FAT toward a velocity model that will prevent cycle skipping of wide-angle reflections during subsequent FWI (see section 3.3.1).

The first-arrival raypaths (Figure 13e) clearly reveal the branching of refracted raypaths along a sequence of smaller thrusts in the active wedge at distances between 57 and 75 km in the detrended FWI model (Figures 9b and 9c). This branching probably contributes to the complex anatomy of the early wavefield in the 10–50 km offset range in Figures 12b and 12c. Moreover, we interpret the P_n wave in the seismograms beyond a 50 km offset (Figures 12b and 12c), which is consistent with the offset range for which the refracted rays from the Moho reach the surface in Figure 13e.

The AVO trend of the real data is well matched by the synthetic data. In particular, good agreement is found for the amplitude rise between offsets of 30 and 50 km that is associated with the postcritical reflections from the lower crust and the Moho (Figure 12d).

3.6.3. Dynamic Image Warping

Although our previous full-waveform modeling show that the synthetic seismograms fit the observed data reasonably well, the “eyeball norm” is unable to provide quantitative information about data fitting. To numerically express the misfit of “well-fitted” data or to evaluate the extent to which one data set “better fits” the observed data compared with another, we need a more quantitative and objective metric. The most obvious solution would be to subtract the recorded seismogram from the modeled one. However, since we do not reproduce the absolute amplitudes good enough, such a quality control would not be informative. Alternatively, cross-correlation techniques could be used to estimate local time shifts between two traces

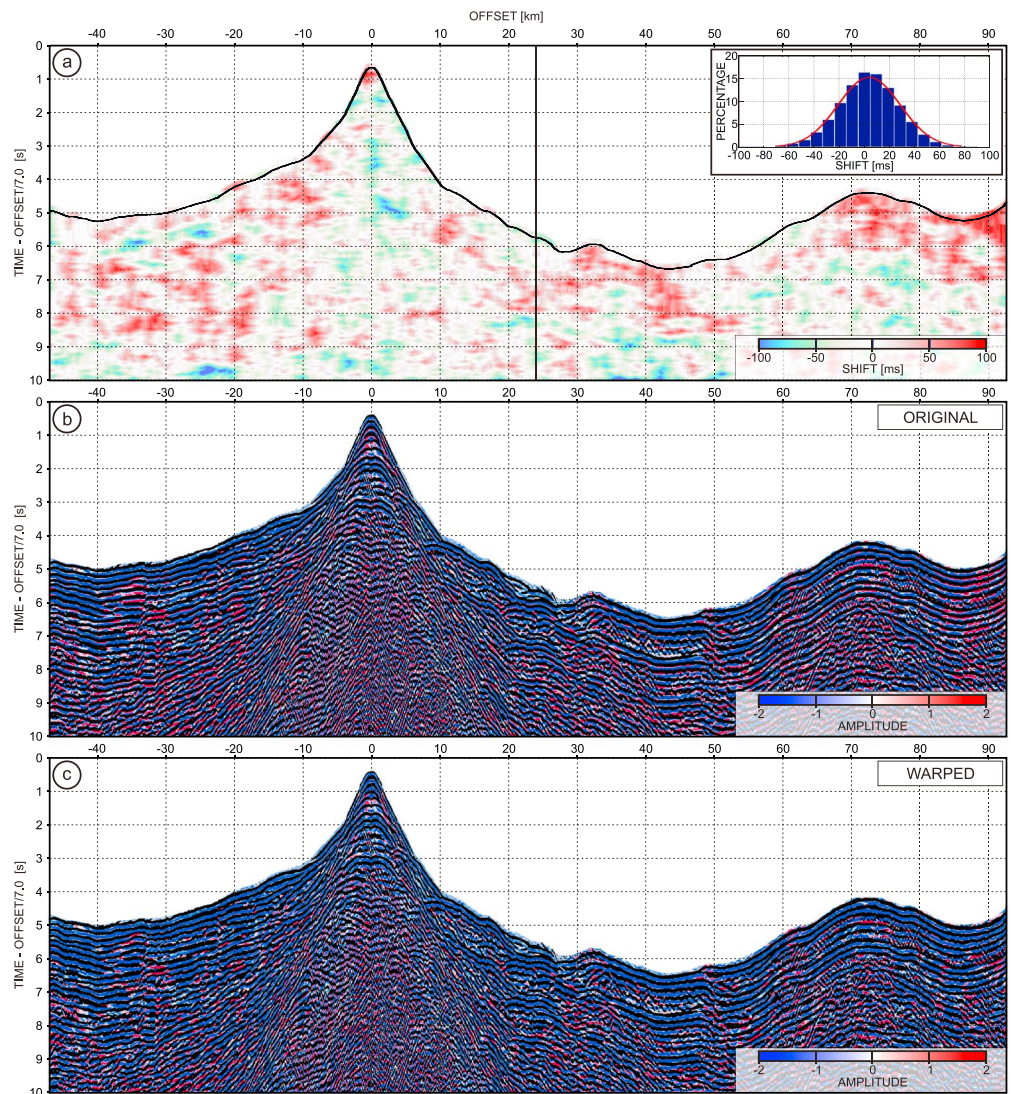


Figure 15. Illustration of DIW for OBS 46. (a) Map of the local shifts of all data samples, with their histogram and fitted distribution shown in the inset. The black curve represents the samples collected 100 ms after the first arrivals, and the black vertical line indicates the location of the trace represented in Figure 14. (b) Observed data displayed using a red-white-blue color scale overlaid with modeled data displayed using a variable-area display. The modeled data were computed using the final FWI model (Figure 12). The two sets of seismograms are in phase when the black color hides the red color. (c) Same as Figure 15b except that the modeled data have been warped using the shifts shown in Figure 15a.

in narrow running windows. Another method that allows one to measure the sample-to-sample mismatch between two given seismograms is Dynamic Image Warping (DIW) which has been shown to outperform local cross-correlation methods in presence of noise and rapidly varying shifts [Hale, 2013].

