GEOPHYSICS, VOL. 78, NO. 2 (MARCH-APRIL 2013); P. R25–R36, 12 FIGS.
10.1190/GEO2012-0104.1

Regularized seismic full waveform inversion with prior model information

Amir Asnaashari[1], Romain Brossier[1], Stéphane Garambois[1], François Audebert[2], Pierre Thore[3], and Jean Virieux[1]

ABSTRACT

Full waveform inversion (FWI) delivers high-resolution quantitative images and is a promising technique to obtain macroscale physical property model of the subsurface. In most geophysical applications, prior information, such as that collected in wells, is available and should be used to increase the image reliability. For this, we propose to introduce three terms in the definition of the FWI misfit function: the data misfit itself, the first-order Tikhonov regularization term acting as a smoothing operator, and a prior model norm term. This last term is the way to smoothly introduce prior information into the FWI workflow. On a selected target of the Marmousi synthetic example, significant improvement was obtained when using the prior model term for noise-free and noisy synthetic data. The prior model term may significantly reduce the inversion sensitivity to incorrect initial conditions. The limited range of spatial wavenumber sampling by the acquisition may be compensated with the prior model information, for multiple-free and multiple-contaminated data. Prior and initial models play different roles in the inversion scheme. The starting model is used for wave propagation and therefore drives the data-misfit gradient, whereas the prior model is never explicitly used for solving the wave equation and only drives the optimization step as an additional constraint to minimize the total objective function. Thus, the prior model is not required to follow kinematic properties as precisely as the initial model, except in zones of poor illumination. In addition, we investigate the influence of a simple dynamic decreasing weighting of the prior model term. Once the cycle-skipping problem has been solved, the impact of the prior model term is gradually reduced within the misfit function to be driven by seismic-data only.

INTRODUCTION

Robust reservoir characterization is a key issue for oil and gas exploration and production. The seismic processing workflow can be roughly summarized in three main steps: velocity model building, migration in time or in depth, and elastic properties characterization through amplitude variation-with-offset (AVO) or amplitude variation-with-angle (AVA) analysis. The velocity model building remains a key step that can be tackled with reflection/refraction tomography in time and/or depth domain. A recent alternative for velocity model building is the full waveform inversion (FWI) that allows to reconstruct high-resolution velocity models of the subsurface through the extraction of the full information content of the seismic data (Tarantola, 1984; Virieux and Operto, 2009).

FWI is a multiscale data-fitting method well-adapted to wide-angle/wide-azimuth acquisition geometries because it uses simultaneously diving and reflected waves. FWI is classically solved with local optimization schemes and is therefore strongly dependent on the starting model definition. This starting model should predict arrival times with errors less than half of the period to cancel the cycle-skipping ambiguity (Virieux and Operto, 2009). The multiscale strategy performed by moving from low to high frequencies during the inversion allows to reduce the nonlinearities and cycle-skipping issues of the inversion and helps convergence toward the global minimum. Recent applications of FWI to real data have shown promising results for exploration projects; see 3D examples in Plessix and Perkins (2010) or Sirgue et al. (2010). Monoparameter reconstruction of the acoustic velocity for exploration is quite impressive even in the anisotropic case (Prieux et al., 2011). Elastic

Manuscript received by the Editor 20 March 2012; revised manuscript received 6 September 2012; published online 4 February 2013.
[1]Université Joseph Fourier & CNRS, Institut des Sciences de la Terre (ISTerre), Grenoble, France. E-mail: amir.asnaashari@ujf-grenoble.fr; romain.brossier@ujf-grenoble.fr; stephane.garambois@ujf-grenoble.fr; jean.virieux@ujf-grenoble.fr.
[2]TOTAL E&P CSTJF, Pau, France. E-mail: francois.audebert@total.com.
[3]TOTAL E&P, Geoscience Research Center, Aberdeen, U. K.. E-mail: pierre.thore@total.com.

parameters could also be recovered for exploration targets (Brossier et al., 2009; Prieux et al., 2012), but elastic inversion applies rather to seismological scales where phases are nicely separated (Fichtner et al., 2010; Tape et al., 2010).

Preconditioning or regularization techniques may alleviate the nonuniqueness of the ill-posed inverse problem. Tikhonov and Arsenin (1977) have proposed a regularization strategy within the optimization step to find the smoothest model that explains the data. Preconditioning techniques acting as a smooth operator on the model update (Operto et al., 2006) may add strong prior features of the expected structure through directive Laplacian preconditioning, as in Guitton et al. (2010). Regularization schemes that preserve edges and contrasts have also been developed for specific FWI applications through an ℓ_1 model penalty (Guitton, 2011) or through a multiplicative regularization (Abubakar et al., 2009) that mimics the total variation scheme (Rudin et al., 1992). Regularization can also be expressed in the curvelet or wavelet domains (Herrmann et al., 2009; Loris et al., 2010). In such domains, the ℓ_1 norm minimization is generally preferred for the model term penalty because it ensures sparsity in the model space.

All the previous regularization techniques allow to stabilize the inversion scheme by assuming a particular representation or structure of the velocity model (smoothness, sparsity, and so on). However, prior model information is generally not used in classical FWI implementation even if Hu et al. (2009) recently suggested to use a prior model in the multiplicative regularization term. Several sources of prior model information are usually available at the exploration stages, such as sonic logs, exploration well data, or geologic information of the field. One may want to use such prior information in the FWI scheme as is done in other velocity building techniques. Taking into account the prior information could also be highly important for monitoring purposes, where many different and precise prior data types have been collected for the target zone. Prior information can be introduced through the generalized Tikhonov regularization using the Bayesian formulation (Greenhalgh et al., 2006; Mead and Renaut, 2009) where the prior model is related to the expected model for the Bayesian interpretation. Strict Tikhonov regularization can be recast into this formalism as well. However, combining strict Tikhonov regularization and generalized Tikhonov regularization may lead to difficulties in defining the respective weights of the different information: prior information and expected smoothness of the model.

Several studies have been done on using two model penalty terms in geophysical electromagnetic applications, such as for the inversion of magnetic stripe data (Farquharson and Oldenburg, 1998) and for the inversion of controlled-source audio-frequency magnetotellurics data to recover a 1D conductivity structure (Routh and Oldenburg, 1999). In this study, we investigate the performances of an FWI scheme based on two model penalty terms in the misfit definition in addition to the data term: the Tikhonov term to ensure smoothness, and a prior model term to drive the inversion in a given direction. In the first part, we present the theoretical framework of our study. Then, through a synthetic application on the Marmousi model, we show the critical effect of the prior model penalty term on the FWI results. We shall highlight how the limited range of wavenumber sampling coming from the limited frequency band and the acquisition geometry may be compensated with the prior model information, for surface multiple-free data and also data containing surface multiples. We shall underline the fundamentally different role of the prior model and of the starting model within the FWI procedure.

THEORY

Full waveform inversion relies on an iterative local optimization problem that is generally introduced as a linearized least-squares problem. The optimization attempts to minimize the residuals between the observed and the modeled wavefields at the receivers. The linearized inverse problem remains ill-posed, and therefore multiple model solutions can provide a satisfactory fit of the observed data. Prior information is generally introduced through regularization in the inverse formalism. However, for specific applications where other sources of information such as sonic logs, stratigraphic data, or geologic constraints are available, it is crucial to take these into account in the inversion process and incorporate them into a prior model, to ensure robust and consistent results. To do so, we briefly introduce the full waveform inversion algorithm including the model norm contribution.

The general definition of the misfit function for solving ill-posed inverse problems could be recast as the Tikhonov function (Tikhonov and Arsenin, 1977)

$$\mathcal{C}(\mathbf{m}) = \mathcal{C}_d(\mathbf{m}) + \lambda \mathcal{C}_m(\mathbf{m}). \tag{1}$$

The data misfit $\mathcal{C}_d(\mathbf{m})$ is based on a norm of the residuals between observed and computed data in the data space, and the model norm $\mathcal{C}_m(\mathbf{m})$ term is based on a norm of a model penalty function in the model space. In the standard Tikhonov approach, this penalty function is based on the first spatial derivative of the current model that should have a minimal norm, thus giving a smooth model. The hyperparameter λ is the regularization parameter, also called trade-off parameter, that balances contributions between the data and the model terms.

For applications where prior information on the model can be established, we add a second penalty term to the misfit function. This term estimates residuals between the current model at a given iteration and the prior model considered at that same iteration. The objective function can now be written as the following expression,

$$\mathcal{C}(\mathbf{m}) = \mathcal{C}_d(\mathbf{m}) + \lambda_1 \mathcal{C}_{1_m}(\mathbf{m}) + \lambda_2 \mathcal{C}_{2_m}(\mathbf{m}), \tag{2}$$

where the Tikhonov term is denoted by $\mathcal{C}_{1_m}(\mathbf{m})$ and the prior model misfit term by $\mathcal{C}_{2_m}(\mathbf{m})$. Two regularization hyperparameters λ_1 and λ_2 are introduced, to allow weighting of the penalty terms with respect to each other and to the data term.

