GEOPHYSICS, VOL. 81, NO. 4 (JULY-AUGUST 2016); P. U25–U38, 11 FIGS.
10.1190/GEO2015-0349.1

A new parameter set for anisotropic multiparameter full-waveform inversion and application to a North Sea data set

Nuno V. da Silva[1], Andrew Ratcliffe[2], Vetle Vinje[3], and Graham Conroy[2]

ABSTRACT

Parameterization lies at the center of anisotropic full-waveform inversion (FWI) with multiparameter updates. This is because FWI aims to update the long and short wavelengths of the perturbations. Thus, it is important that the parameterization accommodates this. Recently, there has been an intensive effort to determine the optimal parameterization, centering the fundamental discussion mainly on the analysis of radiation patterns for each one of these parameterizations, and aiming to determine which is best suited for multiparameter inversion. We have developed a new parameterization in the scope of FWI, based on the concept of kinematically equivalent media, as originally proposed in other areas of seismic data analysis. Our analysis is also based on radiation patterns, as well as the relation between the perturbation of this set of parameters and perturbation in traveltime. The radiation pattern reveals that this parameterization combines some of the characteristics of parameterizations with one velocity and two Thomsen's parameters and parameterizations using two velocities and one Thomsen's parameter. The study of perturbation of traveltime with perturbation of model parameters shows that the new parameterization is less ambiguous when relating these quantities in comparison with other more commonly used parameterizations. We have concluded that our new parameterization is well-suited for inverting diving waves, which are of paramount importance to carry out practical FWI successfully. We have demonstrated that the new parameterization produces good inversion results with synthetic and real data examples. In the latter case of the real data example from the Central North Sea, the inverted models show good agreement with the geologic structures, leading to an improvement of the seismic image and flatness of the common image gathers.

INTRODUCTION

Interest in full-waveform inversion (FWI) as a method for estimating parameters related to wave propagation in the subsurface has increased considerably within the hydrocarbon exploration industry in the past few years (Krebs et al., 2009; Sirgue et al., 2009; Plessix et al., 2012, Warner et al., 2013). The aim is that FWI will make it possible to estimate these parameters in a semiautomatic fashion, decreasing intensive interaction with the data dramatically. We describe this as a semiautomatic process because, even though the FWI process itself is automatic, it still requires finding a suitable starting model and some, even if minimal, data preprocessing.

The concept of determining a velocity model by fitting the entire recording time was introduced more than three decades ago (Lailly, 1983; Tarantola, 1984). Nonetheless, several factors hindered practical applications of FWI. Examples of such factors are the available computational power, the quality of the data, and the need for a good starting model. The latter is related to the nonlinear relationship between data and model parameters, and it was studied during the early development of FWI with the use of multiscale approaches (Bunks et al., 1995; Sirgue and Pratt, 2004). More recently, other techniques have been introduced to mitigate this problem, and in particular the use of dynamic warping (Ma and Hale, 2013), registration (Baek et al., 2014), tomographically enhanced FWI (Tang et al., 2013), wavefield reconstruction inversion (van Leeuwen and Herrmann, 2013), deconvolution-based objective function (Luo and Sava, 2011), adaptive waveform inversion (Warner and Guasch, 2014), phase derivative (Choi and Alkhalifah, 2013), phase

---

First presented at the SEG 84th Annual International Meeting. Manuscript received by the Editor 29 June 2015; revised manuscript received 17 March 2016; published online 6 June 2016.

[1]Formerly CGG, Crawley, UK; presently Imperial College London, London, UK. E-mail: n.vieira-da-silva@imperial.ac.uk.
[2]CGG, Crawley, UK. E-mail: andrew.ratcliffe@cgg.com.
[3]CGG, Oslo, Norway. E-mail: vetle.vinje@cgg.com.

unwrapping (Choi and Alkhalifah, 2015), and combining seismic tomography with FWI (Alkhalifah and Choi, 2014).

However, the most widely used workflow still relies upon the use of seismic tomography techniques to determine a good starting model for FWI that explains the kinematics of the seismic data and allows less than half a cycle skip at the frequency that is being inverted (or at least the starting frequency in the inversion).

The existence of heterogeneities with spatial scales that are much smaller than the smallest wavelength component (one fourth of the wavelength is the Rayleigh criterion) in the propagating energy is responsible for the dependence of wave propagation on direction, or seismic anisotropy (Backus, 1962). This physical phenomenon must be addressed when computing seismic waveforms to determine traveltimes with sufficient accuracy. If seismic anisotropy is not considered, the estimation of velocity from seismic data will then be biased, as the anisotropic effects in the data will be estimated as perturbations in the velocity. Consequently, this can potentially lead to incorrect imaging of the subsurface.

Anisotropic FWI can be approached in several ways. The most widely used method consists of including the anisotropy in the wave propagator to account for correct kinematics and invert for velocity only (Warner et al., 2013). This type of approach requires a good estimate of Thomsen's parameters ε and δ (Thomsen, 1986). This method by itself circumvents a few aspects of jointly estimating parameters, in particular, different sensitivities to different parameters and coupling between the parameters. However, it also imposes a constraint in the inversion because Thomsen's parameters are not allowed to change throughout the inversion. Thus, the inversion scheme aims to determine a velocity model that explains the data for given models of anisotropy.

Some recent efforts have been focused on devising strategies for updating the anisotropy parameters together with the P-wave velocity (Plessix and Cao, 2011; Gholami et al., 2013a, 2013b; Wang et al., 2013; Cheng et al., 2014; Stopin et al., 2014). Incorporating the estimation of the velocity and anisotropy parameters together in the inversion scheme, in principle, allows to improve the data fitting because of the extra degrees of freedom available compared with fixing the anisotropy parameters and also to determine an improved model by removing some of the bias imposed by the fixed nature of the anisotropy parameters. However, a coupling between the parameters may exist and the null space is also much larger, potentially hindering the convergence to artifact-free models.

The use of second-order optimization techniques, in which the off-diagonal elements of the Hessian are considered (Métivier et al., 2014), mitigates the coupling between the parameters as it accounts for the interaction between two parameters. However, second-order methods can be computationally intensive and, for this reason, most of the inversion schemes reported in the literature still use first-order schemes to keep the algorithms feasible, especially when dealing with 3D geometries.

As pointed out by Gholami et al. (2013a), choosing an adequate parameterization can potentially mitigate the coupling between the parameters. An adequate parameterization in this case is the one in which the scattered energy radiates from the perturbation in different directions. However, by doing so, this leads to different spatial scales of the reconstructed parameters (Operto et al., 2013). This means that there is a difference between the characteristic wavelength of perturbation of different parameters. From this, one can conclude that there is a dichotomy between the coupling and also the spatial scale of the perturbations that can be reconstructed from the data.

The problem of choosing a suitable parameterization has been addressed from different perspectives. For example, Plessix and Cao (2011) investigate this problem by decomposing the Hessian matrix into its eigenvalues and eigenvectors to determine a suitable parameterization. The key aspects pointed out from this work are that δ is poorly constrained from surface seismic data and that inverting for the normal moveout (NMO) velocity, v_n, and $\eta = (\varepsilon - \delta)/(1 + 2\delta)$ is sufficient to describe correctly the kinematics of a medium with vertical transverse isotropy (VTI). On the other hand, Gholami et al. (2013a) investigate numerically the radiation patterns, namely the response of a scattered wave with varying angle, for different parameterizations. The main conclusion is that the parameters can be grouped in classes depending on the aims of the inversion and that, in general, parameterizations should be chosen carefully depending on the spread of the recorded data, the resolution required for a given parameter, and the mitigation of the trade-off between the parameters. Another key aspect is that inverting directly for the elements of the tensor of elasticity leads to poor results. Alkhalifah and Plessix (2014) conduct this investigation by deriving radiation patterns analytically from the Born approximation. This work determined that a parameterization based on (v_n, η, δ) is the most suitable for inverting diving and reflected waves. With this parameterization, v_n and η are determined from diving waves and δ is determined from reflected energy, fitting mainly the amplitude. Further, this work also concluded that in hierarchical schemes, (v_h, η, ε) allows to mitigate the coupling between the parameters, whereas ε aims to fit the amplitude. One can conclude that there are different parameterizations that can lead to equally good results, and all previous results are in agreement.

Here for the first time, we investigate the use of an alternative parameterization based on the concept of kinematic equivalent media (Stovas, 2008). The recorded waveforms result from the complex response of the medium, such as reflections, diffractions, and refractions to enumerate a few. Hence, recorded phases do not contain purely information on the traveltime. However, in the case of diving and transmitted waves, this relation is much more linear because interference phenomena in the wavefront are much less complex than in the case of reflected waves, as the latter can result from strong scattering. Further, assumptions associated to the description of acoustic media impose some limitations in the correct description of the dynamics of wave propagation, as the elastic effects are not correctly computed. One of such limitations is, for instance, taking into account the effect of density. This can be addressed by jointly estimating velocity and density. However, such an approach also has inherent drawbacks similar to the estimation of velocity and anisotropy (Prieux et al., 2013). The fact that the source wavelet is also generally unknown is a further determining factor limiting the simulation of the data with the correct amplitude. For these reasons, the tomographic mode of FWI, in which mainly diving waves are inverted, is central for making FWI feasible (as conventionally formulated).

Because the kinematics of the waves are central in acoustic FWI, this suggests that the information on traveltime is also essential, as both perturbations in velocity and anisotropy are associated to perturbations in traveltime, which means that this is encapsulated in different parameters, thus determining the existence of an ambiguity. That means that, for a given parameterization, different numerical combinations of different parameters can lead to the same traveltime. Hence, a parameterization that relates unambiguously the model parameters and the traveltime information potentially has good characteristics for carrying out inversion and also anisotropic FWI with multiparameter updates. It is important to note that such sets do not guarantee the uniqueness of solution (Grechka, 2009); however, if the velocity is determined correctly for example, this implies that the parameter for anisotropy is determined with less ambiguity.