Since our synthetic data reproduce only the amplitude trends and not the real data amplitudes themselves, direct DIW of these seismograms would produce inaccurate results. To overcome this issue, we first extract the instantaneous phase and calculate its envelope for each real and synthetic trace. In Figure 14a, we plot a single trace after transformation (see the vertical black line at an offset of 24 km in Figure 15a for the trace location). The modeled trace (dashed red line) is locally shifted with respect to the observed one (solid black line). DIW provides a set of estimated local time shifts, which are expressed in terms of the number of samples (for a 10 ms sampling interval, Figure 14a, dotted blue line). Warping the synthetic trace by the estimated shifts produces a new trace (dashed green line) that shows much better agreement with the observed envelopes. The same comparison for the original seismic traces after automatic gain control with a 2 s window shows that

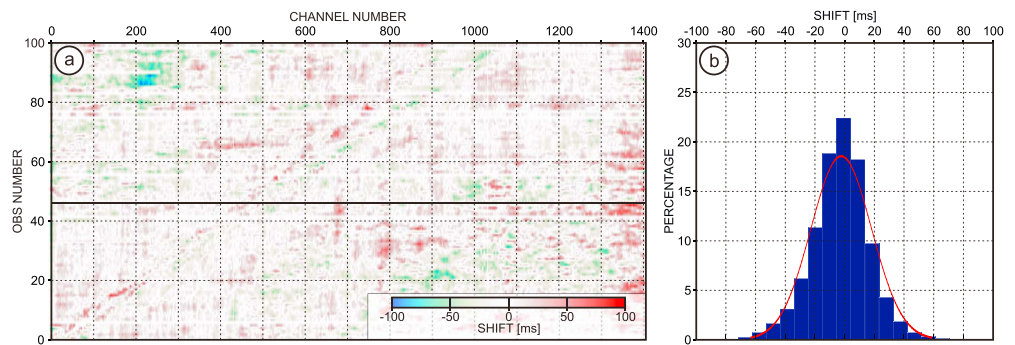


Figure 16. DIW shifts for all traces, plotted in the shot-receiver plane. These DIW shifts were estimated for the samples collected 100 ms after the first arrivals (solid black curve in Figure 15a). (b) Histogram of the shifts shown in Figure 16a and their best fit distribution.

the local shifts inferred from the instantaneous phase envelopes correctly adjust the synthetic trace toward the recorded one (Figure 14b). This proves that DIW can be used to obtain reliable estimates of the mismatches between observed and real data.

The application of DIW can be extended to a full OBS gather to account for the misalignments between not only time samples but also offset samples. The technical aspects of DIW are fully reviewed in Hale [2013]; here we simply explain how we set up the DIW procedure for our application. Our alignment error matrix is computed using instantaneous phase envelopes. Before, the data were band pass filtered to select frequencies between 2 and 8 Hz. Based on the alignment errors, we compute the distance array during the accumulation step. To stabilize the results, at this stage, we apply vertical and horizontal smoothing of the alignment errors using a strain limit of 1/3. This smoothing step allows any mismatches in time and offset to be taken into account. Finally, we estimate the time shifts while backtracking along the time dimension for each separate trace.

Figure 15a presents a time-offset map of the local time shifts estimated for OBS gather 46. The histogram of the shifts (see the inset), together with the fitted distribution curve, reveals a symmetric Gaussian distribution. However, the mean value is slightly higher than 0, being equal to 3.6 ms, while the RMS and standard deviation are 24.9 ms and 24.6 ms, respectively. These statistics quantify the good agreement between the observed and modeled gathers. Using the estimated time shifts, we can warp the synthetic seismograms to check their fit to the recorded data. Comparisons between the recorded and modeled data before and after warping are shown in Figures 15b and 15c. In both cases, the real data are plotted with a red-white-blue color scale, whereas the synthetic data are plotted using a black variable-area display. The two sets of seismograms are in phase if the black color of the variable-area display hides the red color in the plot. From careful inspection of the two panels, we can conclude that the warped synthetic seismograms (Figure 15c) fit the real data better than their unwarped counterparts (Figure 15b) do, as we observe less of the red color in the former panel. These improved alignments are especially visible in areas where the colors in the time-shift map are saturated (Figure 15a).

To build a full picture of the data fit, we estimate the time shifts for all OBS gathers and sort them according to source, receiver, and time sample indexes. This sorting generates a volume of time shifts, which can be sliced along arbitrary surfaces. Here we use a slicing scheme in which the surface created by the time samples is shifted by a given lag with respect to the first-arrival times. Figure 16a shows an example of such a map generated for a time lag of 100 ms from the first break. The iso-lag contour associated with OBS 46 is indicated by the black line in Figure 15a. Small, rarely appearing areas of more intense color may suggest problems with the precision of the data fit in the corresponding parts of the data set. The histogram of the time shifts and the distribution curve exhibit a symmetric Gaussian shape, with a mean value of -2.3 ms, an RMS of 20.5 ms, and a standard deviation of 20.4 ms (Figure 16b).

One may seek to update the velocity model by minimizing these residual time shifts, keeping in mind that some of these time shifts might result from 3-D propagation effects, which cannot be accounted for in our 2-D experiment [Ma and Hale, 2013].