Let us express these three terms in a more explicit way using ℓ_2 norms. The data term may be written as

$$\begin{aligned}\mathcal{C}_d(\mathbf{m}) &= \sum_{ns} \|\mathbf{W}_d(\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m}))\|^2 \\ &= \sum_{ns} \frac{1}{2}\{(\mathbf{d}_{\text{obs}} - \mathbf{d}(m))^T \mathbf{W}_d^T \mathbf{W}_d(\mathbf{d}_{\text{obs}} - \mathbf{d}(\mathbf{m}))\},\end{aligned} \tag{3}$$

where $\mathbf{d}_{\text{obs}}$ and $\mathbf{d}(\mathbf{m})$ are vectors for the observed and computed data, respectively. For our specific investigation, we consider a time-domain approach and each component of these vectors are samples of time-domain seismograms recorded at receiver positions for one seismic source. This misfit function results from a sum over the ns sources of the experiment. The matrix $\mathbf{W}_d$ is a weighting operator on the data. This matrix can also be seen as the inverse of the square-root of the covariance matrix of the data, which contains information on data uncertainties. Considering a constant

measurement quality and uncorrelated traces, we end up with a diagonal matrix of $\mathbf{W}_d = \sigma_d I$, where σ_d is the standard deviation of the data and I is the identity matrix (Tarantola, 2005). The synthetic data $\mathbf{d}(\mathbf{m})$ nonlinearly depend on the model parameters denoted by $\mathbf{m} = \{m_i\}_{i=1,N_m}$, where N_m is the number of unknowns. These model parameters should be determined through the inverse procedure by reducing this data term.

The second term of the misfit function is the Tikhonov term and can be written as

$$\begin{aligned}\mathcal{C}_{1_m}(\mathbf{m}) &= \|\mathbf{B}_x\mathbf{m}\|^2 + \|\mathbf{B}_z\mathbf{m}\|^2 \\ &= \frac{1}{2}\{\mathbf{m}^T\mathbf{B}_x{}^T\mathbf{B}_x\mathbf{m} + \mathbf{m}^T\mathbf{B}_z{}^T\mathbf{B}_z\mathbf{m}\} = \frac{1}{2}\{\mathbf{m}^T\mathbf{D}\mathbf{m}\},\end{aligned} \tag{4}$$

where $\mathbf{B}_x$ and $\mathbf{B}_z$ are the first-order spatial derivative operator matrices with respect to x and z, respectively. In practice, they can be reduced to the second-order Laplacian operator $\mathbf{D}$. We use a classical five-point finite-difference stencil to implement the operator $\mathbf{D}$.

The third term of the objective function is related to the prior model $\mathbf{m}_p$, which can be designed from different information and could be set prior to the seismic inversion, but which could be also adapted iteratively during the inversion procedure. This so-called prior model norm term is computed using the expression

$$\begin{aligned}\mathcal{C}_{2_m}(\mathbf{m}) &= \|\mathbf{W}_m(\mathbf{m} - \mathbf{m}_p)\|^2 \\ &= \frac{1}{2}\{(\mathbf{m} - \mathbf{m}_p)^T\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m} - \mathbf{m}_p)\},\end{aligned} \tag{5}$$

where the matrix $\mathbf{W}_m$ is a weighting operator on the model space. This matrix can also be seen as the inverse of the square-root of the covariance matrix of the model, and contains prior uncertainty information of the prior model parameters. In our implementation, because we want to separate the influence of the diagonal and off-diagonal terms of the covariance matrix, we choose a diagonal $\mathbf{W}_m$ matrix, $\mathrm{diag}(\mathbf{W}_m^T\mathbf{W}_m) = 1/\sigma^2(\mathbf{m})$. The prior weighting model $\sigma^2(\mathbf{m})$ contains the prior model uncertainty (variance) and the potential weighting function, and will be discussed in the application section. The covariances (off-diagonal terms) are implicitly taken into account through the Tikhonov term.

Does the $\mathbf{W}_m$ operator play a critical role in driving the inversion procedure toward a given minimum? This is a question we want to investigate. Note that the misfit function, mixing data and model quantities, is dimensionless due to the introduction of the matrices $\mathbf{W}_d$ and $\mathbf{W}_m$, and through the hyperparameter λ_1 dimension. To have three dimensionless terms in the sum, the hyperparameter λ_1 has a dimension $[\dim(h^2)/\dim^2(\mathbf{m})]$, due to the dimensionality of the $\mathbf{D}$ operator, where the grid size h is for a 2D square regular Cartesian grid. For a model described by velocity, the dimension of λ_1 is second squared ($m^2/(\mathrm{m/s})^2 = s^2$).

Minimizing the misfit function classically leads to the normal equation system, which can be written as

$$\mathcal{H}_m\Delta\mathbf{m} = -\mathcal{G}_m, \tag{6}$$

where the gradient and the Hessian of the misfit function are denoted $\mathcal{G}_m$ and $\mathcal{H}_m$, respectively. The gradient expression can be written with three components as

$$\begin{aligned}\mathcal{G}_m = &-\left(\frac{\partial\mathbf{d}(\mathbf{m})}{\partial\mathbf{m}}\right)^T\mathbf{W}_d^T\mathbf{W}_d(\mathbf{d}_{\mathrm{obs}} - \mathbf{d}(\mathbf{m})) + \lambda_1\mathbf{D}\mathbf{m} \\ &+ \lambda_2\mathbf{W}_m^T\mathbf{W}_m(\mathbf{m} - \mathbf{m}_p).\end{aligned} \tag{7}$$

The sensitivity matrix $\mathbf{J} = \partial\mathbf{d}(\mathbf{m})/\partial\mathbf{m}$ is composed by the Fréchet derivatives of the synthetic data with respect to the model parameters. The data-term gradient is efficiently computed with an adjoint formulation (Plessix, 2006) without an explicit computation of the matrix $\mathbf{J}$. The two terms related to the model penalties are generally straightforward to compute and are easily added to the data-term gradient contribution, leading to negligible computer memory and CPU-time increase.

The Hessian matrix is based on the second derivative of the misfit function and is not computed in our implementation. Instead, we minimize our problem with a bounded quasi-Newton method using the L-BFGS-B routine (Byrd et al., 1995). This routine allows to take into account an approximate nondiagonal inverse Hessian from previous gradient and model vectors, and performs a line-search satisfying Wolfe's conditions. This bounded limited-memory quasi-Newton method is an efficient alternative to preconditioned steepest-descent or conjugate-gradient methods based only on gradients and/or approximate diagonal Hessian approaches. This cheap and efficient estimation of the influence of the inverse Hessian in the optimization improves focusing, partially corrects the descent direction from effects due to limited aperture illumination and frequency bandwidth, and respects dimensionalities of the different parameter values (Brossier et al., 2009).

A major point for real data applications is the source-wavelet estimation. Our FWI is implemented in the time-domain for the forward and the inverse problem. The source-wavelet estimation is, however, straightforwardly implemented in the frequency-domain by a linear inverse problem resolution. The computed and observed time-domain data are Fourier transformed to apply the Pratt (1999) (his equation 17) source estimation equation for each frequency. The Fourier coefficients of the wavelet are then transformed back to the time-domain and appropriately processed (anticausal mute and/or band-pass filtering if required) before performing FWI. This estimation is performed once before the optimization. In the following tests applied to synthetics, we use the exact source wavelet for fair comparisons, such that the results are not biased by potential errors from this estimation.

APPLICATION TO MARMOUSI MODEL

In this section, we study the effect of prior information in FWI. In particular, we show how prior information allows to mitigate the lack of seismic illumination. A selected target zone of the Marmousi II P-wave velocity distribution (Martin et al., 2006) and a homogeneous density model are considered. The target exhibits two gas sand traps (Figure 1a). We consider a shallow-water configuration with a water depth of only 25 m. Our acquisition geometry contains 54 isotropic pressure-sources, located along a horizontal line at 15-m depth, every 50 m. The layout is the same for all shots, one fixed horizontal receiver line at 15-m depth and two fixed vertical lines of receivers inside two exploration wells at $x = 50$ m and $x = 2700$ m with a 10-m interval between sensors. The deepest receivers inside the wells are at $z = 1265$ m. The grid is regular, with the grid size equal to 5 m, and it is consistent for modeling and inversion. We do not consider any sources within the wells because this design is unusual and quite expensive. However, we consider sensors inside the

wells, which could be installed for exploration or monitoring purposes, and allow us to dramatically increase the illumination for velocity reconstruction. Note that our final test will be performed without these well sensors, to mimic a pure surface acquisition. A Ricker wavelet source with a central frequency of 10 Hz is used for all shots. The time seismograms are generated using finite-difference modeling in the time-domain with a fourth-order stencil in space and a second-order integration in time. Perfectly-matching-layer (PML) absorbing boundary conditions (Berenger, 1994) are used for nonreflecting boundaries. The first tests are performed using a PML on top, to mimic multiple-free data. The last test will consider a free-surface condition, modeling surface-multiples. The recorded pressure data are used as observed data, at the surface and in wells. Figure 2a shows an example of a seismogram generated by a shot located at the center of the source line.

In our study, the data weighting matrix $\mathbf{W}_d$ is chosen as identity $\mathbf{W}_d = \mathbf{I}[\dim(\text{data})]^{-1}$, where dim(data) means the unit of pressure data. To have a dimensionless objective function, $\mathbf{W}_d$ should have a dimension which is the inverse of the data dimension. Note that, for all further applications, the Tikhonov regularization parameter is kept fixed to a small value, imposing only a weak smoothing constraint, because we mainly focus on analyzing of the effects of the prior penalty term.