Here, we use the parameter set for kinematic equivalent media as introduced by Vinje et al. (2013) for smoothing velocity models while preserving traveltime. Because these parameters were introduced in the context of preserved traveltime smoothing (PTS), we use this acronym throughout this paper when referring to this set of parameters, even though the specific properties of traveltime preservation and smoothing are not relevant in the context of this work.

The paper is organized as follows: First, we review the equations for acoustic waves in VTI media. We then introduce our new parameter set, discussing its properties using radiation patterns and sensitivity to traveltime, followed by a demonstration on a synthetic example. Finally, we show an application of this new parameterization to a real marine data set acquired in the North Sea, containing an outline of the methodology, results, and geologic interpretation.

THEORY

FWI aims to fit the entire recorded data $\boldsymbol{d}$ to the simulated data $\boldsymbol{u}$, and this is commonly carried out by the minimization of the L_2-norm of the data misfit (Lailly, 1983; Tarantola, 1984):

$$J(\boldsymbol{u},\boldsymbol{d},\boldsymbol{m}) = \frac{1}{2}\sum_{t,r,s}\|u_{t,r,s} - d_{t,r,s}\|_2^2, \tag{1}$$

which is obtained iteratively by a steepest-descent method (Kelley, 1999), preconditioned with the diagonal of the Hessian:

$$\boldsymbol{m}_k = \boldsymbol{m}_k - \alpha_k \boldsymbol{B}_k \nabla_m J, \tag{2}$$

where the subscripts t, r, and s denote time, receiver, and source positions, respectively; $\boldsymbol{m}$ represents the model parameter(s) that are to be estimated; and k denotes the iteration number. The operator $\boldsymbol{B}$ is an approximation of the inverse of the diagonal of the Hessian of the objective function and scales the gradient in space. The scaling of the gradient of the objective function is particularly important to balance the updates appropriately away from the source and receiver locations.

The wavefields are computed using the VTI wave equation (Duveneck et al., 2008; Zhang et al., 2011), which in its compact form can be written as

$$\begin{cases} \frac{1}{v_0^2}\partial_{tt}p_h - \rho(1+2\varepsilon)\nabla_h\left(\frac{1}{\rho}\nabla_h p_h\right) - \rho\sqrt{1+2\delta}\,\partial_z\left(\frac{1}{\rho}\partial_z t_n\right) = s, \\ \frac{1}{v_0^2}\partial_{tt}t_n - \rho\sqrt{1+2\delta}\,\nabla_h\left(\frac{1}{\rho}\nabla_h p_h\right) - \rho\,\partial_z\left(\frac{1}{\rho}\partial_z t_n\right) = s' \end{cases} \tag{3}$$

where $\nabla_h = (\partial_x, \partial_y)$, p_h is the horizontal stress, t_n is the vertical stress, ρ is the density, v_0 is the P-wave (vertical) velocity, s and s' are the source terms, and ε and δ are the anisotropy parameters. For a complete formulation, the wave equation obeys the initial conditions

$$p_h(\boldsymbol{x},t=0) = \partial_t p_h(\boldsymbol{x},t=0) = t_n(\boldsymbol{x},t=0) = \partial_t t_n(\boldsymbol{x},t=0) = 0, \tag{4}$$

and the boundary condition

$$p_h(\boldsymbol{x} \in \partial\Omega, t) = t_n(\boldsymbol{x} \in \partial\Omega, t) = 0 \tag{5}$$

is imposed, where $\partial\Omega$ represents the boundary of the domain. The wave equation 3 is solved with a fourth-order finite-difference scheme in time (Etgen, 1986) and a high-order finite-difference scheme in space that computes derivatives that are accurate to 92% of the Nyquist wavenumber (Zhang et al., 2011). The discrete form of the system of equations 3 is written symbolically as

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{s}, \tag{6}$$

where $\boldsymbol{A}$ is the discrete wave operator, $\boldsymbol{u}$ is the discrete wavefield (or state variable), and $\boldsymbol{s}$ is the discrete source term. The gradient of the objective function in equation 1 is computed with the adjoint-state method (Lions, 1971; Fichtner et al., 2006; Plessix, 2006; Chavent, 2010), and it is given by

$$\nabla_m J(\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{m}) = \boldsymbol{\lambda}^T[\nabla_m \boldsymbol{A}]\boldsymbol{u}, \tag{7}$$

where $\boldsymbol{\lambda}$ is the adjoint variable in the discrete form and the symbol T denotes the matrix transpose operation. This variable is computed with the same time-marching scheme used for the computation of the state variable $\boldsymbol{u}$. The wave propagator uses the Thomsen's parameters, thus the gradient with respect to the anisotropy parameters is computed using the chain rule. Appendix A contains a comprehensive demonstration on how the expression for the gradient of the objective function is computed and on how the time-marching scheme is implemented for computing the gradient.

NEW PARAMETERIZATION

Here, we discuss the new parameterization and some of its more relevant properties in the scope of FWI. This is done using radiation patterns and sensitivity to traveltime. The relation between the kinematically consistent parameters, or PTS as denoted throughout the text, denoted by η_{-1}, η_1, and η_3, and Thomsen's parameters (Stovas, 2008; Vinje et al., 2013) is

$$\begin{aligned} \eta_{-1} &= v_0^{-1} \\ \eta_1 &= v_0(1+2\delta) = v_n^2/v_0 \\ \eta_3 &= v_0^3(1+2\delta)(1+8\varepsilon-6\delta) = v_n^4(1+8\eta)/v_0, \end{aligned} \tag{8}$$

and the reciprocal relation is

$$\begin{aligned} v_0 &= \eta_{-1}^{-1} \\ \delta &= \frac{1}{2}(\eta_{-1}\eta_1 - 1) \\ \varepsilon &= \frac{1}{8}\left(\frac{\eta_3\eta_{-1}^2}{\eta_1} + 3\eta_1\eta_{-1} - 4\right) \\ v_n^2 &= \frac{\eta_1}{\eta_{-1}} \\ \eta &= \frac{1}{8}\left(\frac{\eta_3\eta_{-1}}{\eta_1^2} - 1\right). \end{aligned} \tag{9}$$

It is important to note that the parameters are represented by the variable η with a subscript; however, this should not be confused with $\eta = (\varepsilon - \delta)/(1 + 2\delta)$, often seen in the context of anisotropy and first introduced by Alkhalifah (1998). One can immediately identify η_{-1} as the vertical slowness (in a VTI medium), η_1 as the vertical velocity scaled by δ (similar to v_n), and η_3 as a blend of the contributions of v_0, ε, and δ. These parameters are the integrands of the velocity moments, which encapsulate the combined effect of anisotropy and vertical velocity heterogeneity.

It is recognized that δ is poorly constrained from surface seismic data. In addition, if it does not change significantly laterally, then its effect is essentially a vertical stretch (Alkhalifah and Tsvankin, 1995; Plessix and Cao, 2011). Due to this ambiguity between δ and depth, it is generally chosen not to update this parameter, as reported in previous work using (v_n, η, δ) for parameterizing the medium (Stopin et al., 2014). Here, we follow the same strategy, and we rely upon a good initial estimation of δ, not updating it throughout the inversion. For this reason when carrying out the inversion, the only parameters that are updated are η_{-1} and η_3 because these are sufficient to reconstruct v_0 and ε.

RADIATION PATTERNS

Radiation patterns provide valuable insights on the existence, or not, of coupling between the parameters, and strength of the sensitivity of the data to each one of them, with the changing scattering angle. They can be computed numerically (Gholami et al., 2013a) or analytically (Alkhalifah and Plessix, 2014), and in both cases the assumptions of the Born approximations are used. When the scattered energy corresponding to perturbations in two different parameters is being radiated over an overlapping region of scattering angles, then coupling between the parameters occurs in that region. The scattering angle determines the scale of the perturbations that can be reconstructed from the data according to (Wu and Toksöz, 1987; Aki and Richards, 2002):

$$\boldsymbol{k} = 2k_0 \cos\frac{\theta_s}{2}\boldsymbol{n}, \tag{10}$$

where $\boldsymbol{k}$ is the vertical wavenumber component of the anomaly, $\boldsymbol{n}$ is a unit vector normal to the reflector, k_0 is the wavenumber of the background medium, and θ_s is the scattering angle. If two given parameters are not coupled, then the regions of their scattering angles do not overlap. Parameters which are decoupled in their scattering regions are ideal, in the sense that they present minimal or no coupling. Nonetheless because they are estimated with different resolution, as determined from equation 10, this can lead to subsequent incompatibilities when trying to determine a third parameter as a result of combining the first two. For example, Operto et al. (2013) demonstrate that this issue potentially leads to higher inaccuracy of the inverted models than due to the existence of the coupling.

Because diving waves are important to FWI to specifically constrain the long-wavelength components of the model, we aim to determine which parameterizations are more suitable to be compared with the introduced PTS parameters. The radiation pattern of such parameterizations is characterized by having radiated energy at the larger scattering angles, related to one parameter (at least) that determines the kinematics of wave propagation. Considering the previous published work and references in the beginning of this section, examples of such parameterizations are $(v_0, \delta, \varepsilon)$, (v_n, η, δ), (v_0, δ, v_h), and (v_n, δ, v_h). Figure 1a–1d shows the radiation patterns for each one of these parameterizations respectively, which are included here for a matter of completeness. As it can be immediately observed, and pointed out by Gholami et al. (2013a), parameterizations which are composed of one velocity and two parameters for anisotropy present coupling between the parameters, whereas parameterizations formed with two velocities and one parameter of anisotropy show decoupling between at least two parameters. As mentioned earlier, the latter property is undesired from the perspective of consistency of the wavelength of the reconstructed perturbations and, for this reason, this type of parameterization is not ideal for use in FWI. When the trade-off is strong at the largest angles, it is critical that the starting model for the anisotropy is sufficiently accurate to minimize errors.