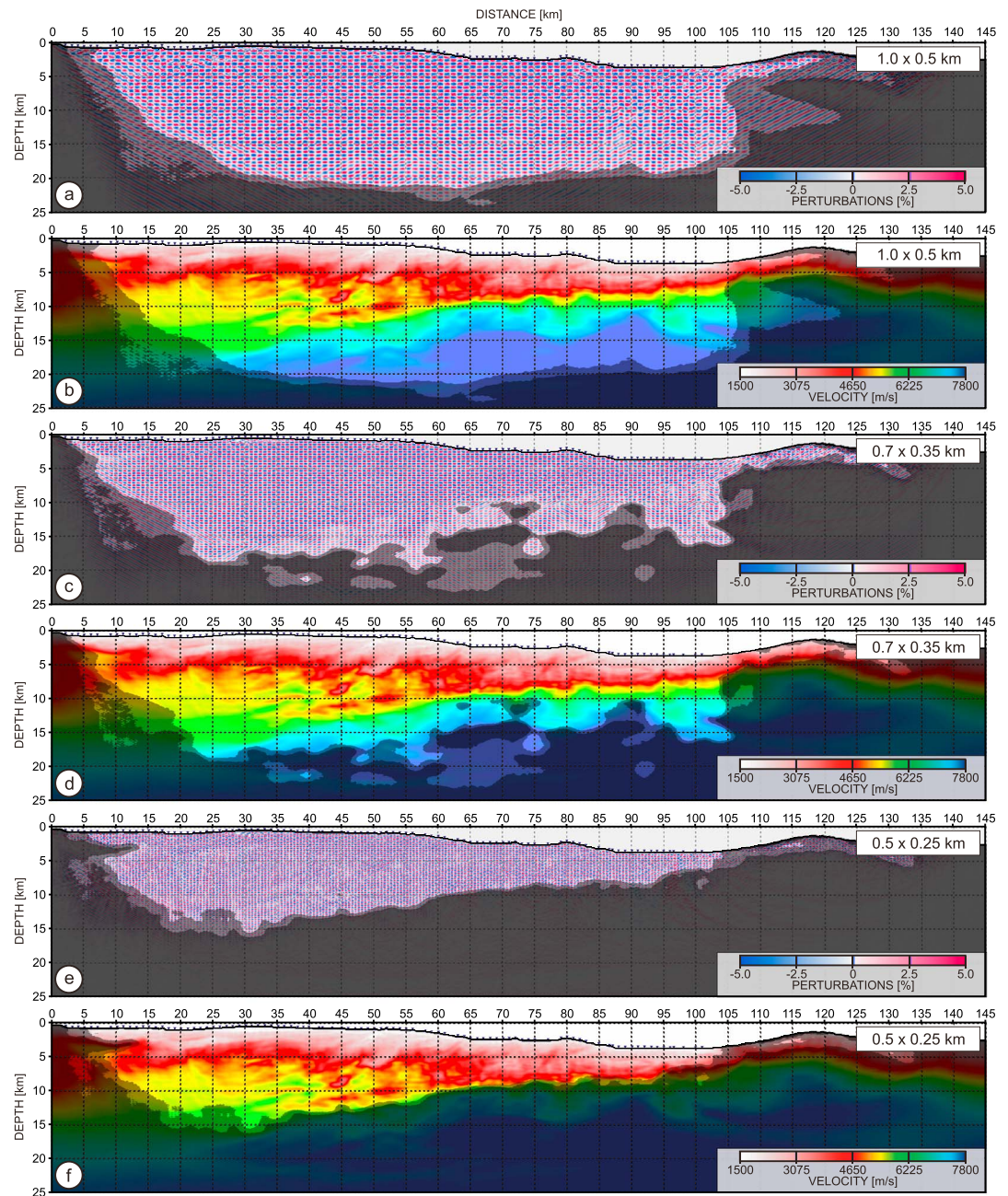


Figure 17. Results of the checkerboard tests: (a, b) Recovered 1.0 × 0.5 km checkerboard pattern with dark gray shading representing $R \leq 0.8$, light gray shading representing $0.8 < R \leq 0.9$, no shading for $R > 0.9$, and velocity model showing area of corresponding resolvability.(c, d) Same as Figures 17a and 17b but for 0.7 × 0.35 km checkerboard pattern. (e, f) Same as Figures 17a and 17b but for 0.5 × 0.25 km checkerboard pattern.

3.6.4. Resolution Tests

We perform a series of checkerboard tests to estimate the resolution of our final velocity model. The resolution of the FWI can be represented by the wave number spectrum covered by the frequency band and the acquisition geometry at some selected subsurface positions. The local wave number k imaged at a diffractor position, perpendicularly to the dip defined by the slowness vectors associated with the rays connecting the source and the receiver to the diffractor, is given by

$$k = \frac{2\omega}{c} \cos\left(\frac{\theta}{2}\right), \quad (10)$$

where ω is the angular frequency c is the local wave speed and θ is the scattering angle [Miller *et al.*, 1987, Figure 18].

Considering average velocities of 2.5 km/s in the shallow basins, 4.5 km/s in the accretionary prism, 6.5 km/s in the oceanic crust up to 8.5 km/s below Moho, a 8.5 Hz frequency, and assuming scattering angles ranging between 110° and 150°, we estimate from equation (10) the characteristic dimensions of the structures which can be resolved by FWI to be 0.20–0.45 km, 0.41–0.91 km, 0.62–1.36 km, and 0.82–1.82 km, respectively.

According to these estimations, we use checkerboards with the following sizes: 1.0 × 0.5 km, 0.7 × 0.35 km, and 0.5 × 0.25 km in the x - z direction. Due to the surface-to-surface geometry of the acquisition, we assume that the vertical resolution is twice as good as the horizontal one. After applying sinusoidal tapers to these rectangular patterns, we add them to the final FWI velocity model (Figure 9a) considering ±5% perturbations.

Synthetic data are generated using the same forward modeling as presented in section 2.1 with the same models for Q and ρ as for real data inversion. Data are subsequently inverted starting from the final FWI model, but with the inversion scheme limited to the frequencies from group 10 (see Table 1, Quasi-Progressive column). The source is kept constant during inversion. We increase number of iterations to 20 per time-damping/offset portion of data to compensate for the fact that we only use last frequency group. Gradient and Hessian preconditioning parameters are the same as those used for the FWI of the real data.

To quantitatively assess the retrieved checkerboards, we used the expression of the resolvability \mathbf{R} at each model cell given by

$$\mathbf{R} = \frac{\sum_{j=1}^M (P_{t_j} + P_{r_j})^2}{2 \sum_{j=1}^M (P_{t_j}^2 + P_{r_j}^2)}, \quad (11)$$

where P_{t_j} and P_{r_j} are the true and recovered velocity perturbations for the j th cell within the neighborhood operator consisting of M cells [Zelt, 1998]. In our case, we choose a neighborhood of 4 × 4 checkerboard patterns. Zelt [1998] empirically concluded that R above 0.7 can be considered as a well-resolved checkerboard pattern. In our tests, R higher than 0.8 was reflecting acceptable resolvability in terms of shape of the checkerboard, while good recovery of the perturbation amplitudes was observed for R above 0.9.

The results of the resolution tests are shown in Figure 17. For the 1.0 × 0.5 km checkerboard, we recover perturbations down to about 22 km (Figure 17a). Unshaded area in Figure 17b roughly corresponds to the first-arrival ray coverage that has been used to build the initial model by FAT, which makes us conclude that the recovery of this size of perturbations is mostly controlled by the acquisition spread.

For the 0.7 × 0.35 km checkerboard used in Kamei *et al.* [2013], the maximum depth of successful recovery varies between 18 and 14 km and roughly follows the geometry of the subducting oceanic crust (Figure 17c). This is consistent with our former estimation of the characteristic scales (0.62–1.36 km) that can be resolved for a wave speed of 6.5 km/s.

Last resolution test was conducted using 0.5 × 0.25 km checkerboard (Figure 17e). We obtained satisfactory recovery of perturbations between 14 and 10 km depth from the landward to the seaward side of the model. This subsurface area covers the structures overlying the subducting oceanic crust (sedimentary basins and wedges) as highlighted by the good correlation between the boundary indicating a resolvability lower than 0.9 and the top of the subducting oceanic crust (Figure 17f).