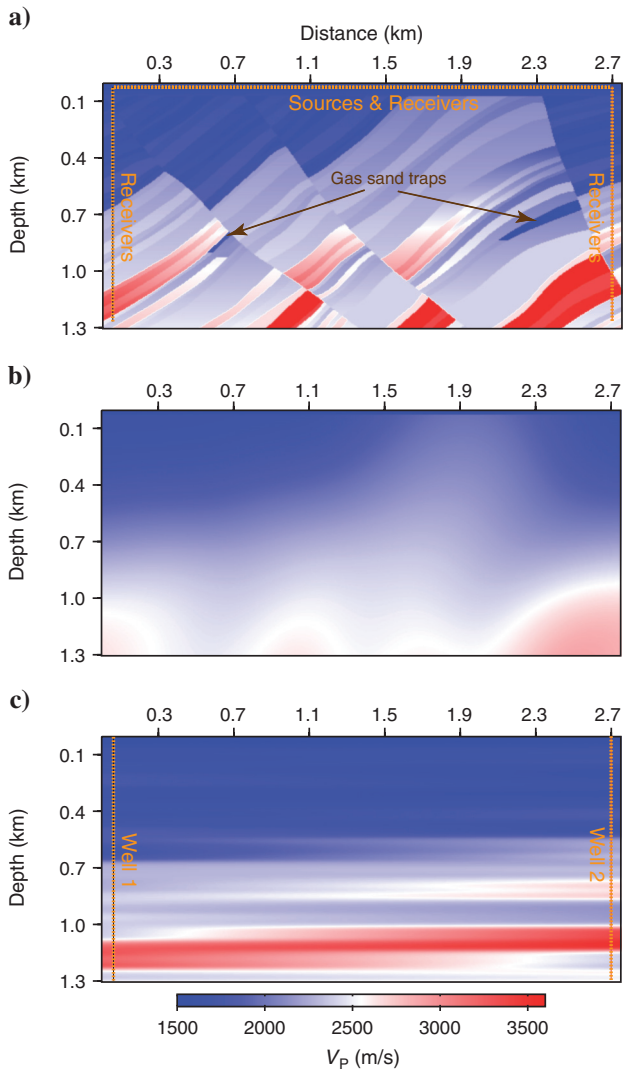

Figure 1. (a) The true V_P velocity model which is a small part of the Marmousi model and the acquisition geometry; (b) initial model for inversion which is a smooth model of the true model; (c) the prior model created by linear distance weighted interpolation in the x-direction between the exact values inside two exploration wells and then gently smoothed.

A smooth velocity model (Figure 1b), which mimics a time-tomography velocity model based on first arrivals and reflected events, and referred to henceforth as "smoothed velocity model", is used as the initial model for FWI. A time-domain FWI approach is used, involving all the frequencies of the spectrum (maximum 30 Hz in this case). No additional hierarchical approach, such as the frequency-continuation approach of Bunks et al. (1995), is used in these examples. This means that the weighting of each frequency is directly linked to its amplitude in the spectrum. A first investigation (Figure 4a) is performed with noise-free data and a standard regularized FWI method, without considering a prior model (equivalent to $\lambda_2 = 0$). The result shows that the optimization is trapped in a local minimum. This issue can be related to cycle-skipping ambiguities due to the starting model inaccuracy, especially in the deeper part below 700 m and on the left part of the model until the second fault. Due to these inaccuracies, the target zones composed of the two reservoir areas are not well-recovered with this configuration.

FWI with prior model and impact of prior weighting matrix ($\mathbf{W}_m$)

In our framework, where well information does exist, the FWI method should use this nonseismic information as prior information for the inversion. We first need to build the prior velocity model $\mathbf{m}_p$ and the model weighting matrix $\mathbf{W}_m$ that contains the prior model uncertainty. In our study, we consider that the sonic-log measurements acquired in the two exploration wells provide an accurate estimation of the local vertical velocity. A prior model could have been created from interpolation of the well velocity, following picked horizons in the migrated section. Instead, we build a crude prior velocity model based on a linear interpolation between the two well locations without any migration and picking approach. This interpolated model (Figure 1c) from only the well data, henceforth called "interpolated velocity model", despite being far from the true 2D structure of the Marmousi model, will be used as a prior velocity model for regularized inversion. As shown in the following test, this crude prior model allows to significantly help the inversion to converge, and when applied to real data, the more accurate the prior is, the better the FWI results will be: For example, if the prior model is derived using standard quantitative interpretation techniques.

The prior model has to be associated with the weighting matrix $\mathbf{W}_m$, to weight the penalty associated to the model residual $(\mathbf{m} - \mathbf{m}_p)$. As already mentioned, we use a diagonal weighting matrix containing the uncertainty and some weighting. From how the prior model is built, we know that, quantitatively, the interpolated velocity values should be accurate close to the well positions, but they can be erroneous far from the wells, because the structure is highly heterogeneous. Therefore, we decided to build a weighting shape whose uncertainty values follow a Gaussian function with weak values near the wells and increasing values in the center of the area (Figure 3a). This is the prior weighting model A.

A key point in all additive regularized optimization schemes is the selection of the weighting hyperparameters. As already mentioned, the λ_1 value chosen is small enough to ensure a slight smooth-

ing of the results. In practice, to select the λ_2 hyperparameter, we compute the misfit function for the starting model for $\lambda_2 = 1$. Based on the ratio γ between the prior-model misfit $\lambda_2 \mathcal{C}_{2_m}(\mathbf{m})$ and the data-term misfit $\mathcal{C}_d(\mathbf{m})$, we adjust the λ_2 value such that $10^{-3} < \gamma < 10^{-2}$. Therefore, by selecting this reasonable ratio of prior-model and data misfit terms, the FWI is prevented to minimize the model norm heavily at early iterations. In fact, an even stronger weight is applied to the data term in the global objective function. In this test, we choose to have the ratio $\gamma = 10^{-2}$.

Figure 4c shows the reconstructed velocity model after FWI, starting from the smoothed velocity model, and using the interpolated velocity model as a prior model. We can see that the shallow left part of the reconstructed model has been much improved compared to Figure 4a. However, the deeper part of the result remains dominated by the footprint of the interpolated velocity model used as a prior model. This footprint can be interpreted as an inappropriate relative weight between the prior penalty term and the data misfit term, for waves that illuminate this deeper part. The consistency of the two terms at shallow depth leads to an improved reconstruction. To visually see the relative amplitude of the different terms of the gradient, the absolute value of the data-term gradient (Figure 5a), the prior-model term gradient (Figure 5b), and their ratio (prior-model/data) (Figure 5c) are computed at the first iteration. With increasing depth, the amplitude of the data-term gradient decreases because the associated wave amplitudes, mainly in a reflection regime, decrease due to geometrical spreading, intrinsic attenuation, and energy partitioning at interfaces. The ratio between the prior-model and the data gradients therefore shows that the deep part of the gradient is driven by the prior-model at the expense of the data term because of the homogeneous weighting term with depth in the $\mathbf{W}_m$ matrix. To overcome this unfavorable balance between the data and the prior terms in the optimization, a weighting is required and can be implemented in two different ways. We can either modify the $\mathbf{W}_m$ matrix to decrease the weight in depth or change the data-term

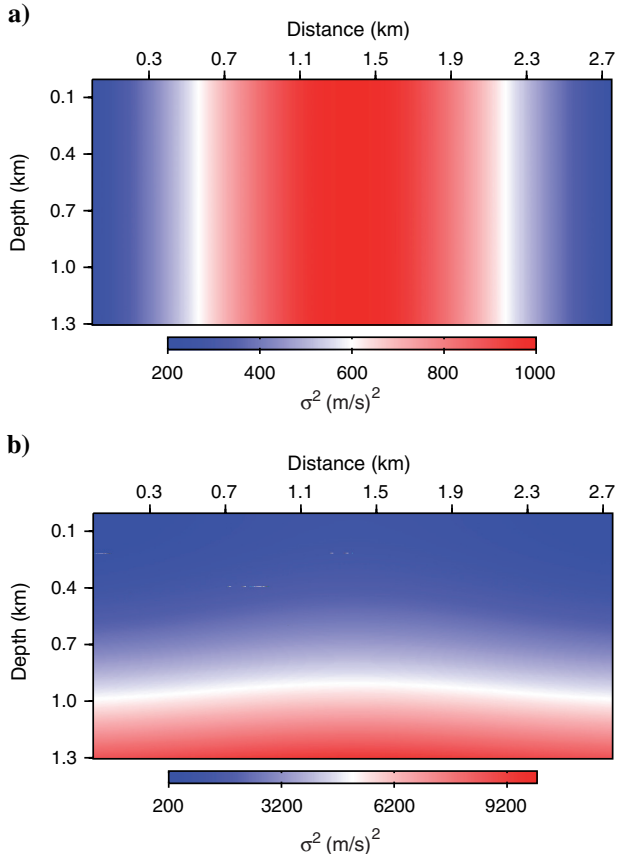

Figure 3. Two types of prior weighting model used for regularized inversion: (a) model A, the Gaussian function varying only in the x-direction between two wells with maximum value at the center of model, and (b) model B, the same variation in x complemented by a quadratic evolution in the z-direction (the Gaussian lateral variation could be seen now in the undulation of the white interface).