The novelty introduced here is the radiation pattern for the PTS parameters. First, equation 3 is transformed into an equivalent expression that couples vertical pressure and its deviation generated by the existence of anisotropy as in Zhou et al. (2006) and Plessix and Cao (2011). If the medium is isotropic, the anomalous part is zero and one obtains the wave equation for isotropic media. The normal part is computed from the Green's function for the background medium, which is homogeneous and isotropic, and the anomalous part is computed directly from the Born approximation. This is accomplished by the transformation $t_n = p_n/\sqrt{1+2\delta}$, corresponding to a vertical stretch (Alkhalifah et al., 2001) yielding

$$\begin{cases} \frac{1}{v_0^2}\partial_{tt}p_h - \rho(1+2\varepsilon)\nabla_h\left(\frac{1}{\rho}\nabla_h p_h\right) - \rho\sqrt{1+2\delta}\,\partial_z\left(\frac{1}{\rho}\partial_z\frac{1}{\sqrt{1+2\delta}}p_n\right) = \mathrm{s}, \\ \frac{1}{v_0^2}\partial_{tt}p_n - \rho(1+2\delta)\nabla_h\left(\frac{1}{\rho}\nabla_h p_h\right) - \rho\sqrt{1+2\delta}\,\partial_z\left(\frac{1}{\rho}\partial_z\frac{1}{\sqrt{1+2\delta}}p_n\right) = \mathrm{s}, \end{cases} \tag{11}$$

where s is the source term. Equation 11 can be further analyzed defining the pressure field as $p = p_n$ and its deviation $q = p_h - p_n$, leading to

$$\begin{cases} \frac{1}{v_0^2}\partial_{tt}p - \rho(1+2\delta)\nabla_h\left(\frac{1}{\rho}\nabla_h(p+q)\right) - \rho\sqrt{1+2\delta}\,\partial_z\left(\frac{1}{\rho}\partial_z\frac{1}{\sqrt{1+2\delta}}p\right) = \mathrm{s}, \\ \frac{1}{v_0^2}\partial_{tt}q - 2\rho(\varepsilon-\delta)\nabla_h\left(\frac{1}{\rho}\nabla_h(p+q)\right) = 0. \end{cases} \tag{12}$$

Equation 12 is divided by $1 + 2\delta$ and recast in terms of the PTS parameters, giving

$$\begin{cases} -\frac{\eta_{-1}}{\eta_1\rho}\omega^2 p - \nabla_h\cdot\frac{1}{\rho}\nabla_h(p+q) - \frac{1}{\sqrt{\eta_{-1}\eta_1}}\partial_z\frac{1}{\rho}\partial_z\frac{1}{\sqrt{\eta_{-1}\eta_1}}p = s, \\ -\frac{\eta_{-1}}{\eta_1\rho}\omega^2 p - \frac{1}{4}\left(\frac{\eta_{-1}\eta_3}{\eta_1^2} - 1\right)\nabla_h\cdot\frac{1}{\rho}\nabla_h(p+q) = 0. \end{cases} \tag{13}$$

For carrying out the analysis using the Born approximation, the fields are decomposed in primary and scattered components $p = p_0 + p_1$ and $q = q_0 + q_1$, where 0 in the subscript denotes the primary component and 1 denotes the scattered field. The medium parameters $(\eta_{-1}, \eta_1, \eta_3)$ are each decomposed into background and perturbation components

$$\begin{aligned}\eta_{-1} &= \eta_{0,-1}(1+r_{-1})\\ \eta_1 &= \eta_{0,1}(1+r_1)\\ \eta_3 &= \eta_{0,3}(1+r_3).\end{aligned} \tag{14}$$

In the background medium $r_{-1} = r_1 = r_3 = 0$, yielding $\eta_{0,-1} = v_0^{-1}$, $\eta_{0,1} = v_0$, and $\eta_{0,3} = v_0^3$. These definitions lead to an equation for an isotropic background medium (equation B-1 in Appendix B), from which p_0 and q_0 can be determined. The radiation pattern is computed by determining the shape of the envelope of the energy propagation for the scattered field, and the derivation of the formulas is presented in Appendix B. From the Born approximation, the scattered field p_1 is given by

$$p_1(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega) = -s(\omega)\omega^2 \int d\boldsymbol{x}\frac{G(\boldsymbol{x},\boldsymbol{x}_s,\omega)G(\boldsymbol{x},\boldsymbol{x}_r,\omega)}{v_0^2\rho}\boldsymbol{r}\cdot\boldsymbol{w}, \tag{15}$$

where the dot in equation 15 indicates the scalar dot product and

$$\boldsymbol{r} = \begin{pmatrix} r_{-1}\\ r_1\\ r_3\end{pmatrix} \quad\text{and}\quad \boldsymbol{w} = \begin{pmatrix} -1-\cos^2\frac{\theta_s}{2}+\frac{1}{4}\sin^2\frac{\theta_s}{2}\\ 1-\cos^2\frac{\theta_s}{2}-\frac{1}{2}\sin^2\frac{\theta_s}{2}\\ \frac{1}{4}\sin^4\frac{\theta_s}{2}\end{pmatrix}. \tag{16}$$

The vector $\boldsymbol{r}$ contains the perturbation of the PTS parameters as defined in equation 14, θ_s is the scattering angle, and $\boldsymbol{w}$ defines the radiation pattern for each one of the parameters. Figure 1e shows the radiation pattern for the PTS parameters.

The radiation pattern of the parameters η_1 and η_3 determines that these parameters can be sensed at the larger scattering angles, thus they are suitable for describing the kinematics of the medium, which is expected because these parameters are constructed specifically to have that property. The radiation pattern for η_{-1} shows that its associated scattered energy is radiated across the whole range of scattering angles. However, its strength is dimmed at the larger angles. It is important to note that this is not the same as stating that this parameter cannot be sensed at the larger angles, as in fact the strength of the scattered energy for η_{-1} is still stronger than the other two parameters. The same does not happen when parameterizing with two velocities and one parameter for anisotropy, where the strength of the energy envelope reaches zero. Contrary to the parameterizations $(v_0, \delta, \varepsilon)$ and (v_n, η, δ), which show the decoupling of at least one of the parameters of anisotropy, the radiation pattern of the PTS parameters shows that the anisotropy can only be estimated with low resolution; thus, it is the kinematics of the waves that mainly drives the inversion. This is ideal in acoustic inversion. However, trade-off exists between all the parameters at the wider angles. That means that it is necessary to have a correct model of the background for the anisotropy to minimize errors in the inverted model.

SENSITIVITY TO TRAVELTIME

FWI aims to fit all the events in the seismic trace, which is understood as fitting all the traveltimes and amplitudes resulting from excitation of the subsurface. As mentioned earlier, the traveltime information has a crucial role in the inversion of diving waves using FWI. For this reason, the study of the dependence of the sensitivity of traveltime with the perturbation in the model parameter space gives insight on the relation between traveltime and the combination of parameters. In this section, we discuss the coupling of the traveltime with the parameters describing the medium for the same parameterizations discussed in the section on radiation patterns. For simplicity and clarity, we define a model with one homogeneous layer with a compressional velocity, and a single reflector at a depth of 1000 m. The model parameters are $\eta_{-1} = 5.0\times 10^{-4}$ s/m, $\eta_1 = 2160$ m/s, and $\eta_3 = 1.00224\times 10^{10}(\text{m/s})^3$, which corresponds to a compressional velocity of 2000 m/s, $\delta = 0.04$, and $\varepsilon = 0.05$. The traveltimes are computed from (Stovas, 2008)

$$t^2(x) = t_0^2 + \frac{x^2}{v_n^2} + \frac{(1-S_2)x^4}{4v_n^4}, \tag{17}$$

where

$$\begin{aligned}t_0 &= \frac{2z}{v_0}\\ v_n^2 &= v_0^2(1+2\delta)\\ S_2 &= 1+\frac{8(\varepsilon-\delta)}{1+2\delta} = 1+8\eta.\end{aligned} \tag{18}$$

The traveltime misfit is quantified by the sum over the receivers of the square of the difference between the traveltime computed for the base model and the perturbed models:

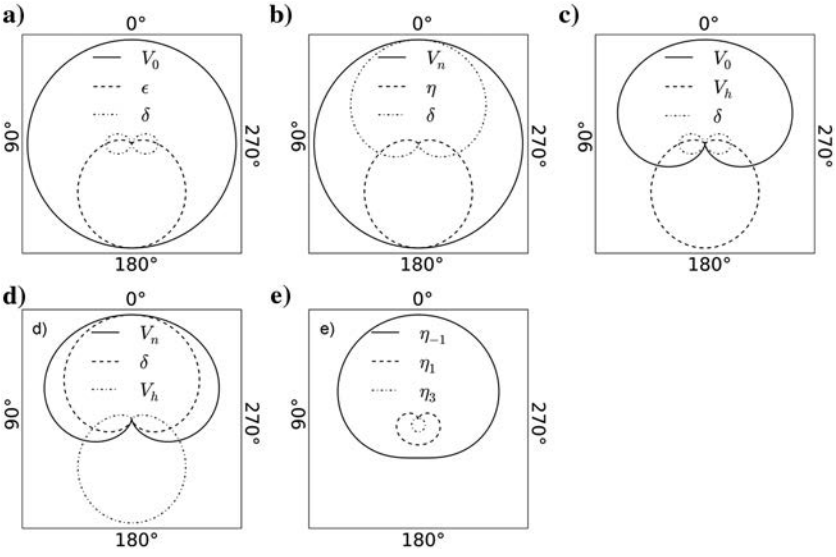

Figure 1. Radiation patterns for (a) $(v_0, \delta, \varepsilon)$, (b) (v_n, η, δ), (c) (v_0, δ, v_h), (d) (v_n, δ, v_h), and (e) the PTS parameters, assuming a horizontal reflector.

$$E = \sum_i (t_{p,i} - t_{0,i})^2, \tag{19}$$

where $t_{p,i}$ and $t_{0,i}$ are the traveltimes at the ith receiver for the perturbed and base medium, respectively. The minimum offset for the receiver position is 200 m, and our comparison is made for two different values of the maximum offset 1500 and 5000 m. This aspect is important because the influence of ε on the data will depend upon the propagation path of the energy, and generally the effect of the anisotropy is more pronounced in the data recorded at the largest offsets.