4. Discussion

We have proposed a new FAT + FWI workflow for deep crustal imaging from dense long-offset OBS data. In this framework, a key challenge is to avoid cycle skipping during FWI by improving the accuracy of the starting model built with FAT, generating low frequencies in the data and designing a multiscale layer-stripping strategy for use during FWI. Three main ingredients of our workflow allow us to largely achieve this goal: (i) early and extensive quality control of the starting model achieved through iterative refinement of the first breaks and FAT; (ii) Laplace-Fourier inversion with narrow time damping, which allows inversion to begin at a

frequency as low as 1.5 Hz; and (iii) a multiscale layer-stripping strategy based on nested frequency continuation from low values to higher ones, progressive relaxation of time damping, and progressive incorporation of long offsets.

4.1. Coping With the Limitations of FAT

For the initial model building, FAT is the most natural choice. The anatomy of the wavefield prompted us to limit the useful signal for traveltimes tomography to the first breaks. One drawback of FAT is the incomplete illumination of deep structures by diving waves. To address this issue, a layer-stripping strategy combined with time windowing (or damping) during FWI is useful to fit late wide-angle-reflected waves once the accuracy of the medium above the deep reflectors has been improved during the previous FWI stages. Another key aspect to consider is that the picked first-arrival traveltimes can be (and most probably are) affected by a certain degree of error due to the subjective nature of manual picking and the interpreter's ability to identify first breaks in the complex parts of the wavefield. This, in turn, may lead to a starting model that is not sufficiently accurate for FWI. We emphasize that careful quality control, including model-driven refinement of the picks during FAT if necessary, is essential for building a kinematically consistent long-wavelength model.

4.2. Coping With the Nonlinearity of Long-Offset Data FWI

In contrast to time-domain FWI, frequency-domain FWI does not allow for data-driven preconditioning strategies based on time windowing. Without this functionality, all arrivals with different resolution power are processed simultaneously, thereby increasing the nonlinearity of the inverse problem. Laplace-Fourier inversion overcomes this limitation by applying inversion on monochromatic data that have been damped in time. The benefit of this time damping is multifold. First, it allows arrivals to be empirically selected according to their aperture (or scattering) angles. Because the aperture angle and the frequency together control the resolution of FWI, progressive relaxation of this time damping introduces a second level of scale continuation in the FWI reconstruction in addition to the one provided by progressive inversion of higher frequencies. Second, it generates artificial low frequencies in the data through the convolution of the recorded data with a strongly oscillating sinc-like function in the frequency domain.

4.3. On the Reliability of the Acoustic Approximation

While accounting for elastic effects during waveform inversion is almost mandatory in earthquake seismology, the oil and gas exploration community has extensively demonstrated the relevance of the acoustic FWI in many marine environments [e.g., Plessix *et al.*, 2012, 2013; Warner *et al.*, 2013; Operto *et al.*, 2015; Benfield *et al.*, 2017]. The V_p models built by acoustic FWI are routinely assessed by checking the flatness of the common image gathers computed in the FWI models and direct comparison with in situ velocity measurements in the wells (check shots, vertical seismic profiling). Qualitatively, acoustic FWI of seismic data can be justified by the fact that a V_p perturbation in an elastic medium generates P wave scattering but no S wave scattering [Tarantola, 1986, Figure 4]. Conversely, S wave (V_s) velocity perturbations generate both P and S wave scatterings. However, the P wave scattering, which is that exploited by acoustic FWI, mainly impacts the AVO variations of the P - P reflections at critical and postcritical incidences. Acoustic FWI was shown to be relevant in geological environments involving soft elastic impedance contrasts such as North Sea. In such case, FWI is driven to the first order by kinematic effects, even if the AVO of the P wave reflections is not accurately modeled. On the opposite, hard impedance contrasts such as those present in North Oman generate significant mode conversions and amplitude effects that can easily drive the acoustic FWI toward a local minimum [Barnes and Charara, 2009; Marelli *et al.*, 2012; Pérez Solano *et al.*, 2013; Takougang and Calvert, 2013]. The fact that the velocity contrasts in the Tokai area are moderate due to the presence of sedimentary basins and wedges atop the crust warrants the use of the acoustic FWI in our study. Moreover, these moderate contrasts as well as the complex geology and rough interfaces have made the identification of the P - S converted phases in our data set challenging, suggesting that the amplitudes of these waves are weak. The level of waveform fit in terms of phase and amplitude strongly supports the relevance of these statements (Figure 12). To thoroughly assess the benefit of elastic FWI for our case study, viscoelastic inversion of the vertical and horizontal geophones could be performed to update the V_s parameter by fitting the AVO variations of the P wavefield and the P - S waveforms if these latter have a significant footprint in the data [Sears *et al.*, 2008, 2010; Prioux *et al.*, 2013]. Moreover, the joint update of the V_p and V_s models can be combined with that of ρ and Q , with the additional challenge to manage parameter cross talks between wave speeds and the second-order ρ and Q parameters.

4.4. On the Reliability of the Isotropic Approximation

Transverse isotropy (TI) with vertical (VTI) or tilted (TTI) symmetry axis is systematically considered at smaller scales for oil exploration applications since accurate depth registration of geological structures is crucial for drilling. The sensitivity of acoustic FWI to VTI anisotropy has been investigated by Plessix and Cao [2011]. A first conclusion is that the Thomsen's parameter δ [Thomsen, 1986] cannot be estimated from surface data because of the intrinsic ambiguity between this parameter and depth. Therefore, any erroneous prior knowledge of this parameter translates to vertical stretching of the subsurface structures. However, if the data set contains only diving waves, the inversion is only sensitive to the horizontal velocity (V_h). Conversely, if a long-offset acquisition records all types of waves (diving waves and precritical and postcritical reflections), the inversion can potentially retrieve two parameters such as the normal moveout (NMO) velocity (V_{nmo}) and V_h , using δ as a passive parameter. If an isotropic FWI is applied to such data set, the reconstructed isotropic velocities range between the V_{nmo} and V_h with stretching effects generated by the nonconsideration of δ and lateral artifacts that accommodate anisotropy. The anatomy of our data set, which is dominated by diving waves and post-critical reflections, suggests a setting intermediary between a pure transmission regime and a full-aperture experiment. Therefore, we expect the footprint of the anisotropy in our inversion to be reasonably weak and the updated velocities to be more closely related to V_h . Our adaptive tuning of the damping factor in the inversion preconditioner, ϵ in equation (7), might have further mitigated the footprint of anisotropy by strongly damping the contribution of the shallow subvertical wave paths at the benefit of the deep subhorizontal wave paths during the long-offset data inversion (see the discussion in section 3.4.4 and Figure 8). This strategy helped us to prevent shallow velocity artifacts during the inversion of the long-offset data. These artifacts might have been generated by the inconsistency between shallow wave speeds constrained by wave paths of different incidences during the different offset-dependent inversions. To thoroughly assess the footprint of the TI anisotropy in our study, multiparameter inversion for joint V_{nmo} and V_h updates setting $\delta = 0$ would be necessary.