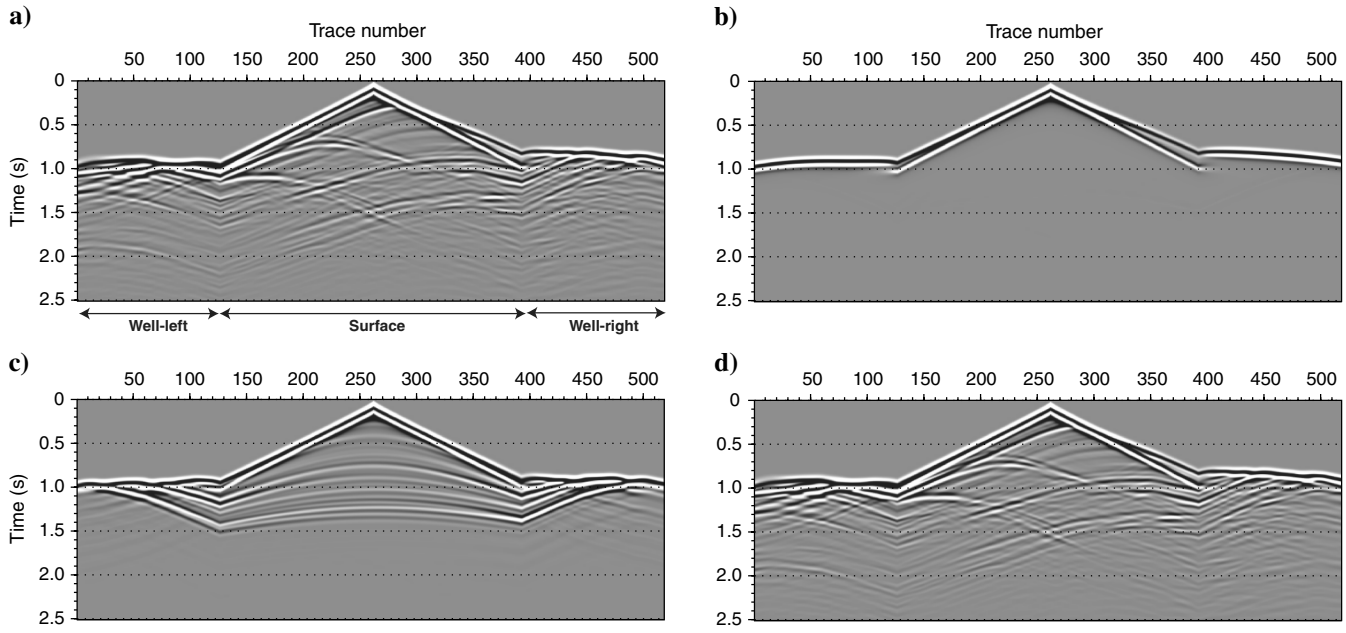

Figure 2. Seismograms of pressure data for the source located almost at the center of the Marmousi model $x = 1.4$ km: (a) recorded inside the true model; (b) calculated inside the smooth initial model; (c) calculated inside the interpolated velocity model; and (d) computed inside the final model obtained through our dynamic approach.

weighting $\mathbf{W}_d$ matrix such that the late arrivals have more weight in the data misfit and the data-gradient terms. This second weighting can be linked to the metric choice of the misfit function norm for the data term as defined in Jin et al. (1992).

In our study, we choose to involve the depth weighting in the model space and we use a rough but efficient approximation of the geometrical spreading to change the $\mathbf{W}_m$ matrix in depth: We propose to make the operator $\mathbf{W}_m^T\mathbf{W}_m$ decrease by a simple $1/z^2$ with respect to the depth z, to compensate for the propagating decay of the wave amplitude. This kind of depth weighting has been used in the controlled source electromagnetic method (Plessix and Mulder, 2008) and gravity inversion applications (Li and Oldenburg, 1998). Plessix and Mulder (2008) have proposed the depth weighting matrix to compensate for the exponential decay of the amplitude of electromagnetic waves and also geometrical spreading. This depth weighting was used as a preconditioning of the model parameter (Plessix and Mulder, 2008) and to counteract the geometric decay of the kernels in inversion (Li and Oldenburg, 1998). In our application, we use the same general principle, but our main goal is to make an appropriate balance between the prior-model norm and data misfit in depth. We combine this weighting and the uncertainty associated to distance away from the wells to build a new weighting matrix (Figure 3b), referred to as "prior weighting model B."

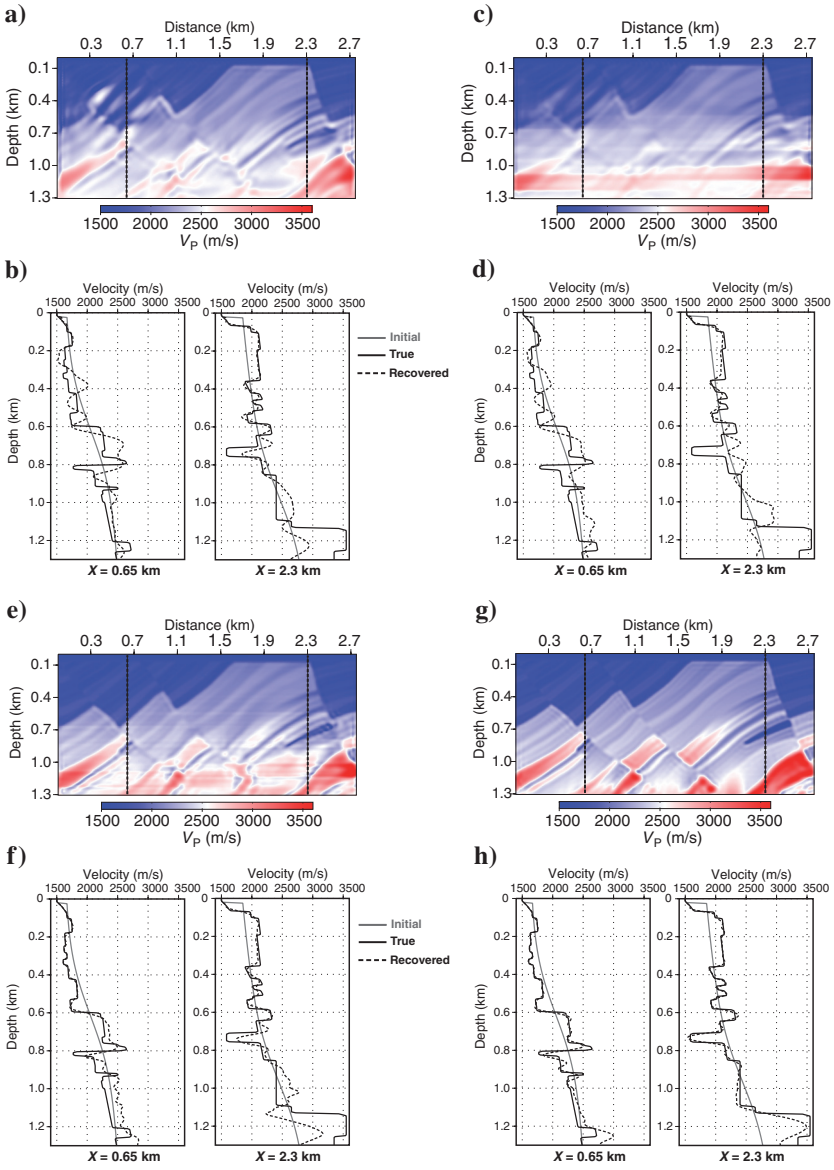

Figure 4. The recovered V_P models by FWI and two QC vertical logs passing through the two target areas at $x = 0.65$ km and $x = 2.3$ km: (a) reconstructed model starting from the smooth initial model and without a prior model, small λ_1 and $\lambda_2 = 0$; (b) two vertical logs corresponding to the model (a); (c) reconstructed model starting from the smooth initial model and with the prior model, small λ_1, fixed λ_2, the prior weighting model A and the ratio between prior-model and data misfit terms $\gamma = 1 \times 10^{-2}$; (d) two vertical logs corresponding to the model (c); (e) reconstructed model starting from the smooth initial model and with the prior model, small λ_1, fixed λ_2 same as case (c), the prior weighting model B. Note the γ-ratio is now decreased to 3×10^{-3}; (f) two vertical logs corresponding to the model (e); (g) reconstructed model starting from the smooth initial model and with the prior model, small λ_1, initial value of $\gamma = 3 \times 10^{-3}$, the prior weighting model B and using the dynamic prior approach. The prior model is removed from the inversion at the end of the procedure; (h) two vertical logs corresponding to the model (g).

FWI is now applied using the smoothed velocity model as the initial model and the interpolated velocity model as the prior model, together with the prior weighting model B. The same hyperparameter λ_2 is used, but note that the γ-ratio value between prior-model and data misfit is decreased to around 3×10^{-3}. This weighting model B allows to successfully balance the gradient energy with the depth as shown in Figure 5d and 5e. The reconstructed velocity is shown in Figure 4e and exhibits a significant improvement compared to Figure 4a. This result first shows the importance of the prior weighting, which should contain appropriate uncertainty information, but should also ensure an appropriate balance between the prior misfit term and the data misfit term in the optimization. In this case, the prior term significantly helps the inversion to converge to the global minimum of the optimization problem, mitigating the cycle-skipping issues that the data misfit term cannot handle alone. In fact, adding the prior model penalty allows to change successive descent directions and helps the inversion to converge to the correct global minimum valley of the misfit function. This test shows that prior information allows to constrain inversion and, therefore, mitigates the nonuniqueness issue of ill-posed inverse problems.