A practical question in this analysis is the choice of the scan range for each parameter in each parameter set investigated. Here, we are left with a dilemma: Do we choose a specific percentage scan for a given parameter set and propagate the upper and lower bounds of these parameters to the other parameter sets, or do we just scan a specific percentage of each parameter in each parameter set individually? We acknowledge there are advantages and disadvantages in both approaches, but we must make a choice. Hence, here the true v_0 and ε parameters are perturbed between ±4%, and the respective upper and lower bounds of the other parameterizations are computed to ensure that the same space of parameters is being investigated. Finally, for the reasons mentioned before, δ is not updated, thus the update of η_1 is not necessary, and the sensitivity study in the PTS parameters is based upon the perturbation of η_{-1} and η_3 only.

Figure 2 depicts the effect of the perturbation of the model parameters on the traveltime. In the left column we show the plots of the misfit with model perturbation using the data simulated up to an offset of 1.5 km, and the right column shows the plots of the misfit with model perturbation using the data simulated up to an offset of 5 km. It is clear that at short offsets, the perturbation of the parameters related to anisotropy have no effect on the traveltime because there is a range along which the traveltime does not change. This is true for all the parameterizations investigated here. At the larger offsets, the behavior of the misfit function changes substantially, where clearly both varying parameters affect the misfit. Thus, the analysis of the misfit at the larger offsets provides the best information to determine which parameterization shows less ambiguity in terms of perturbation of traveltime versus perturbation in the parameters. By inspection of these misfit plots, one concludes that all the parameterizations tested, with the exception of PTS, show strong ambiguity. This means that there is a relatively wide combination of any two parameters that lead to the same misfit of traveltime. The parameterization with PTS is the only one (among the parameterizations tested) that does not show strong ambiguity when the velocity and anisotropy have a stronger effect in the data. This property is important as this allows, at least in principle, to constrain the combination of parameters that explains the data.

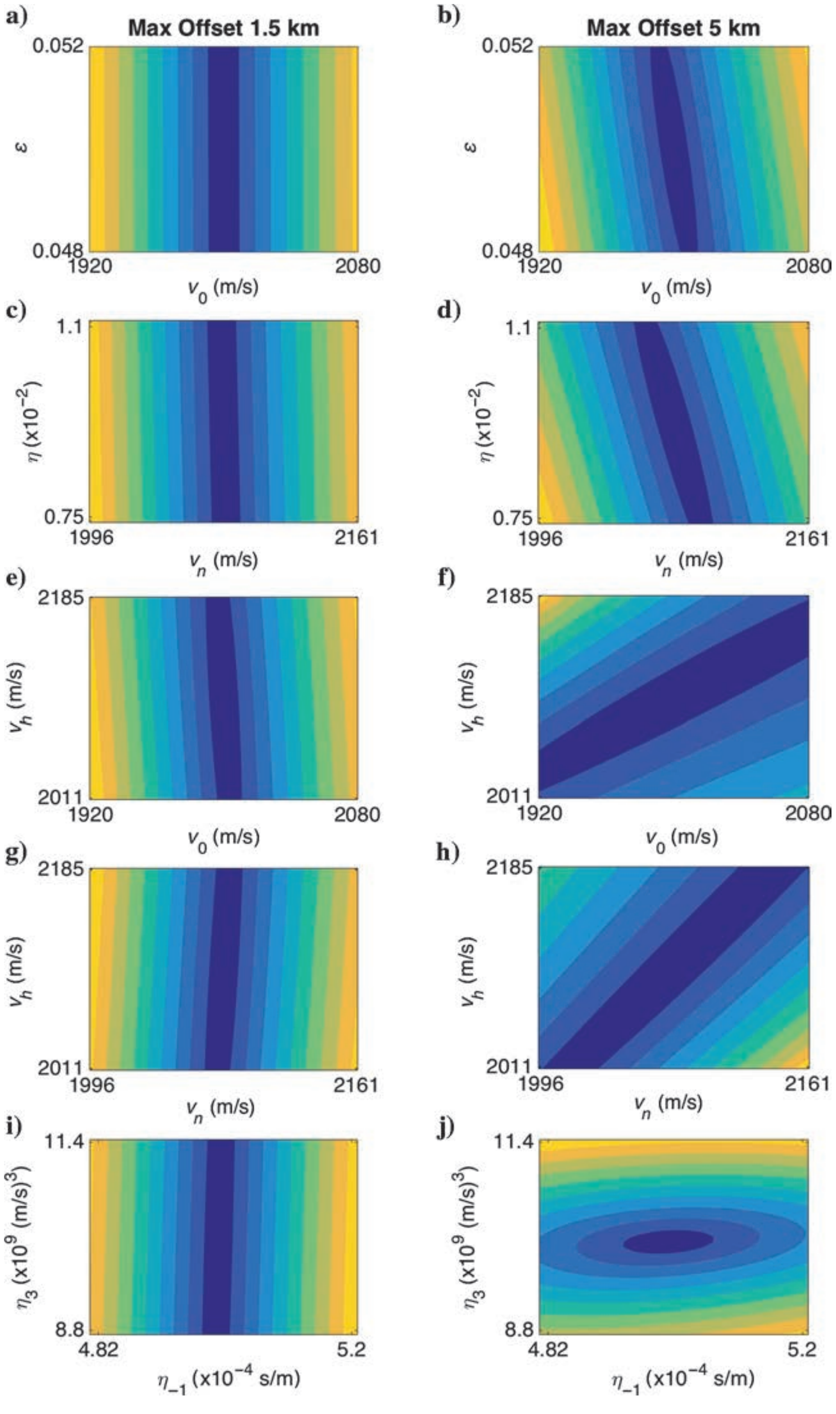

Figure 2. Plot of the misfit of traveltime with perturbation of (v_0, ε) for a maximum offset of (a) 1.5 and (b) 5 km; with perturbation of (v_n, η) for a maximum offset of (c) 1.5 and (d) 5 km; with perturbation of (v_0, v_h) for a maximum offset of (e) 1.5 and (f) 5 km; with perturbation of (v_n, v_h) for a maximum offset of (g) 1.5 and (h) 5 km; and with perturbation of (η_{-1}, η_3) for a maximum offset of (i) 1.5 and (j) 5 km. The color scale is normalized for each plot and ranges between 0 (blue) and 1 (yellow).

SYNTHETIC EXAMPLE

In this section, we present a synthetic example where an inversion is carried out using the new parameterization, and the results are compared with the ones obtained using (v_n, η, δ). The discussion in the previous section and previously published work indicates that the parameterizations based on two velocities and one Thomsen's parameter do not lead to the best inversion results, in general. In addition, as concluded by Plessix and Cao (2011) and Alkhalifah and Plessix (2014), a parameterization based on (v_n, η, δ) is the most suitable for describing VTI media as their kinematics is correctly represented with v_n and η. For this reason, we compare this synthetic example study with this parameterization only. It is expected that these two parameterizations should lead to similar inversion results, when not updating δ and using only diving waves. This is because the radiation patterns for both parameterizations have the highest sensitivity to ε at the largest incidence angles.

One of the issues when performing multiupdates is the fact that the parameters have different dynamic ranges. This dynamic range is even more pronounced in the case of the PTS parameters. Thus, the appropriate scaling of the parameters is crucial to improve the conditioning of the inversion and avoid that one parameter has a stronger influence on the inversion result than the other, thus potentially biasing the result of the inversion. In both parameterizations, the parameters are scaled by their background. Hence, in the case of the PTS parameters, this is obtained by $\tilde{\eta}_{-1}=(\eta_{-1}-\eta_{0,-1})/\eta_{0,-1}$ and $\tilde{\eta}_3=(\eta_3-\eta_{0,3})/\eta_{0,3}$. In the case of (v_n, η, δ), only the NMO velocity is scaled with $\alpha = (v_n - v_{0,n})/v_{0,n}$, as reported in Stopin et al. (2014).

The synthetic example discussed in this section is chosen in such a way that the diving and reflected waves are simulated. Nonetheless, the data are mainly dominated by diving waves because the background of the true vertical velocity model (Figure 4a) is a positive gradient of velocity with depth. The true model for ε follows the same structure as the velocity model (Figure 3b). Thus, if the anomalies are not coincident in space, it is likely that the errors in the estimated models increase, and these will be higher if the background of the starting model is incorrect.