4.5. The 2-D Versus 3-D Acquisitions

It is likely that our 2-D imaging has been hampered by 3-D effects due to the complex geology of the eastern Nankai Trough. This raises the issue of how to design the next generation of 3-D dense sea bottom acquisitions for deep crustal investigations [Morgan *et al.*, 2013, 2016]. Although various advances have been achieved in seismic instrumentation and the ability of the academic community to perform OBS acquisitions is increasing, the gathering of large OBS pools remains challenging. Moreover, although available computer resources enable imaging of moderate size targets by 3-D time-domain or frequency-domain FWI [Warner *et al.*, 2013; Operto *et al.*, 2015], applications on large computational meshes required by deep crustal investigations remain a computational challenge keeping in mind that the computational overhead generated by 3-D problems is 2 orders of magnitude greater than that in the 2-D case. To evaluate the feasibility of 3-D OBS surveys for FWI applications, an assessment of the OBS spacing required for reliable imaging is necessary. A first attempt to answer this question based on our 2-D experiment is presented in the supporting information (Text S5 and Figure S8).

4.6. The Role of the FWI Misfit Function: Amplitudes and Automation

One disadvantage of our workflow is the requirement for human intervention to keep cycle skipping under control. Second, the use of the acoustic approximation calls into question the role of the amplitudes in our inversion. In this study, we have used a conventional difference-based L_2 misfit function. It is widely acknowledged that this misfit function may potentially have many local minima in representative FWI applications and is inherently influenced by the amplitudes. We did not attempt to test phase-only inversion using a logarithmic misfit function, as done by Kamei *et al.* [2014]. However, we emphasize that our modeled data fairly well reproduced the sharp AVO variations observed in the recorded gathers (Figure 12e and supporting information Figures S2e, S4e, S6e, and S7e). Many other misfit functions with better convexity properties have been published in the literature over the past decade in the context of exploration geophysics, such as cross-correlation-based misfit functions [van Leeuwen and Mulder, 2010], matching-filter-based misfit functions [Warner and Guasch, 2016; Zhu and Fomel, 2016], or optimal transport misfit functions [Métivier *et al.*, 2016]. Such an alternative misfit function could be used to update the FAT model and hence to provide an improved initial model for conventional FWI. However, the robustness of these functions for such long-offset data dominated by diving waves and supercritical reflections has yet to be demonstrated. Other promising approaches rely on extending the model space by means of wavefield reconstruction [van Leeuwen and Herrmann, 2013] or on FWI with traveltimes constraints [Treister and Haber, 2016].

5. Conclusions

We have revisited the TKY-21 OBS data set from the eastern Nankai Trough with the aim of developing a practical frequency-domain FWI workflow for the construction of crustal-scale velocity models. The main challenge was to minimize the occurrence of cycle skipping, considering the large number of propagated wavelengths acquired with ultralong offsets. To achieve this goal, we integrated into the FWI workflow three data-driven continuation strategies based on the attributes of frequency, traveltime, and offset. The hierarchical management of the frequencies and traveltimes (or, equivalently, aperture angles) defines a multiscale approach, whereas the management of the offsets defines a layer-stripping strategy. Such offset continuation is not frequently applied in FWI, and care should be taken when combining it with time damping. Here this approach allows us to progressively incorporate deeply propagating arrivals recorded at long offsets into the inversion once the accuracy of the upper part of the crustal model has been improved through the inversion of shorter-offset data. A second key ingredient is the generation of low frequencies (down to 1.5 Hz) by means of the time damping of the seismograms. We have concluded that reliable FWI results can be obtained in such a complex environment only if very careful quality control of the first-break picking and FAT is performed. This step, however, requires rather tedious human intervention. The development of more convex FWI misfit functions should allow the need for an accurate FAT velocity model to be relaxed in the future.

We have also investigated a complete workflow for the quality control of FWI results which includes full-waveform seismic modeling, source wavelet estimation, refraction, and reflection ray tracing modeling and time-mismatch estimation via DIW. This quality control workflow enables a precise and quantitative assessment of the traveltime and amplitude agreement. The data fit achieved here is surprisingly good considering the wave physics approximation used and reflects the significant amount of high-resolution geological information contained in the velocity model. A more thorough geodynamic interpretation of this velocity model is currently the objective of ongoing work.

Acknowledgments

This work has been partially funded by (i) the Polish National Science Center, grant 2011/03/D/ST10/05128; (ii) an internal IG PAS grant 1A/IGF PAN/2016; (iii) the IG PAS, Leading National Research Center (KNOW) in Earth Sciences for the years 2014–2018; and (iv) statutory activities 3841/E-41/S/2016 of the Ministry of Science and Higher Education of Poland. We thank R. Brossier (ISTerre) and L. Métivier (ISTerre/LJK) for fruitful discussions and for providing us with the TOY2DAC FWI code developed within the framework of the SEISCOPE Consortium (available at <https://seiscope2.osug.fr>) and Wei Zhou (ISTerre) for providing us with the DIW code. We thank JAMSTEC for providing us with the SFJ-OBS data. This study was granted access to the HPC resources of SIGAMM hosted by the Observatoire de la Côte d'Azur. We would like to thank the Editor Uri ten Brink, Joanna Morgan, and two anonymous reviewers for their very constructive comments.