Roles of initial versus prior models

In many geophysical inversions, it has been proven successful to choose the initial model equal to the prior model (see Oldenburg [1994] and Routh and Oldenburg [1999] for electrical/electromagnetic inversion, and Miller et al. [2008] and Routh and Anno [2008] for time-lapse inversion), when it is chosen sufficiently accurate. In this part, we address the relative role

of the prior and initial models in the inversion procedure, when only partial information is contained in the available models. A natural first idea could be the use of the prior velocity model (Figure 1c) as the initial model of FWI. Because this model helps the FWI when it is used as prior model, it could be a good candidate for the initial model of the inversion. Fundamental differences exist when using a particular velocity model as a prior model which has no direct impacts on the modeling of synthetic data or when using it as an initial model with direct consequences on the synthetic data. Figure 6 illustrates the inversion result derived using the classical regularized FWI (same tuning as Figure 4a) and the interpolated velocity model as initial model. We can clearly see that the inversion converges toward a local minimum, leaving it far from being a satisfactory result. Moreover, in this case, the optimization process stops after only a few iterations. The shallow part on the right side of the model seems satisfactory, but the left side and the deeper parts seem to be badly handled by this initial model, built from interpolation in this strongly laterally varying structure. One interpretation of this failure is related to the major difference in the meaning of the initial and of the prior model: The initial model must be localized in the attraction valley of the global minimum of the misfit function, often related in seismic as being kinematically accurate and not generating erroneous arrivals in the synthetic data computed using the wave equation (see Figure 2a and 2c). Indeed, it is much more difficult for the inversion workflow to suppress or shift a structure than to create a new one. On the contrary, the prior model is never explicitly used as an input for solving the wave equation and is only used to drive the optimization step to minimize the total objective function. It can, therefore, contain any structure, complementary to the information contained in the initial model, that can drive inversion toward expected zones of the model space. In our case, the prior model allows FWI to be driven and partially fills in the lack of low wavenumbers that cannot be extracted from only the data. The smoothed velocity model and the interpolated velocity model contain pieces of partial information on the velocity model, that, when used alone, are not sufficient to converge toward the global minimum. Only an appropriate combination of both pieces of information, through the initial and prior models, allows to exploit the partial information included in both, and allows to significantly improve the results. Note, however, that for regions of poor seismic illumination, as the optimization is driven by the prior model, this model is required to be as accurate as possible (kinematically correct) to ensure good results. In this case, the prior model puts its imprint on the current model and then is partially used for solving the wave equation.

Dynamic prior regularization parameter

In complex environments, the prior model derived from extra information on the target zone may be far away from the exact model we never reach. Even if the prior model can significantly improve results by driving the inversion in an appropriate direction, the final model can keep a significant footprint of the prior model structure and may prevent a significant expression of the data itself. As shown in Figure 4e, the result exhibits ghost interfaces coming from the interpolated prior model. These footprints of the prior setting do not honor the data itself. However, keeping a fixed hyperparameter

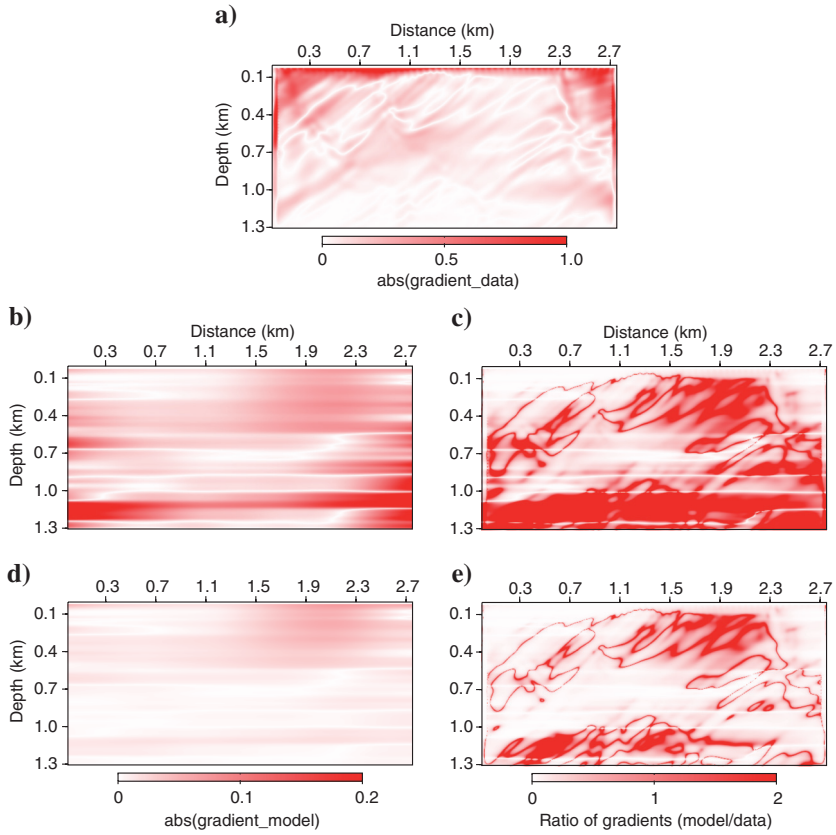

Figure 5. At first iteration of optimization, the absolute value of (a) the data-term gradient, (b) the prior-model term gradient, (c) the ratio between prior-model and data gradients, in case of using the prior weighting model A; (d) and (e) same as (b) and (c), respectively, but in the case of using the prior weighting model B.

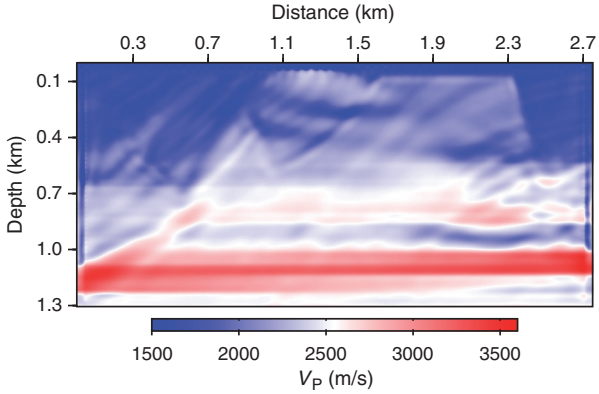

Figure 6. The recovered V_P model by standard FWI starting from an initial model equal to the interpolated velocity model, small λ_1 and $\lambda_2 = 0$.

on the prior term of the misfit function prevents the results from being improved because the prior model is intrinsically wrong in such a case.

Thus, one can investigate a dynamic weighting of the prior information, to decrease the weight of the prior term (λ_2) during iterations of the optimization. We suggest a simple dynamic approach, considering a starting λ_2 value that gradually decreases until it reaches zero. This method allows to drive FWI toward the global minimum valley of the objective function at the beginning, due to the prior-model influence, and to finally leave only the data term to drive the final iterations of the optimization, by gradually decreasing the prior weight. The Tikhonov regularization term is kept constant as we only discuss here the reciprocal influences of the data misfit term and of the prior term. Our heuristic approach is based on the decrease of the objective function with iterations. When the slope of the objective function curve becomes smaller than a specific threshold, the current λ_2 value is divided by a factor of two to reinforce the weight of data misfit term. Our criterion is based on the first derivative of the misfit function with respect to iterations, computed with a simple finite-difference stencil. The derivative value at each iteration is normalized by the first derivative value. During the optimization procedure, the corresponding derivative value is compared to the fixed threshold at each iteration. Every time that the derivative is smaller than the threshold, meaning that the misfit function curve with iterations is becoming too flat, the hyperparameter λ_2 is changed. The key issue of this strategy is the value of the threshold at which the hyperparameter λ_2 term must be decreased. We find that a few trials can narrow this value quite rapidly from variation of the misfit function. This threshold value should be of the order of $10^{-4} \sim 10^{-3}$. Note that, in our implementation, the L-BFGS-B optimization is restarted each time the hyperparameter λ_2 is changed.

The dynamic method has a similar property to the multiplicative regularization and cooling regularization approaches (van den Berg et al., 1999, 2003). In total variation (TV) as the multiplicative constraint, the data objective function itself is defined as the weight of TV. Therefore, the regularization term has a large weighting parameter in the beginning of the optimization process, and gradually decreases as the objective function is minimized and the data fitted.