The grid spacing here is chosen to be 50 m as well as the source and receiver spacing. The data are generated using a finite-bandwidth source wavelet, limited between 1.5 and 10 Hz, for a split-spread configuration, and the offset range is 10 km in each direction. The inversions are carried out in a multiscale fashion (Bunks et al., 1995) from 2 up to a maximum frequency of 10 Hz. All the inversions start from a smooth version of the true model for vertical velocity (Figure 3b) and ε (Figure 4b). As pointed out by Podgornova et al. (2015), data generated by acoustic scattering, which do not contain a DC compo-

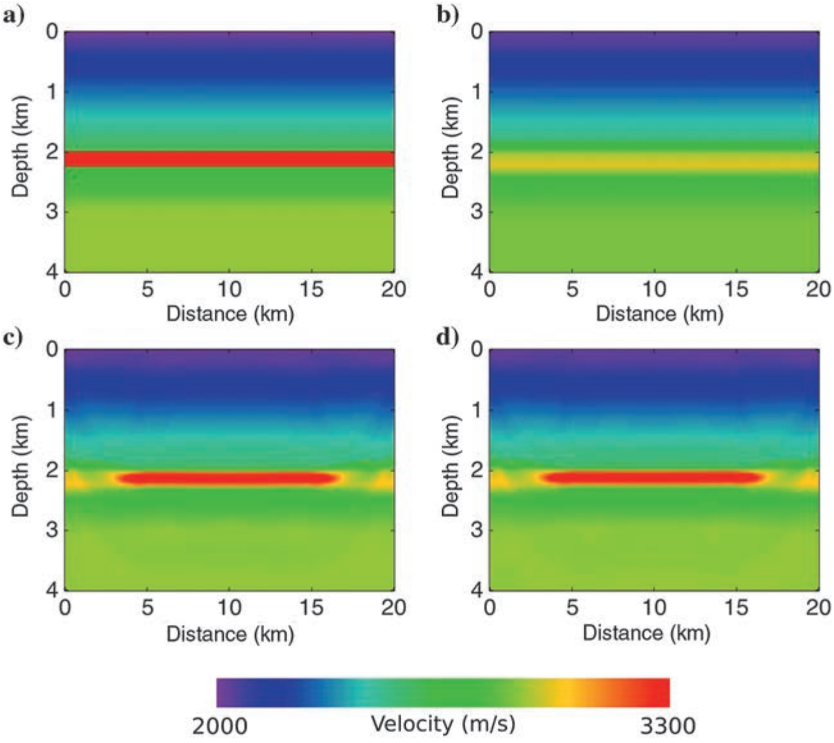

Figure 3. Vertical velocity models for the synthetic example: (a) true model, (b) starting model, (c) velocity model inverted using (v_n, η, δ), and (d) velocity model inverted using PTS.

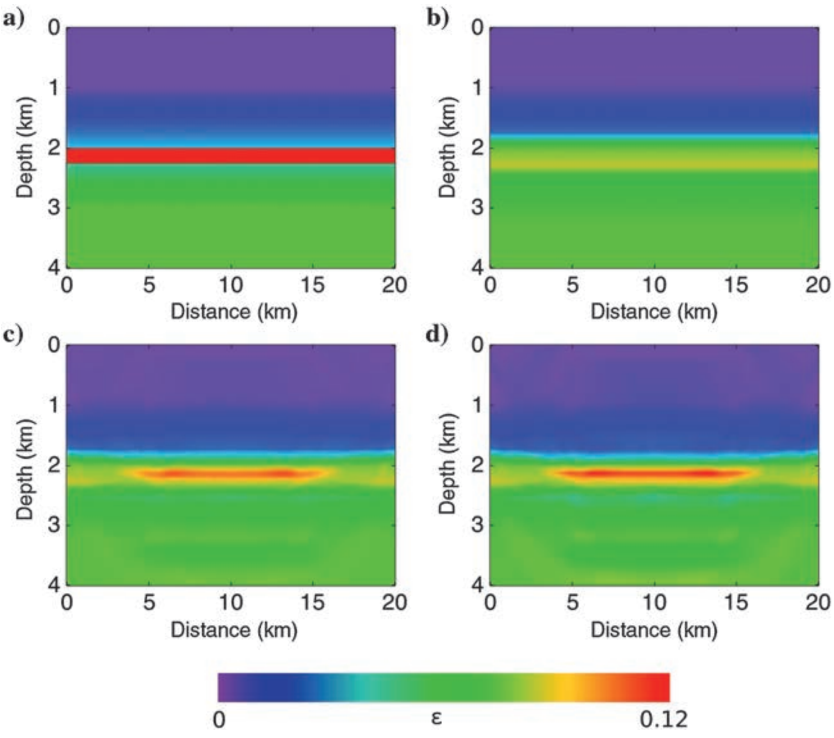

Figure 4. Synthetic models of ε: (a) true model, (b) starting model, (c) inverted ε model using (v_n, η, δ), and (d) inverted ε model using PTS.

nent, only have sensitivity to one parameter at long wavelengths. For this reason, it is crucial that the starting model for ε has these spatial scales well-represented.

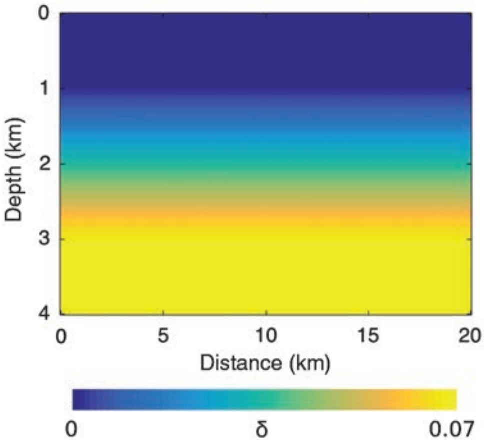

Figure 5. Synthetic model of the δ parameter. This model is used to generate the data and does not change throughout the inversion.

In this example, we only show the models for vertical velocity and ε because these are the parameters with which we parameterize our wave propagator, as in equation 3. The models in terms of the optimization parameters, either for (v_n, η, δ) or PTS, are not stored as these are converted on the fly to the parameters that are used by the wave propagator. This is straightforward to do because the gradients with respect to any given parameter(s) can be computed from the chain rule. As we are inverting the data using different parameterizations, it is important to compare the same quantities. For these reasons, we only outline the results in terms of vertical velocity and ε. Figure 3c shows the inverted velocity model using (v_n, η, δ), and Figure 3d shows the inverted velocity model using PTS. Both models are well-reconstructed, and the differences between both are negligible. The reflective anomaly is less well-reconstructed in the edges as a result of poor illumination from the finite-acquisition (source and receiver) aperture. The inverted models of ε using (v_n, η, δ) and PTS are depicted in Figure 4c and 4d, respectively. The anomaly is reconstructed using both parameterizations. Nonetheless, in both cases, the geometric shape is not as well-recovered, and the edges of the anomaly are not as sharp as in the case of the velocity model. Also, at the edges of the model, there is almost no update due to the poor coverage of sources and receivers, as in the case of the inverted velocity models. The model of ε reconstructed with the PTS parameters shows a stronger anomaly than the one inverted with (v_n, η, δ). Also some artifacts are visible in both models (Figure 4c and 4d) in the shallow part of the model. These are stronger in the model resulting from the PTS parameters. Because the reconstructed velocity models are very similar using both parameterizations, this suggests that the anomaly at the center of the inverted ε with PTS is not a result of stronger trade-off between the parameters. It is rather a result of the parameterization itself that captures the effect of anisotropy. This is expected as η_3 has a much larger range than η_{-1}. In addition, the radiation pattern for v_n covers the entire range of incident angles with the same strength, whereas the radiation pattern for η_{-1} shows less strength in the region of wider angles, thus the effect of anisotropy will be stronger in η_3. The parameter δ is chosen to be increasing with depth (Figure 5), and is not updated in the inversion. The δ model depicted in Figure 5 is the same used to generate and invert the synthetic data in this example.

This example shows that the PTS parameters are suitable for estimating anisotropy from surface recorded seismic data, and can compete with other parameterizations that have very good characteristics to address this problem, as in the case of (v_n, η, δ).

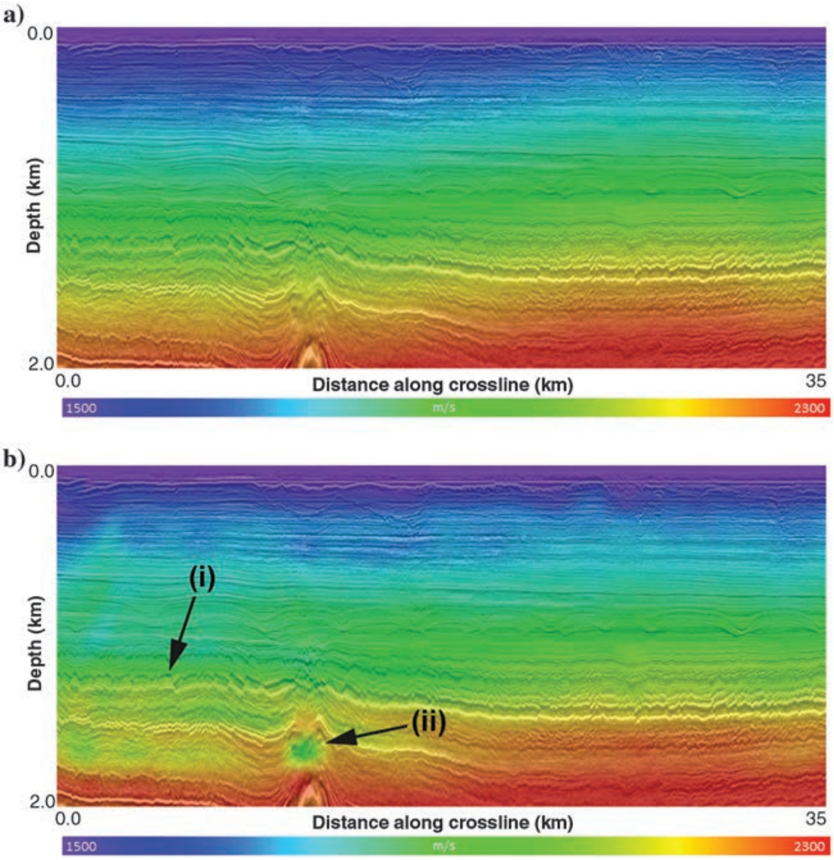

Figure 6. (a) Starting velocity model and (b) FWI inverted velocity model using the new PTS parameterization. Both models are overlaid with a stacked image generated from a Kirchhoff migration using the relevant model. We see very good agreement in the FWI model with geologic structures, such as the shallow channels in the first few hundred meters depth, the dewatering faults (indicated by (i)), and the low-velocity anomaly (indicated by (ii)) thought to be a small gas cloud above a salt diapir.