References

- Adamczyk, A., M. Malinowski, and A. Górszczyk (2015), Full-waveform inversion of conventional Vibroseis data recorded along a regional profile from SE Poland, *Geophys. J. Int.*, *203*(1), 351–365.
- Amundsen, L., L. T. Ikelle, and L. E. Berg (2001), Multidimensional signature deconvolution and free-surface multiple elimination of marine multicomponent ocean-bottom seismic data, *Geophysics*, *66*(5), 1594–1604.
- Ando, M. (1975a), Source mechanisms and tectonic significance of historical earthquakes along the Nankai Trough, Japan, *Tectonophysics*, *27*, 119–140.
- Ando, M. (1975b), Possibility of a major earthquake in the Tokai district, Japan and its pre-estimated seismotectonic effects, *Tectonophysics*, *27*, 69–85.
- Barnes, C., and M. Charara (2009), The domain of applicability of acoustic full-waveform inversion for marine seismic data, *Geophysics*, *74*(6), WCC91–WCC103, doi:10.1190/1.3250269.
- Benfield, N., V. Rambaran, J. Dowlath, T. Sinclair, M. Evans, J. Richardson, A. Ratcliffe, and A. Irving (2017), Extracting geologic information directly from high-resolution full-waveform inversion velocity models—A case study from offshore Trinidad, *Leading Edge*, *36*(1), 67–74.
- Bleibinhaus, F., R. W. Lester, and J. A. Hole (2009), Applying waveform inversion to wide-angle seismic surveys, *Tectonophysics*, *472*, 238–248.
- Brenders, A. J., and R. G. Pratt (2007a), Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model, *Geophys. J. Int.*, *168*, 133–151.
- Brenders, A. J., and R. G. Pratt (2007b), Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data, *Geophys. J. Int.*, *168*, 152–170.
- Brossier, R., S. Operto, and J. Virieux (2009), Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, *74*(6), WCC105–WCC118, doi:10.1190/1.3215771.
- Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent (1995), Multiscale seismic waveform inversion, *Geophysics*, *60*(5), 1457–1473.
- Byrd, R. H., P. Lu, and J. Nocedal (1995), A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Stat. Comput.*, *16*, 1190–1208.
- Candès, E., L. Demanet, D. Donoho, and L. Ying (2006), Fast discrete curvelet transforms, *Multiscale Model. Simul.*, *5*(3), 861–899.
- Dessa, J. X., S. Operto, S. Kodaira, A. Nakanishi, G. Pascal, K. Uhira, and Y. Kaneda (2004a), Deep seismic imaging of the eastern Nankai Trough (Japan) from multifold ocean bottom seismometer data by combined traveltime tomography and prestack depth migration, *J. Geophys. Res.*, *109*, B02111, doi:10.1029/2003JB002689.
- Dessa, J. X., S. Operto, S. Kodaira, A. Nakanishi, G. Pascal, J. Virieux, and Y. Kaneda (2004b), Multiscale seismic imaging of the eastern Nankai Trough by full waveform inversion, *Geophys. Res. Lett.*, *31*, L18606, doi:10.1029/2004GL020453.
- Duff, I. S., A. M. Erisman, and J. K. Reid (1986), *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, U. K.
- Gardner, G. H. F., L. W. Gardner, and A. R. Gregory (1974), Formation velocity and density—The diagnostic basics for stratigraphic traps, *Geophysics*, *39*, 770–780.
- Górszczyk, A., A. Adamczyk, and M. Malinowski (2014), Application of curvelet denoising to 2D and 3D seismic data—Practical considerations, *J. Appl. Geophys.*, *105*, 78–94.
- Ha, W., and C. Shin (2012), Proof of the existence of both zero- and low-frequency information in a damped wavefield, *J. Appl. Geophys.*, *83*, 96–99.
- Hale, D. (2013), Dynamic warping of seismic images, *Geophysics*, *78*(2), S105–S115.
- Henry, P., S. Operto, S. Lallemand, J.-X. Dessa, V. Martin, M. Noble, H. Tokuyama, S. Kodaira, S. Kuramoto, and M. Kinoshita (2014), Deformation processes and earthquakes in Nankai, in *2004 GeoHazards Workshop Abstract Book*, ENSMP, Shizuoka, Japan.