Figure 4g shows the recovered model obtained by this dynamic method, using the smoothed velocity model as the starting model, the interpolated velocity model as a prior model, and the optimal weighting matrix B. As with the fixed λ_2 strategy, we can see that the reconstructed model is dramatically improved when compared

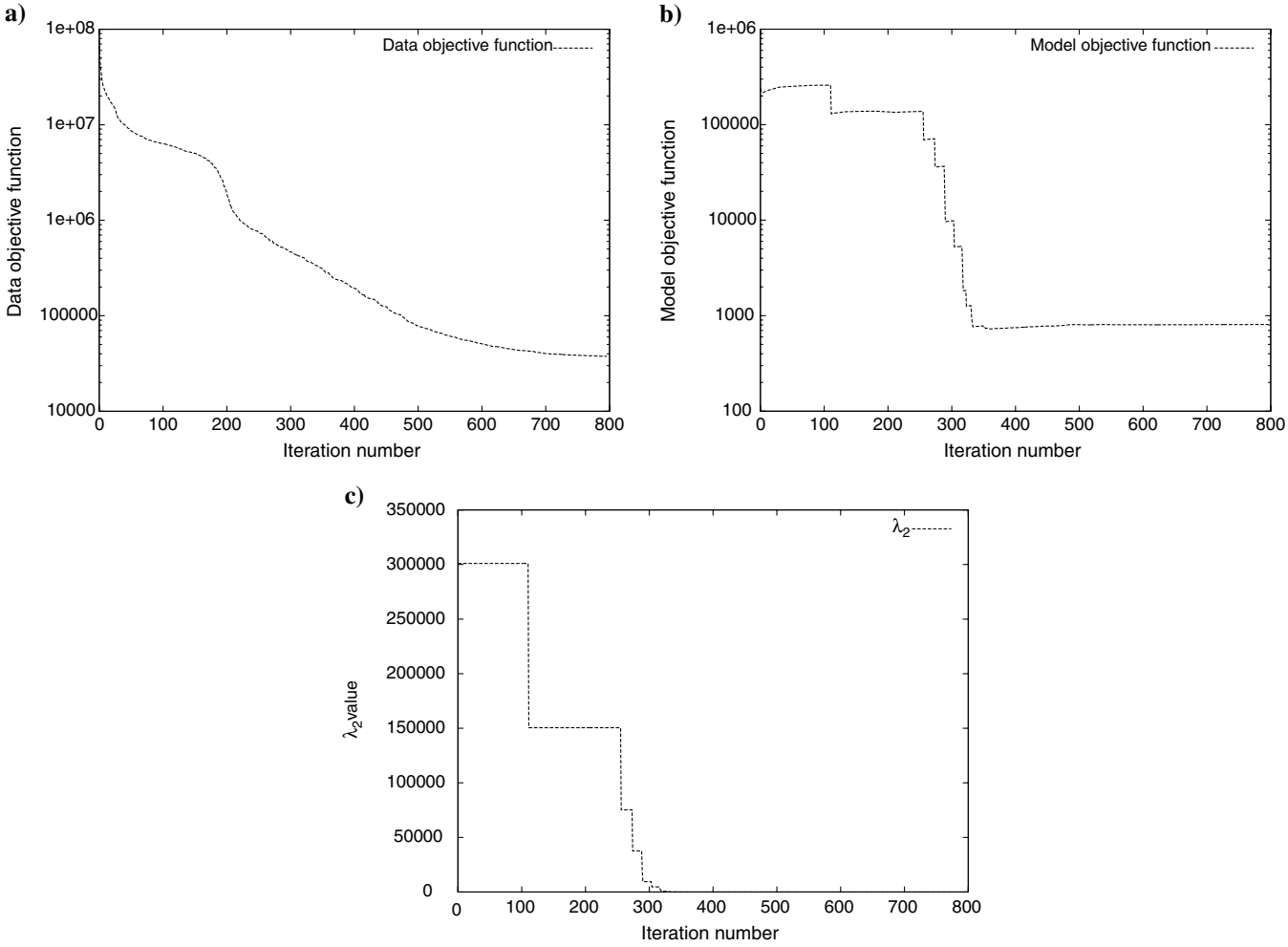

Figure 7. Evolution of (a) the data objective function, (b) the model objective function, and (c) λ_2 value with iterations in case of using the dynamic approach. Note that (a) and (b) curves are shown in logarithmic scale.

to the one obtained by standard FWI. Moreover, the dynamic approach allows to mitigate the footprint of the prior model, because during the optimization the prior penalty weight decreases with respect to the data misfit term. Thus, the effect of the prior model is being reduced slowly and the data misfit term helps inversion to converge to a quasi-perfect final model. As a quality control, vertical profiles taken through the two gas sand traps (for the initial, true, and recovered models) illustrate that the target velocity is recovered accurately (Figure 4h), compared to the result from standard FWI using the same initial model (Figure 4b). In addition, the computed seismogram inside the final model shows that the full seismic arrivals have been exploited during the optimization (see Figure 2d).

The objective function curves for the data term, the model norms, and the λ_2 curve as a function of iteration are shown in Figure 7. In this case, the data are without noise and a very small stopping criterion is selected to fit the data as much as possible, leading to a large number of iterations. This stopping criterion, based on the flatness of the misfit function for two successive iterations, is the same for all the inversion tests, so that the results are comparable. The data objective function always decreases, and by reducing λ_2, we try to prevent giving a high weight to the prior penalty term. Therefore, by reducing the model objective function value, an appropriate contribution of the prior model is kept during optimization. Note that there is no change on the small λ_1 value and that the Tikhonov term always exists, leading to a nonzero model objective.

The convergence of standard FWI without any prior model and FWI with dynamic prior weighting shows dramatic differences of the evolution of the total objective function (Figure 8). The standard FWI gets trapped in a local minimum and stops the optimization after 87 iterations. By including the prior model to the optimization, the path of descent is changed and the optimization procedure is not trapped by local minimum attraction basins. The beginning of the optimization appears quite equivalent for both approaches until iteration 40, even showing better convergence speed for the standard FWI. Please note, however, that even if the convergence rate is the same, the results may be different because the null-space is different between an inversion with and without the prior model penalty. After this step, due to different descent paths, standard FWI slows down convergence speed and rapidly stops. For FWI with dynamic prior weight (blue curve), we can observe a large decrease in the objective function between iterations 170 and 230. Looking at the updated model history for these iterations shows a significant improvement, associated with the prior model penalty use, in the shallow left part of the model, leading to a large decrease in the data misfit. In standard FWI, the data and Tikhonov terms of the misfit gradient alone are not able to solve this problem in this shallow left part of the model.

Noisy data

In the presence of noise, the ill-posedness of the inverse problem is increased. Therefore, we need to study the effect of noise on our proposed regularized FWI including the prior model penalty. We keep the same acquisition configuration, whereas an artificial Gaussian noise in the range of 0–30 Hz, the bandwidth of the source, is added to the true noise-free data. The signal-to-noise ratio (S/N) is around 7 dB. Figure 9 illustrates an example of shot gathers used for FWI (we have used the suaddnoise procedure of Seismic Unix [Cohen and Stockwell, 2008]).

Three inversion tests starting from the smoothed velocity model (Figure 1b) are performed. The first one uses the standard FWI without the prior model (Figure 10a). The second test uses the interpolated velocity model as prior model with a fixed λ_2 value (Figure 10c), and the third one uses the dynamic prior weighting (Figure 10e). All the parameters are chosen identical to those of the noise-free data set case, except the hyperparameter λ_2, which is now increased to account for the noise energy in data. The λ_2

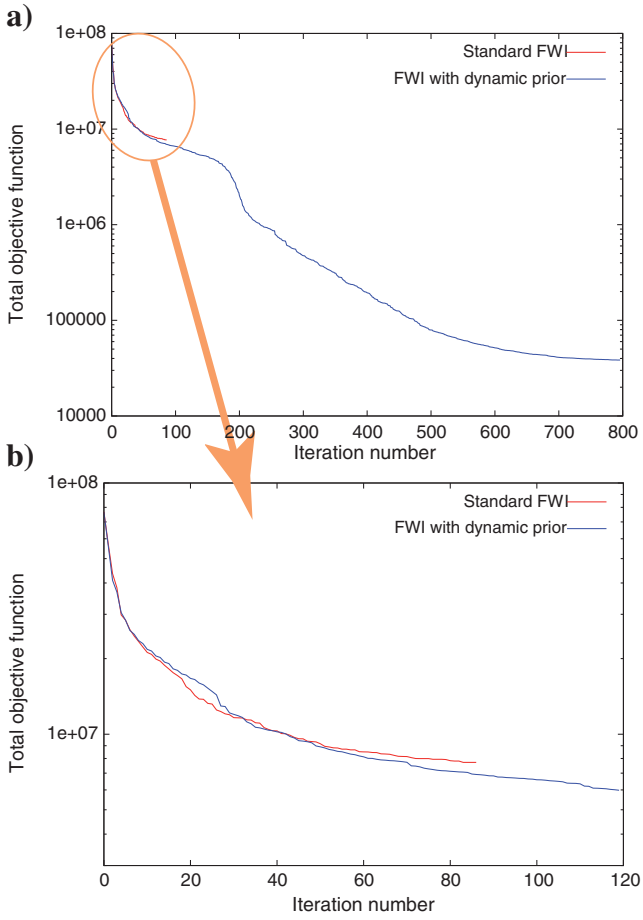

Figure 8. Comparison of the total objective function curves in the case of using the standard FWI and the dynamic prior weighting FWI: (a) for all iterations; and (b) at early iterations.

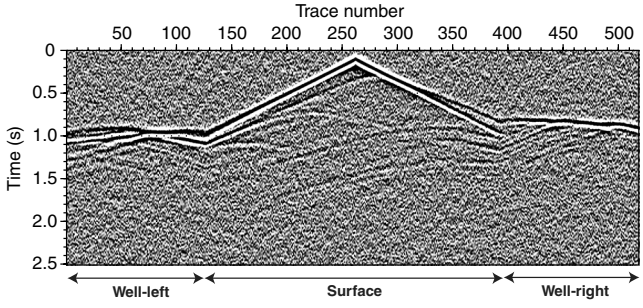

Figure 9. Noisy seismograms of pressure data for the source located almost at the center of the Marmousi model $x = 1.4$ km; random Gaussian noise added to the synthetic seismograms in bandwidth of 0–30 Hz and S/N= 7 dB.

value is chosen such that the ratio between the prior-model penalty and the data misfit remains the same in the global misfit function, and equal to the noise-free case ($\gamma = 3 \times 10^{-3}$). In the presence of noise, the data misfit function has a larger value than the noise-free case; therefore, a higher λ_2 value is required.