REAL DATA EXAMPLE

Geologic setting of the Central North Sea data set

The Central North Sea is characterized by the existence of recent glacial channels and in-filled canyon systems. These channels are filled with a wide range of lithologies, including sand-rich deposits with biogenic gas which, in combination, produce strong lateral variations in the seismic velocity and absorption effects. Deeper in this section, there are contourite features trending parallel to the basin and deposited during long periods of very consistent current flow down the axis of the basin. The shallow channels, canyons, and contourites cause velocity anomalies that give rise to pull-up and push-down distortions in the seismic image, affecting its quality, which can be problematic for an accurate characterization of potential geologic targets. Also present in this region are salt diapirs, which indirectly can affect the shallow section due to the existence of gas chimneys and/or pockets created by hydrocarbons leaking from the crest of the diapir, or faults filled with gas migrating from the crest of the diapir to the top of the section.

Data acquisition and preprocessing

The data set was acquired with 10 solid streamers (Dowle, 2006) separated by 75 m, each 6 km long, comprising 480 × 12.5 m groups of receivers. The inline shot spacing is 18.75 m, and shots are fired in flip-flop sequence. The nominal acquisition bin-size is 6.25 × 18.75 m. The variable-depth streamer configuration introduces notch diversity by having the receiver ghosts in different regions of the spectrum over different offsets (Soubaras and Dowle, 2010). This type of acquisition is easily accommodated in our FWI scheme by honoring the depths of the receivers and including the free-surface boundary condition in the modeling engine for generating the ghost effect in the correct regions of the spectrum (Jupp et al., 2012).

The inversion is carried out with minimal preprocessing to ensure that the recorded information is as preserved as possible. Prior to running the inversion, swell noise is attenuated and the raw shot data were band-pass filtered from 4 to 8 Hz. In this frequency band, the variable-depth streamer data have an excellent signal-to-noise ratio, especially at the longer offsets. In addition, an inner and outer mute are applied to enhance the transmitted energy in the data, thus meeting an ideal condition for running FWI. The source wavelet was modeled from a standard industry package, but with no source or receiver ghosts present, and was subsequently low-pass filtered with a cutoff at 10 Hz. Consequently, the free-surface boundary condition in the modeling adds the necessary ghosts to the modeled data to match the seismic wavelet in the real data.

Multiparameter FWI results

This part of the Central North Sea is known to exhibit VTI; thus, a VTI parameterization of the subsurface is appropriate when inverting this data set. The inversion covers an area of approximately 25 × 43 km, which corresponds to 1075 km^2, down to a depth of 2 km. The model parameters are discretized on a 56.25 m grid. The grid spacing is chosen to meet the stability and numerical dispersion requirements to use our finite-difference scheme in space and time. The seismic data set comprises approximately 116,000 sources and approximately 186 million

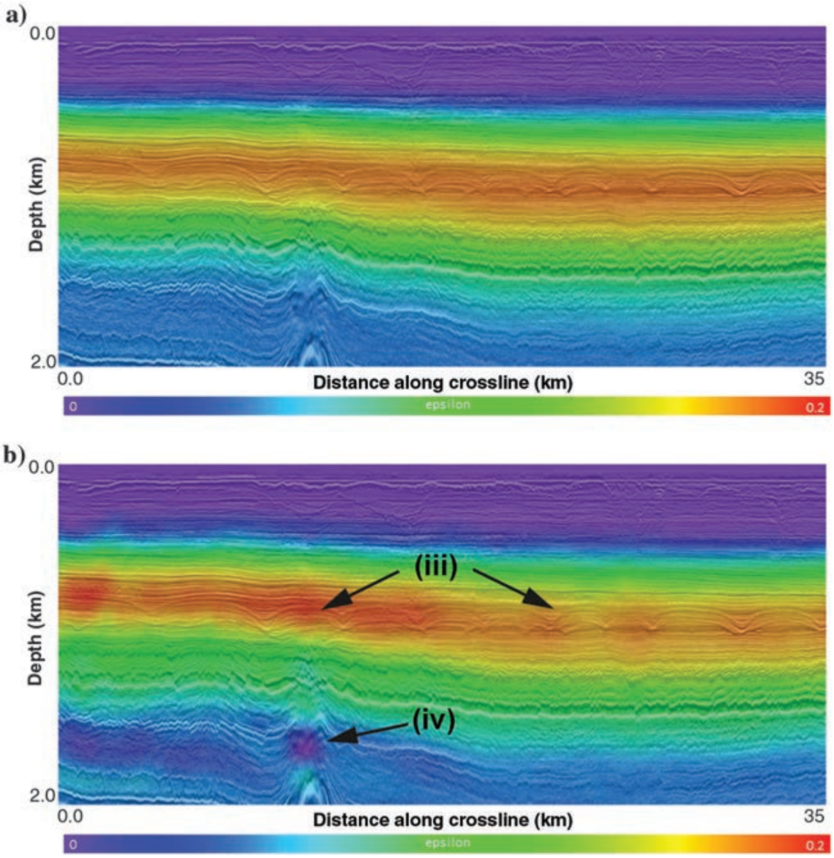

Figure 7. (a) Starting ε model and (b) FWI inverted ε model using the new PTS parameterization. Both models are overlaid with a stacked image generated from a Kirchhoff migration using the relevant model. We see very good agreement in the FWI models with geologic structures, (iii) building anomalies with higher ε in the center of the contourites and (iv) also capturing the gas pocket, showing a significant decrease here.

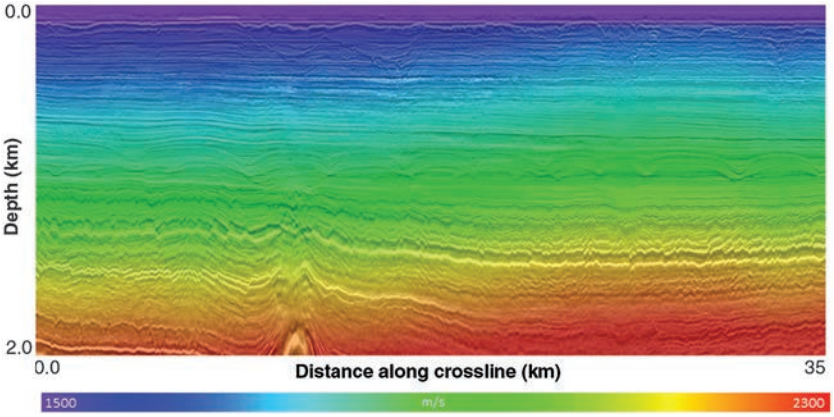

Figure 8. Starting δ model overlaid with a reference Kirchhoff stacked image.

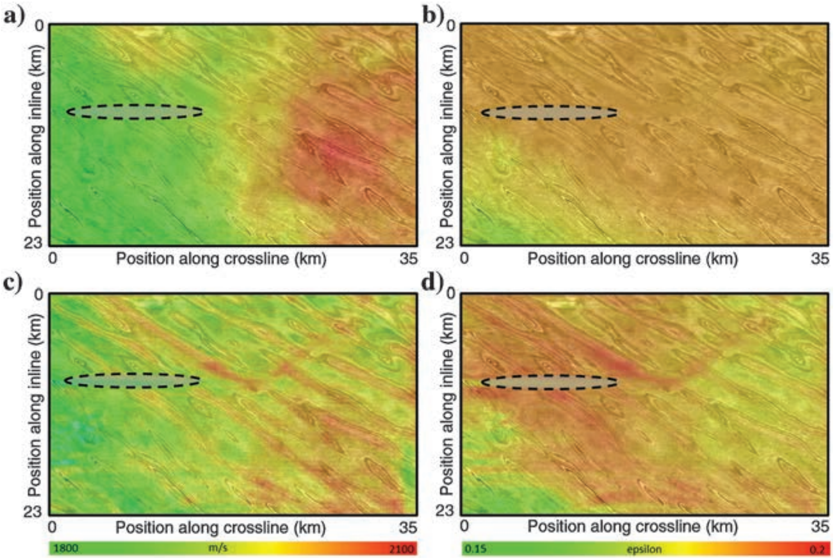

Figure 9. Depth slices at 920 m of: (a) starting velocity model, (b) starting ε model, (c) inverted velocity model, and (d) inverted ε model. Note that all displays are overlaid with a reference Kirchhoff stacked image, and the highlighted zone represents a void in the acquisition due to the existence of infrastructure in the area.