- Hicks, G. J. (2002), Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions, *Geophysics*, *67*, 156–166.
- Hustedt, B., S. Operto, and J. Virieux (2004), Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling, *Geophys. J. Int.*, *157*, 1269–1296.
- Jo, C. H., C. Shin, and J. H. Suh (1996), An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator, *Geophysics*, *61*, 529–537.
- Kamei, R., and R. G. Pratt (2013), Inversion strategies for visco-acoustic waveform inversion, *Geophys. J. Int.*, *194*, 859–894.
- Kamei, R., R. G. Pratt, and T. Tsuji (2012), Waveform tomography imaging of a megasplay fault system in the seismogenic Nankai subduction zone, *Earth Planet. Sci. Lett.*, *317–318*, 343–353.
- Kamei, R., R. G. Pratt, and T. Tsuji (2013), On acoustic waveform tomography of wide-angle obs data—Strategies for preconditioning and inversion, *Geophys. J. Int.*, *192*, 1250–1280.
- Kamei, R., R. G. Pratt, and T. Tsuji (2014), Misfit functionals in Laplace-Fourier domain waveform inversion, with application to wide-angle ocean bottom seismograph data, *Geophys. Prospect.*, *62*, 1054–1074.
- Kamei, R., T. Miyoshi, R. G. Pratt, M. Takahashi, and S. Masaya (2015), Application of waveform tomography to a crooked-line 2D land seismic data set, *Geophysics*, *80*(5), B115–B129.
- Kodaira, S., A. Nakanishi, J. O. Park, A. Ito, T. Tsuru, and Y. Kaneda (2003), Cyclic ridge subduction at an inter-plate locked zone off Central Japan, *Geophys. Res. Lett.*, *30*(6), 1339, doi:10.1029/2002GL016595.
- Kurzmann, A., A. Przybindowska, D. Kohn, and T. Bohlen (2013), Acoustic full waveform tomography in the presence of attenuation: A sensitivity analysis, *Geophys. J. Int.*, *195*(2), 985–1000.
- Lailly, P. (1984), The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering*, edited by P. Lailly, pp. 206–220, Soc. Ind. appl. Math, SIAM, Philadelphia.
- Lallemant, S., P. Henry, V. Martin, G. Pascal, M. Noble, X. Le Pichon, and H. Tokuyama (2001), Subduction of oceanic highs and splay faults in eastern Nankai subduction zone, *Eos Trans. AGU*, *81*(48), T52E–12.
- Le Pichon, X., S. Lallemant, H. Tokuyama, F. Thoué, P. Huchon, and P. Henry (1996), Structure and evolution of the backstop in the eastern Nankai Trough area (Japan): Implications for the soon-to-come Tokai earthquake, *Island Arc*, *5*, 440–454.
- Lester, R., H. J. A. V. Avandok, K. McIntosh, L. Lavie, C.-S. Liu, T. K. Wang, and F. Wu (2014), Rifting and magmatism in the northeastern South China Sea from wide-angle tomography and seismic reflection imaging, *J. Geophys. Res.*, *119*, 2305–2323, doi:10.1002/2013JB010639.
- Ma, Y., and D. Hale (2013), Wave-equation reflection travelt ime inversion with dynamic warping and full waveform inversion, *Geophysics*, *78*(6), R223–R233.
- Malinowski, M. (2013), Models of the Earth's crust from controlled-source seismology—Where we stand and where we go?, *Acta Geophys.*, *61*(6), 1437–1456.
- Malinowski, M., S. Operto, and A. Ribodetti (2011), High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full waveform inversion, *Geophys. J. Int.*, *186*(3), 1179–1204.
- Marelli, S., H. Maurer, and E. Manukyan (2012), Validity of the acoustic approximation in full-waveform seismic crosshole tomography, *Geophysics*, *77*(3), R129–R139.
- Martin, V. (2003), Structure et tectonique du prisme d'accrétion de Nankai dans la zone de Tokai par imagerie sismique en trois dimensions, PhD thesis, Univ. de Paris Sud.
- Mazzotti, S., P. Henry, X. L. Pichon, and T. Sagiya (1999), Strain partitioning in the zone of transition from Nankai subduction to Izu-Bonin collision (Central Japan): Implications for an extensional tear within the subducting slab, *Earth Planet. Sci. Lett.*, *172*, 1–10.
- Mazzotti, S., S. Lallemant, P. Henry, X. L. Pichon, H. Tokuyama, and N. Takahashi (2002), Intraplate shortening and underthrusting of a large basement ridge in the eastern Nankai subduction zone, *Mar. Geol.*, *187*, 63–88.
- Métivier, L., and R. Brossier (2016), The seiscop optimization toolbox: A large-scale nonlinear optimization library based on reverse communication, *Geophysics*, *81*(2), F11–F25.
- Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux (2014), Full waveform inversion and the truncated Newton method: Quantitative imaging of complex subsurface structures, *Geophys. Prospect.*, *62*, 1353–1375, doi:10.1111/1365-2478.12136.
- Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux (2016), Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion, *Geophys. J. Int.*, *205*(1), 345–377.
- Miller, D., M. Oristaglio, and G. Beylkin (1987), A new slant on seismic imaging: Migration and integral geometry, *Geophysics*, *52*(7), 943–964.
- Morgan, J., M. Warner, R. Bell, J. Ashley, D. Barnes, R. Little, K. Roelle, and C. Jones (2013), Next-generation seismic experiments: Wide-angle, multi-azimuth, three-dimensional, full-waveform inversion, *Geophys. J. Int.*, *195*(3), 1657–1678.
- Morgan, J., M. Warner, G. Arnoux, E. Hooft, D. Toomey, B. VanderBeek, and W. Wilcock (2016), Next-generation seismic experiments—II: Wide-angle, multi-azimuth, 3-D, full-waveform inversion of sparse field data, *Geophys. J. Int.*, *204*(2), 1342–1363.
- MUMPS team (2015), MUMPS—Multifrontal Massively Parallel Solver users' guide—version 5.0.1 (July 23, 2015), ENSEEIHT-ENS Lyon. Available at: <http://www.enseeiht.fr/apo/MUMPS/> or <http://graal.ens-lyon.fr/MUMPS>.
- Nocedal, J., and S. J. Wright (2006), *Numerical Optimization* 2nd edn., Springer, New York.
- Operto, S., J. Virieux, J. X. Dessa, and G. Pascal (2006), Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: Application to the eastern Nankai Trough, *J. Geophys. Res.*, *111*, B09306, doi:10.1029/2005JB003835.
- Operto, S., R. Brossier, Y. Gholami, L. Métivier, V. Prieux, A. Ribodetti, and J. Virieux (2013), A guided tour of multiparameter full waveform inversion for multicomponent data: From theory to practice, *Leading Edge*, *9*, 1040–1054.
- Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux (2015), Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: Application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophys. J. Int.*, *202*(2), 1362–1391.
- Park, J. O., G. Fujie, L. Wijerathne, T. Hori, S. Kodaira, Y. Fukao, G. F. Moore, N. L. Bangs, S. Kuramoto, and A. Taira (2010), A low-velocity zone with weak reflectivity along the Nankai subduction zone, *Geology*, *38*(3), 283–286.
- Pérez Solano, C. A., A. Stopin, and R.-E. Plessix (2013), Synthetic study of elastic effects on acoustic full waveform inversion, in *Proceedings of the 75th EAGE Conference and Exhibition Incorporating SPE EUROPEC 2014*, EAGE.
- Plessix, R. E. (2006), A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, *167*(2), 495–503.
- Plessix, R. E. (2007), A Helmholtz iterative solver for 3D seismic-imaging problems, *Geophysics*, *72*(5), SM185–SM194.
- Plessix, R. E., and Q. Cao (2011), A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium, *Geophys. J. Int.*, *185*, 539–556, doi:10.1111/j.1365-246X.2011.04957.x.
- Plessix, R.-E., G. Baeten, J. W. de Maag, and F. Ten Kroode (2012), Full waveform inversion and distance separated simultaneous sweeping: A study with a land seismic data set, *Geophys. Prospect.*, *60*, 733–747.