The results with noisy data remain consistent with the noise-free tests: The prior penalty term still drives the inversion toward a more realistic and accurate final model, albeit more noisy. The dynamic approach also remains an appropriate strategy as can be shown from the model reconstruction and from the two displayed vertical profiles (Figure 10f). This test for data with the presence of noise confirms the robustness of the approach for nonperfect data.

Surface acquisition and multiple-contaminated data

In this section, we apply our scheme to a less favorable frame: A free surface condition is used, meaning that surface multiples are now present in the data. Moreover, we suppress the sensors located in the wells, leading to a pure surface acquisition. Note that the sensors in the wells previously allowed to dramatically increase the illumination in this selected target of Marmousi, where the ratio maximum offset/depth is about two, instead of from three to four as in classical FWI applications to exploit diving waves. In such a configuration with a small offset compared to the depth, the FWI behaves generally like a nonlinear migration technique, because the low part of the wavenumber domain can not be retrieved (Plessix and Mulder, 2004) and we may question how the prior information may fill in this part of wavenumber domain.

The observed data used in this test are shown in Figure 11. The surface-related multiples can clearly be seen compared to the previous data set.

Two inversion tests starting from the smoothed velocity model are performed. The first one uses the standard FWI without the prior model (Figure 12a). The second test uses the interpolated velocity model as the prior model, the optimal prior weighting model B, and the dynamic approach (Figure 12c). Here, we assume that the sonic logs are still available and that we can use them for building the prior model (like in previous tests).

All the parameters are chosen identical to those of the previous tests. The starting λ_2 value is adapted, due to a difference in the data energy and trace number, to keep the same $\gamma = 3 \times 10^{-3}$ ratio between the prior-model and data misfit terms at the first iteration.

The result of standard FWI (Figure 12a) contains many anomalies and ringing artifacts in the shallow part. These effects could be associated with surface-related multiples coming from the free surface. Even if the main structures are recovered in the shallow part, the velocity model is heavily contaminated by artifacts at all depths leading to erroneous velocity values at the two reservoir depths (see the QC logs in Figure 12b). Moreover, due to the limited aperture coverage of the acquisition, the low part of the wavenumber spectrum is not recovered, and the structure seems depth stretched due to the initial model inaccuracy. Adding the prior model and the dynamic weighting allows to significantly improve the result (Figure 12c), canceling the shallow ringing effects associated to the surface-related multiples. In the shallow part of the target where illumination remains strong, the result is almost perfect. Deeper because of the lack of illumination, some artifacts appear, but the prior model allows, at least partially, to fill in the low part of the wavenumber spectrum that can not be retrieved from short-spread reflection only. This leads to well-positioned structures, until at least 1-km depth (Figure 12d).

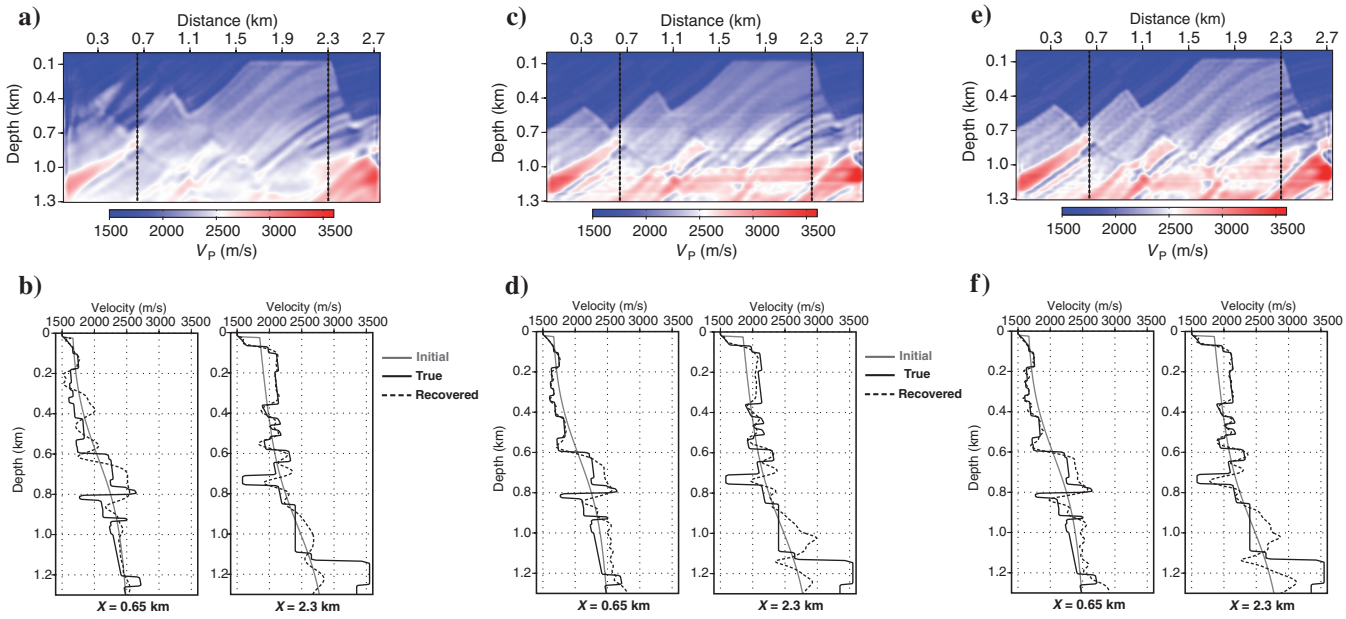

Figure 10. The recovered V_P models by FWI of the noisy data and two QC vertical logs passing through the two target areas at $x = 0.65$ km and $x = 2.3$ km: (a) reconstructed model starting from the smooth initial model and without a prior model, small λ_1 and $\lambda_2 = 0$; (b) two vertical logs corresponding to the model (a); (c) reconstructed model starting from the smooth initial model and with the prior model, small λ_1, fixed adapted λ_2 to have the ratio $\gamma = 3 \times 10^{-3}$ at first iteration, and the prior weighting model B; (d) two vertical logs corresponding to the model (c); (e) reconstructed model starting from the smooth initial model and with the prior model, small λ_1, same initial λ_2 before it is decreased to zero (or same initial ratio $\gamma = 3 \times 10^{-3}$), the prior weighting model B and using the dynamic approach. The prior model is removed from the inversion at the end of the procedure; (f) two vertical logs corresponding to the model (e).

DISCUSSION

So far, FWI has been considered essentially as a data-driven procedure with negligible contribution of prior model information and has therefore been investigated for seismic exploration purposes. As the knowledge of the target zone is increased, we may need to introduce more model-driven features in the optimization procedure, especially when we have poor illumination of the target zones.

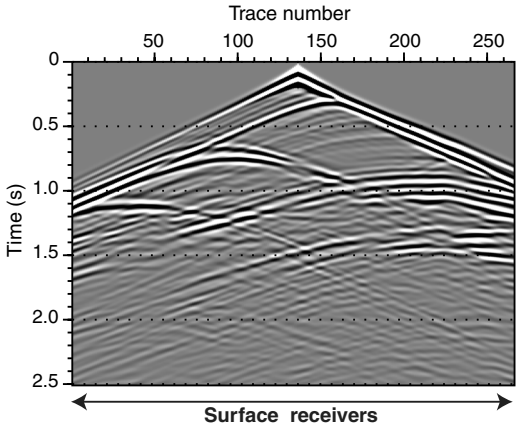

Figure 11. Seismograms of pressure data for the source located almost at the center of the Marmousi model $x = 1.4$ km, recorded inside the true model with free surface condition and using the receivers only at the surface.

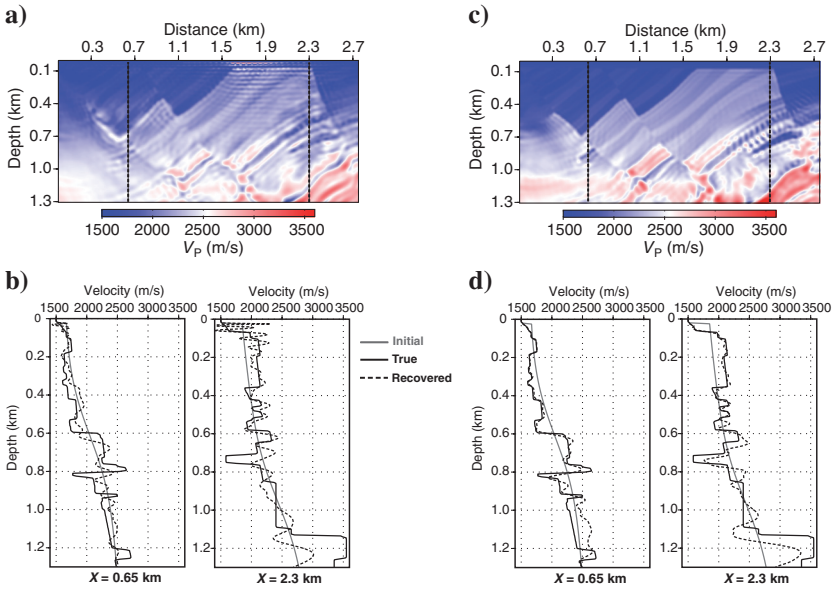

Figure 12. The recovered V_P models by FWI and two QC vertical logs passing through the two target areas at $x = 0.65$ km and $x = 2.3$ km, in case of free surface condition and using only the receivers at the surface: (a) reconstructed model starting from the smooth initial model and without a prior model, small λ_1 and $\lambda_2 = 0$; (b) two vertical logs corresponding to the model (a); (c) reconstructed model starting from the smooth initial model and with interpolated velocity model (Figure 1c) as a prior model, small λ_1, initial λ_2 value chosen to have the ratio $\gamma = 3 \times 10^{-3}$ at first iteration, the prior weighting model B and using the dynamic approach; (d) two vertical logs corresponding to the model (c).