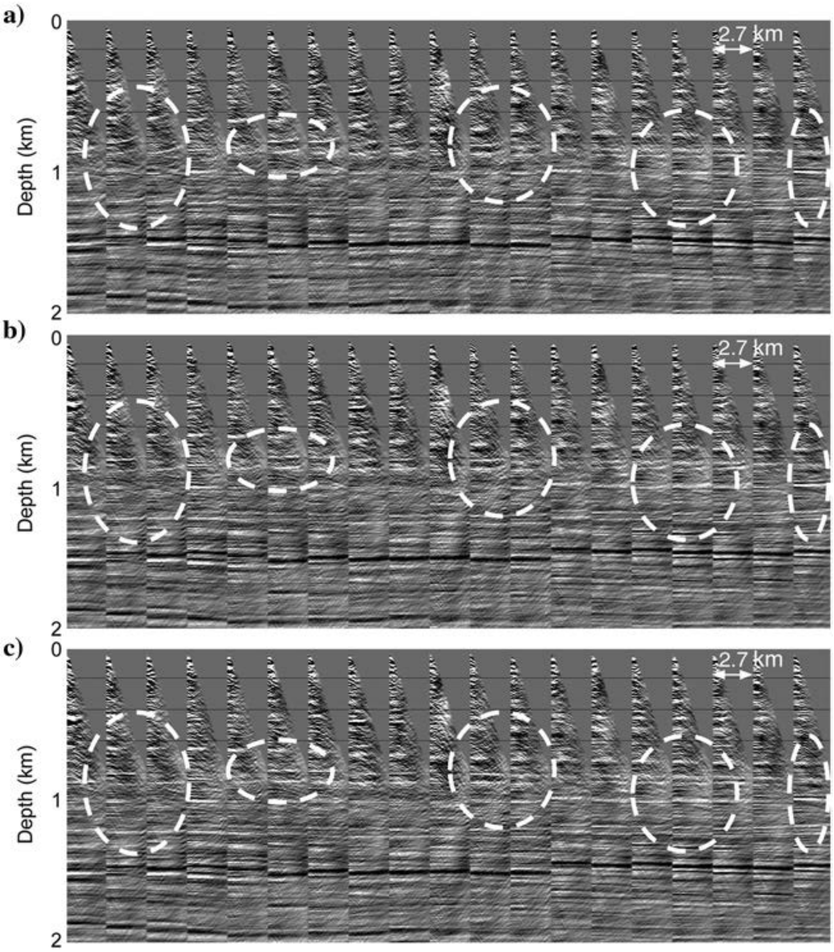

Figure 10. CIGs obtained using: (a) starting model, (b) velocity only inversion, and (c) inversion using the PTS parameters. The maximum offset is 2.7 km, and the maximum depth is 2 km.

traces in total. In the case of a streamer data set, the number of individual sources is very large, and the use of reciprocity has no benefit because the number of receivers is even larger than the number of sources. Consequently, a shot-skipping scheme is used (Warner et al., 2013), skipping 7 out of 8 shots in every iteration. Such an approach alleviates the computational burden for simulating all the data by a factor of 8 and also avoids the introduction of a bias in the inversion due to decimation, thus allowing for an excellent trade-off between computational cost and inversion quality.

The starting model is built using ray-based tomography and sonic information from wells existing in the area, allowing us to constrain vertical velocity and δ. This procedure builds a VTI initial model that is essentially a 1D model. Figures 6a, 7a, and 8 show these starting models for velocity, ε, and δ, respectively, overlaid with a Kirchhoff migrated section for quality control. The starting velocity model has a positive gradient with depth not showing any noticeable correlation with the detailed features in the seismic image, such as the shallow channels. In addition, it also does not show any large contrasts that can be correlated with the reflections in the seismic image.

Prior to running the inversion, the model response is computed for an arbitrary set of shots, and these are compared with the respective recorded shots within the lower bandwidth from which the inversion starts. This aims to control any potential cycle skipping between recorded and synthetic data, to guarantee convergence to a meaningful model. The inversion is carried out with a multiscale approach for three frequency bands with high cutoff filters applied at 5, 6, and 7 Hz, respectively.

Figures 6b and 7b show the inverted velocity and ε models, respectively, computed via the PTS parameters, with an overlay of the appropriately migrated Kirchhoff section (namely using the inverted models). The inverted velocity model in Figure 6b shows very good agreement with geologic structures, such as the shallow channels, the de-watering faults (i) and the low-velocity anomaly thought to be a small gas cloud (ii) above a salt diapir. The inverted ε model in Figure 7b also captures the gas pocket (iv), showing a significant decrease here. It is expected that the gas causes this structure to have physical properties closer to the isotropic case and, in addition, the velocity of the pressure waves also decreases. Hence, the velocity and ε decreased in this region, and this is in very good agreement with the geologic interpretation. The inversion also builds anomalies with higher ε in the center of the contourites (iii). This is in agreement with the expectation that these geologic features are filled with a more shale-rich sediment than the neighboring regions.

Figure 9a–9d shows depth slices at 920 m of the starting velocity, starting ε, inverted velocity and inverted ε models, respectively, overlaid with a reference Kirchhoff migration (note the highlighted void due to infrastructure in the area). FWI has captured, in a 3D sense, the structural details associated with the contourites in the velocity and ε models. The wavelengths of the anomalies in the ε model are generally larger than the wavelengths of the inverted anomalies in the velocity model. This is in agreement with the radiation pattern for the PTS parameters, as it is expected to recover ε with lower resolution than vertical velocity. Figure 6b also shows evidence of large wavelengths of velocity updates, demonstrating that even though the radiation pattern for η_{-1} shows less strength in the wider angles, this does not preclude the update of the longer wavelengths of this parameter. One can also observe regions where velocity and ε increase or decrease together, as well as regions where they increase or decrease in opposition to one another. Although this observation is good evidence of a meaningful update, we acknowledge that this is not definitive proof that crosscoupling between the parameters does not occur.

Figure 10a–10c depicts the common-image gathers (CIGs) obtained using the starting model, the velocity model inverted without updating the anisotropy (in this case, the inversion is carried out using the starting model for ε, in Figure 4a, and keeping it unchanged), and jointly updating velocity and anisotropy. The CIGs resulting from inverting for velocity only (Figure 10b) show an overall improvement in the flatness of the gathers when compared with the ones obtained from the starting model (Figure 10a). Nonetheless, they still show residual moveout at the longer offsets, which is highlighted by the white ellipses. This residual moveout is further improved as a result of updating the model for ε, jointly with the velocity model (Figure 10c).

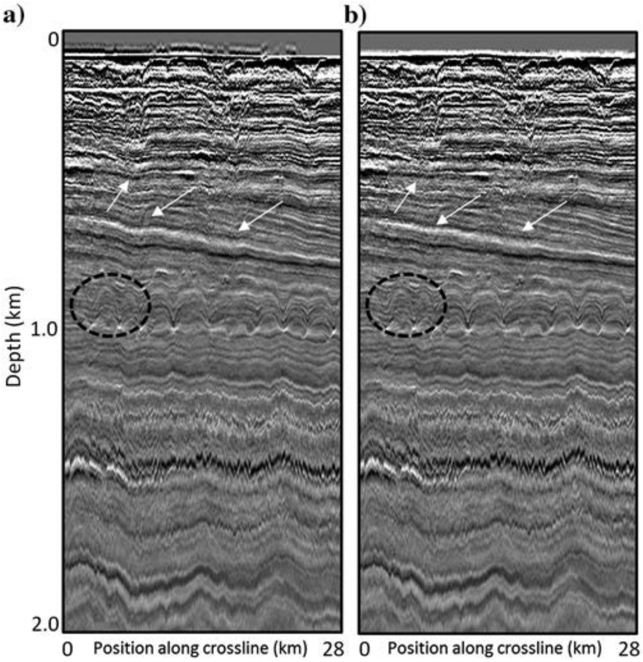

Figure 11. Kirchhoff stack image generated by migration with: (a) the initial model, and (b) the final model inverted using the joint FWI process. The white arrows show areas where the pull-up and push-down distortions in the image have been corrected, whereas the black circle highlights an area with an improved image response.

The overall improvement of the flatness of the gathers, when jointly inverting for velocity and ε, demonstrates that the update of anisotropy is meaningful even though the sensitivity to this parameter is very weak due to the band-limited nature of the source, as previously discussed. This demonstrates that the long wavelengths of ε in the starting model are accurate enough for carrying out the inversion and circumvent the lack of a zero-frequency, DC component, in the source wavelet (Podgornova et al., 2015).

Figure 11a and 11b shows sections that have been Kirchhoff migrated using the starting and jointly inverted models. Here, we see a clear uplift in the image quality (see white arrows) as the FWI model has fixed the pull-up and push-down distortions caused by the shallow channels, as well as improving the strength of the reflectors in places (in particular, see the region highlighted by the black circle).

CONCLUSIONS

A new parameterization (PTS) was introduced for anisotropic FWI with multiparameter update for VTI models. The PTS parameters are well-suited to invert anisotropy from diving waves. The new parameterization, as in the case of parameterizations with one velocity and two parameters for anisotropy, has maximum trade-offs at the largest scattering angles. For this reason, the estimated anisotropy has low resolution, and the starting model of anisotropy needs to be accurate enough to mitigate errors due to trade-off between the parameters, as well as the lack of a DC component in the source wavelet. In the case of the parameterization with two velocities and one parameter of anisotropy, the perturbation of the parameters leads to a perturbation in the traveltime. This can potentially lead to stronger changes in the updates of anisotropy in FWI. Thus, our new parameterization combines some of the characteristics of parameterizations using one velocity and two parameters of anisotropy, and parameterizations using two velocities and one parameter of anisotropy. The parameterization with PTS was demonstrated to work with a synthetic example and a real data example, showing its practical use.