- Plessix, R. E., P. Milcik, H. Rynja, A. Stopin, K. Matson, and S. Abri (2013), Multiparameter full-waveform inversion: Marine and land examples, *Leading Edge*, 32(9), 1030–1038.
- Pratt, R. G. (1999), Seismic waveform inversion in the frequency domain. Part I: Theory and verification in a physical scale model, *Geophysics*, 64, 888–901.
- Pratt, R. G. (2008), Waveform tomography—Successes, cautionary tales, and future directions, paper presented at the 70th Annual EAGE Conference and Exhibition, pp. WO11, Full-Waveform Inversion: Current status and perspectives, Roma.
- Pratt, R. G., and M. H. Worthington (1990), Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method, *Geophys. Prospect.*, 38, 287–310.
- Pratt, R. G., Z. M. Song, P. R. Williamson, and M. Warner (1996), Two-dimensional velocity models from wide-angle seismic data by waveform inversion, *Geophys. J. Int.*, 124, 323–340.
- Pratt, R. G., C. Shin, and G. J. Hicks (1998), Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, 133, 341–362.
- Prieux, V., R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. Barkved, and J. Kommedal (2011), On the footprint of anisotropy on isotropic full waveform inversion: The Valhall case study, *Geophys. J. Int.*, 187, 1495–1515, doi:10.1111/j.1365-246X.2011.05209.x.
- Prieux, V., R. Brossier, S. Operto, and J. Virieux (2013), Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 2: Imaging compressional and shear-wave velocities, *Geophys. J. Int.*, 194(3), 1665–1681, doi:10.1093/gji/ggt178.
- Ravaut, C., S. Operto, L. Improt, J. Virieux, A. Herrero, and P. dell'Aversana (2004), Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt, *Geophys. J. Int.*, 159, 1032–1056.
- Rondenay, S., M. G. Bostock, and K. M. Fischer (2005), Multichannel inversion of scattered teleseismic body waves: Practical considerations and applicability, in *Seismic Earth: Array Analysis of Broadband Seismograms*, *Geophys. Monogr. Ser.*, vol. 157, edited by G. N. Alan Levander, pp. 187–203, AGU, Washington, D. C.
- Sears, T., S. Singh, and P. Barton (2008), Elastic full waveform inversion of multi-component OBC seismic data, *Geophys. Prospect.*, 56(6), 843–862.
- Sears, T. J., P. J. Barton, and S. C. Singh (2010), Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to alba field, U. K. North Sea, *Geophysics*, 75(6), R109–R119.
- Seno, T., S. Stein, and A. E. Gripp (1993), A model for the motion of the Philippine Sea plate consistent with NUVEL-1 and geologic data, *J. Geophys. Res.*, 98, 17,941–17,948.
- Shah, N., M. Warner, T. Nangoo, A. Umpleby, I. Stekl, J. Morgan, and L. Guasch (2012), Quality assured full-waveform inversion: Ensuring starting model adequacy, paper presented at the 82nd Annual International Meeting, SEG Technical Program Expanded Abstracts, pp. 1–5, Soc. of Explor. Geophys., Tulsa, Okla., doi:10.1190/segam2012-1228.1.
- Shin, C., and Y. H. Cha (2009), Waveform inversion in the Laplace-Fourier domain, *Geophys. J. Int.*, 177, 1067–1079.
- Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin (2001), Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion, *Geophysics*, 66(6), 1856–1863.
- Shipp, R. M., and S. C. Singh (2002), Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.*, 151, 325–344.
- Shragge, J., M. G. Bostock, and S. Rondenay (2001), Multiparameter two-dimensional inversion of scattered teleseismic body waves: 2. Numerical examples, *J. Geophys. Res.*, 106(12), 30,783–30,793.
- Sirgue, L. (2006), The importance of low frequency and large offset in waveform inversion, in *Proceeding of the 68th EAGE Conference and Exhibition*, pp. A037, EAGE, Vienna.
- Sirgue, L., and R. G. Pratt (2004), Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics*, 69(1), 231–248.
- Song, Z., P. Williamson, and G. Pratt (1995), Frequency-domain acoustic-wave modeling and inversion of crosshole data. Part 2: Inversion method, synthetic experiments and real-data results, *Geophysics*, 60(3), 786–809.
- Takougang, E. M. T., and A. J. Calvert (2011), Application of waveform tomography to marine seismic reflection data from the Queen Charlotte basin of western Canada, *Geophysics*, 76(2), B55–B70.
- Takougang, E. M. T., and A. J. Calvert (2013), Seismic waveform tomography across the Seattle Fault Zone in Puget sound: Resolution analysis and effectiveness of visco-acoustic inversion of viscoelastic data, *Geophys. J. Int.*, 193(2), 763–787.
- Tarantola, A. (1984), Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49(8), 1259–1266.
- Tarantola, A. (1986), A strategy for non linear inversion of seismic reflection data, *Geophysics*, 51(10), 1893–1903.
- Thomsen, L. A. (1986), Weak elastic anisotropy, *Geophysics*, 51, 1954–1966.
- Toksöz, M. N., and D. H. Johnston (1981), *Geophysics Reprint Series, No. 2: Seismic Wave Attenuation*, Soc. of Explor. Geophys., Tulsa, Okla.
- Treister, E., and E. Haber (2016), 78th EAGE Conference and Exhibition 2016, Full Waveform Inversion IV, EAGE, doi:10.3997/2214-4609.201601542.
- Tsuji, T., J.-O. Park, G. Moore, S. Kodaira, Y. Fukao, S. Kuramoto, and N. Bangs (2009), Intraoceanic thrusts in the Nankai Trough off the Kii Peninsula: Implications for intraplate earthquakes, *Geophys. Res. Lett.*, 36, L06303, doi:10.1029/2008GL036974.
- Tsuji, T., S. Kodaira, J. Ashi, and J.-O. Park (2013), Widely distributed thrust and strike-slip faults within subducting oceanic crust in the Nankai Trough off the Kii Peninsula, Japan, *Tectonophysics*, 600, 52–62.
- van Leeuwen, T., and F. J. Herrmann (2013), Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, 195(1), 661–667, doi:10.1093/gji/ggt258.
- van Leeuwen, T., and W. A. Mulder (2010), A correlation-based misfit criterion for wave-equation traveltome tomography, *Geophys. J. Int.*, 182(3), 1383–1394.
- Vigh, D., and E. W. Starr (2008), Comparisons for waveform inversion, time domain or frequency domain?, in *SEG Technical Program Expanded Abstracts*, pp. 1890–1894, Soc. of Explor. Geophys., Las Vegas, Nev.
- Virieux, J., and S. Operto (2009), An overview of full waveform inversion in exploration geophysics, *Geophysics*, 74(6), WCC1–WCC26.
- Virieux, J., S. Operto, H. Ben Hadj Ali, R. Brossier, V. Etienne, F. Sourbier, L. Giraud, and A. Haidar (2009), Seismic wave modeling for seismic imaging, *Leading Edge*, 28(5), 538–544.
- Wang, Y., and Y. Rao (2009), Reflection seismic waveform tomography, *J. Geophys. Res.*, 114, 1978–2012, doi:10.1029/2008JB005916.
- Warner, M., and L. Guasch (2016), Adaptive waveform inversion: Theory, *Geophysics*, 81(6), R429–R445.
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand (2013), Anisotropic 3D full-waveform inversion, *Geophysics*, 78(2), R59–R80.
- Williamson, P. (1991), A guide to the limits of resolution imposed by scattering in ray tomography, *Geophysics*, 56, 202–207.
- Wu, R. S., and M. N. Toksöz (1987), Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, 52, 11–25.

- Zelt, C., and P. J. Barton (1998), Three-dimensional seismic refraction tomography: A comparison of two methods applied to data from the Faeroe basin, *J. Geophys. Res.*, *103*(B4), 7187–7210.
- Zelt, C., and R. B. Smith (1992), Seismic travelt ime inversion for 2-D crustal velocity structure, *Geophys. J. Int.*, *108*, 16–34.
- Zelt, C. A. (1998), Lateral velocity resolution from three-dimensional seismic refraction data, *Geophys. J. Int.*, *135*(3), 1101–1112.
- Zhu, H., and S. Fomel (2016), Building good starting models for full-waveform inversion using adaptive matching filtering misfit, *Geophysics*, *81*(5), U61–U72, doi:10.1190/geo2015-0596.1.