The description of the misfit function with three terms should increase potential perspectives of the FWI because we may relax the illumination constraints of this approach at the expense of a better knowledge of expected features of the model we want to reconstruct.

The design of the hyperparameters and, more specifically, the dynamic evolution during the inversion procedure, could be improved and robustness should be analyzed. One can say that this tuning is based on the misfit evolution for different damping laws. From the synthetic example that has been studied, we have found that behaviors of the FWI for different acquisition configurations, namely the one with receivers at the surface and within two wells, and the one with receivers only at the surface with the surface-related multiples, are quite similar and, therefore, the tuning of hyperparameters should not be highly sensitive to the application.

In addition, deep targets could benefit as well from the prior information once the overburden structure has been defined. In seismic exploration, sub-basalt and subsalt imaging is quite challenging and any extra piece of information could help to improve the illumination of the target. The introduction of prior information would allow to help recover poorly illuminated zones, thus broadening the application of the full waveform inversion.

CONCLUSIONS

We have proposed a regularized FWI scheme that includes prior information as an optimization penalty term. Aside from the data misfit term, our misfit definition is composed of two penalty terms: the Tikhonov term to ensure smoothness, and the prior model term to help the convergence toward expected models. Generally, this latter prior penalty term is not used in classical FWI implementation, but we show that adding this information reduces the nonuniqueness issue of the inverse problem that is a well-known difficulty of the full waveform inversion. This prior information can be deduced from nonseismic data, well-logging, and geologic constraints that are generally available for specific exploration applications and for monitoring during production. We show with that this prior information improves the well-posedness of the problem as compared to the standard FWI approaches, and allows to partially mitigate potential kinematic inaccuracy of the starting model as well as illumination issues. The prior weighting operators and the prior model require an appropriate design. One has to properly balance the prior model term and the data term during the inversion. We have shown that balancing both pieces of information well is crucial during the optimization process. The prior model definition may include structures that can help drive the inversion toward the global minimum valley without being kinematically as accurate as the initial model for wave propagation. Moreover, these structures may prevent convergence in the final steps of the inversion and one could wish to decrease the importance of this prior information with respect to the data information. We have proposed a dynamic weighting of the prior term during the

inversion to smoothly reduce the impact of the prior information, leaving the floor at the end only to the data itself (regardless of the smooth Tikhonov term).

During this investigation, we have shown the striking differences between the roles of the initial model and the prior model in this constrained FWI framework where generally only partial information is available: The initial model must respect the wave equation and the related kinematic features to be positioned in the global minimum valley of the misfit function; whereas the prior model does not have such restrictive obligations like the initial model. The construction of the initial model is quite delicate whereas the construction of the prior model could allow more freedom.

Future investigations will be focused on using prior model for time-lapse applications to accurately obtain the physical parameter variations in a target zone. The design of the prior model for more complex environments and real data applications should also consider geostatistical approaches and/or standard quantitative interpretation techniques to honor the geologic structures.

ACKNOWLEDGMENTS

We would like to thank TOTAL Exploration & Production and SEISCOPE consortium for supporting this study. This work was performed with access to the high-performance computing facilities of CIMENT (Université de Grenoble) and to the HPC resources of GENCI-CINES under Grant 2011-046091. We acknowledge these facilities and the support of their staff. We would like to gratefully thank the Associate Editor, Aria Abubakar, three anonymous reviewers, and Partha Routh, for their very constructive comments on the manuscript.

REFERENCES

Abubakar, A., W. Hu, T. M. Habashy, and P. M. van den Berg, 2009, Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data: Geophysics, **74**, no. 6, WCC47–WCC58, doi: 10.1190/1.3250203.

Berenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200, doi: 10.1006/jcph.1994.1159.

Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: Geophysics, **74**, no. 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**, 1457–1473, doi: 10.1190/1.1443880.

Byrd, R., P. Lu, and J. Nocedal, 1995, A limited memory algorithm for bound constrained optimization: SIAM Journal on Scientific and Statistical Computing, **16**, 1190–1208, doi: 10.1137/0916069.

Cohen, J. K., and J. J. W. Stockwell, 2008, CWP/SU: Seismic Unix release no. 41: An open source software package for seismic research and processing, Center for Wave Phenomena, Colorado School of Mines.

Farquharson, C. G., and D. W. Oldenburg, 1998, Non-linear inversion using general measures of data misfit and model structure: Geophysical Journal International, **134**, 213–227, doi: 10.1046/j.1365-246x.1998.00555.x.

Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2010, Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle: Earth and Planetary Science Letters, **290**, 270–280, doi: 10.1016/j.epsl.2009.12.003.

Greenhalgh, S., B. Zhou, and A. Green, 2006, Solutions, algorithms and inter-relations for local minimization search geophysical inversion: Journal of Geophysics and Engineering, **3**, 101–113, doi: 10.1088/1742-2132/3/2/001.

Guitton, A., 2011, A blocky regularization scheme for full waveform inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2418–2422.

Guitton, A., G. Ayeni, and G. Gonzales, 2010, A preconditioning scheme for full waveform inversion: 80th Annual International Meeting, SEG, Expanded Abstracts, 1008–1012.

Herrmann, F. J., Y. A. Erlangga, and T. T. Y. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, no. 4, A35–A40, doi: 10.1190/1.3115122.

Hu, W., A. Abubakar, and T. M. Habashy, 2009, Simultaneous multifrequency inversion of full-waveform seismic data: Geophysics, **74**, no. 2, R1–R14, doi: 10.1190/1.3073002.

Jin, S., R. Madariaga, J. Virieux, and G. Lambaré, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, 575–588, doi: 10.1111/gji.1992.108.issue-2.

Li, Y., and D. W. Oldenburg, 1998, 3-D inversion of gravity data: Geophysics, **63**, 109–119, doi: 10.1190/1.1444302.

Loris, I., H. Douma, G. Nolet, I. Daubechies, and C. Regone, 2010, Nonlinear regularization techniques for seismic tomography: Journal of Computational Physics, **229**, 890–905, doi: 10.1016/j.jcp.2009.10.020.

Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: The Leading Edge, **25**, 156–166, doi: 10.1190/1.2172306.

Mead, J. L., and R. A. Renaut, 2009, A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems: Inverse Problems, **25**, 025002, doi: 10.1088/0266-5611/25/2/025002.

Miller, C. R., P. S. Routh, T. R. Brosten, and J. P. McNamara, 2008, Application of time-lapse ERT imaging to watershed characterization: Geophysics, **73**, no. 3, G7–G17, doi: 10.1190/1.2907156.

Oldenburg, D. W., 1994, Practical strategies for the solution of large-scale electromagnetic inverse problems: Radio Science, **29**, 1081–1099, doi: 10.1029/94RS00746.

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: Journal of Geophysical Research, **111**, doi: 10.1029/2005JB003835.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503, doi: 10.1111/gji.2006.167.issue-2.

Plessix, R. E., and W. A. Mulder, 2004, Frequency domain finite difference amplitude preserving migration: Geophysical Journal International, **157**, 975–987, doi: 10.1111/gji.2004.157.issue-3.

Plessix, R. E., and W. A. Mulder, 2008, Resistivity imaging with controlled-source electromagnetic data: Depth and data weighting: Inverse Problems, **24**, 034012, doi: 10.1088/0266-5611/24/3/034012.

Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seismometer data set: First Break, **28**, 71–78.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I: Theory and verification in a physic scale model: Geophysics, **64**, 888–901, doi: 10.1190/1.1444597.

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: Geophysical Journal International, **187**, 1495–1515, doi: 10.1111/gji.2011.187.issue-3.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2012, Two-dimensional anisotropic visco-elastic full waveform inversion of wide-aperture 4C OBC data from the Valhall Field: Presented at the 74th Annual International Conference and Exhibition, EAGE.

Routh, P. S., and P. D. Anno, 2008, Time-lapse noise characterization by inversion: 68th Annual International Meeting, SEG, Expanded Abstracts, 3143–3147.

Routh, P. S., and D. W. Oldenburg, 1999, Inversion of controlled source audio-frequency magnetotellurics data for a horizontally layered earth: Geophysics, **64**, 1689–1697, doi: 10.1190/1.1444673.

Rudin, L., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: Physica D, **60**, 259–268, doi: 10.1016/0167-2789(92)90242-F.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, 65–70.

Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2010, Seismic tomography of the Southern California crust based on spectral-element and adjoint methods: Geophysical Journal International, **180**, 433–462, doi: 10.1111/gji.2010.180.issue-1.

Tarantola, A., 1984, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015, doi: 10.1111/gpr.1984.32.issue-6.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Tikhonov, A., and V. Arsenin, 1977, Solution of ill-posed problems: Winston.

van den Berg, P., A. Abubakar, and J. Fokkema, 2003, Multiplicative regularization for contrast profile inversion: Radio Science, **38**, 23.1–23.10, doi: 10.1029/2001RS002555.

van den Berg, P. M., A. L. van Broekhoven, and A. Abubakar, 1999, Extended contrast source inversion: Inverse Problems, **15**, 1325–1344, doi: 10.1088/0266-5611/15/5/315.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.