The synthetic case demonstrates that the proposed parameterization is capable of reconstructing the models of velocity and ε reasonably well, given that the conditions necessary to perform FWI are met, and that these results are comparable when compared with (v_n, η, δ), which is a parameterization that is proven to lead to sensible inversions. The anomalies of velocity and ε are reconstructed with reasonably accurate magnitude, demonstrating that the scaling of the PTS parameters by its background improves the conditioning of this set of parameters.

The results of the real data inversion demonstrate that this parameterization is effective, as the final overall inversion resulted in the improvement of the seismic imaging and flatness of the CIGs. The results are in agreement with the radiation patterns, as the inverted velocity model shows updates with short and long wavelengths, and the inverted ε model shows predominantly updates with long wavelengths. Overall, a very good agreement was also obtained between the inverted models and the geologic structures observed in the seismic image.

ACKNOWLEDGMENTS

We thank CGG for permission to publish this work and CGG's Multi-Client & New Ventures business line for permission to show the Central North Sea data set. We would also like to thank our colleagues in CGG for their support, especially S. Thompson for his geologic interpretation, and G. Body and A. Fernandes for their help with the real data example. We very gratefully acknowledge the contributions of associate editor T. Alkhalifah and reviewers S. Operto, J.-W. Oh, and an anonymous reviewer, that substantially improved the quality of this paper. Finally, we would like to thank A. Stovas at NTNU in Trondheim, Norway for interesting discussions on this work.

APPENDIX A

ADJOINT-STATE METHOD FOR THE DISCRETIZED VTI WAVE EQUATION

Here, an adjoint method for computing the gradient of the objective function (equation 1) with respect to the model parameters is derived. There are essentially two approaches for obtaining the adjoint variable for an optimization problem: optimize then discretize (OTD) or discretize then optimize (DTO). Both approaches are equally valid. However, when using the OTD approach one needs to make sure that the solution obtained is consistent with a discrete form. Here, we outline the general DTO approach. However, some considerations are made in regards to the points that are specific of a wave equation for VTI anisotropy. We first introduce the Lagrangian functional:

$$\mathcal{L}(\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{m}) = J(\boldsymbol{u}(\boldsymbol{m}),\boldsymbol{d}) + \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{u} - \boldsymbol{s}), \tag{A-1}$$

where $\boldsymbol{A}$ is the forward modeling operator, $\boldsymbol{u}$ is the discretized state variable corresponding to the wavefield, $\boldsymbol{s}$ represents the source term, $\boldsymbol{\lambda}$ is the Lagrange multiplier corresponding to the adjoint variable, $\boldsymbol{d}$ are the recorded data, $\boldsymbol{m}$ are the model parameters, J is the L_2-norm of the square of the differences between the measured and simulated data, and the symbol T corresponds to the matrix transpose operation. Note that the Lagrangian equation A-1 is defined over the discretized quantities, the same approach can be used in a continuous form (wave operator, wavefield, and model parameters). The optimally conditions require $\nabla_{\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{m}}\mathcal{L}(\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{m}) = \boldsymbol{0}$, leading to

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{s}, \tag{A-2}$$

$$\boldsymbol{A}^T\boldsymbol{\lambda} = -\nabla_{\boldsymbol{u}}J(\boldsymbol{u},\boldsymbol{d}), \tag{A-3}$$

$$\nabla_{\boldsymbol{m}}J(\boldsymbol{u},\boldsymbol{\lambda},\boldsymbol{m}) = \boldsymbol{\lambda}^T[\nabla_{\boldsymbol{m}}\boldsymbol{A}]\boldsymbol{u}. \tag{A-4}$$

These relations result from determining the critical points of the Lagrangian. Equation A-2 is the state equation, equation A-3 is the adjoint equation, and equation A-4 is the gradient with respect to the model parameters for updating the model properties.

To apply this expression for minimizing the residual of the data misfit, one needs to write the wave operator in a compact form and determine its transpose, to obtain an expression for computing the adjoint wavefield. For this, we analyze the discrete form of the wave equation in equation 3. This wave equation is integrated in time with a fourth-order scheme as reported in equation 26 of Zhang et al. (2011).

Representing the state variable discretized in space at the nth time step by $\boldsymbol{u}^n$, the time stepping is represented in matrix form as

$$\begin{pmatrix} \boldsymbol{I} & & & & & & \\ \boldsymbol{D} & \boldsymbol{I} & & & & & \\ \boldsymbol{I} & \boldsymbol{D} & \boldsymbol{I} & & & & \\ & \boldsymbol{I} & \boldsymbol{D} & \boldsymbol{I} & & & \\ & & & & \ddots & & \\ & & & & \boldsymbol{I} & \boldsymbol{D} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^0 \\ \boldsymbol{u}^1 \\ \boldsymbol{u}^2 \\ \vdots \\ \vdots \\ \boldsymbol{u}^N \end{pmatrix} = \begin{pmatrix} \boldsymbol{s}^0 \\ \boldsymbol{s}^1 \\ \boldsymbol{s}^2 \\ \vdots \\ \vdots \\ \boldsymbol{s}^N \end{pmatrix}, \tag{A-5}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{D}$ is the discrete form of the main factor in the square brackets on the right side of equation 26 of Zhang et al. (2011), specific for our finite-difference representation, and $\boldsymbol{s}^n$ is the source term in the spatial grid at the nth time step. In the compact form, equation A-5 takes the form

$$\boldsymbol{A}\boldsymbol{u} = \boldsymbol{s}. \tag{A-6}$$

One can identify immediately that the structure of equation A-5 allows for a recursive solution, which is equivalent to the time-stepping procedure used for integrating the wave equation in time, with boundary conditions embedded in the structure of the matrix. The matrix representation of the time stepping is now substituted in the adjoint equation A-3, yielding

$$\boldsymbol{A}^T\boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{D}^T & \boldsymbol{I} & & & \\ & \boldsymbol{I} & \boldsymbol{D}^T & \boldsymbol{I} & & \\ & & \boldsymbol{I} & \boldsymbol{D}^T & \boldsymbol{I} & \\ & & & \ddots & & \\ & & & & \boldsymbol{I} & \boldsymbol{D}^T \\ & & & & & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda}^0 \\ \boldsymbol{\lambda}^1 \\ \boldsymbol{\lambda}^2 \\ \vdots \\ \vdots \\ \boldsymbol{\lambda}^N \end{pmatrix} = -\nabla_{\boldsymbol{u}}J(\boldsymbol{u},\boldsymbol{d}). \tag{A-7}$$

As in the case of the time stepping for the forward modeling, the solution of equation A-7 is also obtained recursively, starting from the adjoint wavefield at the final time step (which is generally referred to back propagating the data residuals). Also, the boundary conditions are assumed to be same for the adjoint field as they are embedded in $\boldsymbol{A}$ and the adjoint field is computed with final condition $\boldsymbol{\lambda}(t = t_{\max}) = \boldsymbol{0}$. It is important to note that the systems $\boldsymbol{A}$ and $\boldsymbol{A}^T$ are never explicitly formed throughout the inversion scheme, they simply express the recursive nature of time stepping the forward and adjoint simulations. In practice, the wavefield is computed for each shot and then correlated with the adjoint field at each time-step, accumulating and storing the result of this correlation, when computing the gradient of the misfit function.

APPENDIX B

RADIATION PATTERN FOR THE NEW PARAMETER SET

The background medium is isotropic; hence, r_{-1}, r_1, and r_3 are identically zero (as well as the Thomsen's parameters), reducing equation 13 to

$$\begin{cases} -\frac{\omega^2}{v_0^2\rho_0}p_0 - \frac{1}{\rho_0}\nabla^2 p_0 = s, \\ q_0 = 0. \end{cases} \tag{B-1}$$

Expanding the system of equations 13 in unperturbed and perturbed quantities (for the field and the parameters), and neglecting second-order terms, one can relate the primary field p_0 with the scattered field p_1 and q_1:

$$\begin{cases} -\frac{\omega^2}{v_0^2\rho}p_1 - \frac{1}{\rho}\nabla^2 p_1 = \frac{\omega^2}{v_0^2\rho}r_{-1}p_0 - \frac{1}{2\rho}\partial_z^2 r_{-1}p_0 - \frac{1}{2\rho}r_{-1}\partial_z^2 p_0 - \frac{v_0^2}{4\omega^2\rho}\nabla_H^2 r_{-1}\nabla_H^2 p_0 \\ -\frac{\omega^2}{v_0^2\rho}r_1 p_0 - \frac{1}{2\rho}\partial_z^2 r_1 p_0 - \frac{1}{2\rho}r_1\partial_z^2 p_0 + \frac{v_0^2}{2\omega^2\rho}\nabla_H^2 r_1 \nabla_H^2 p_0 \\ -\frac{v_0^2}{4\omega^2\rho}\nabla_H^2 r_3 \nabla_H^2 p_0, \end{cases} \tag{B-2}$$

where $\nabla_H^2 = (\partial_x^2, \partial_y^2)$ and the scattered field q_1 is eliminated through the relation

$$-\frac{\omega^2}{\nu_0^2\rho}q_1 = -\frac{1}{4}(r_{-1} - 2r_1 + r_3)\nabla_H^2 p_0. \tag{B-3}$$

Equation B-3 is obtained by applying the Born approximation to the second equation in equation 13. The pressure field p_0 in the background is determined from

$$p_0 = s(\omega)G(\mathbf{x}, \mathbf{x}_s, \omega), \tag{B-4}$$

where $G(\mathbf{x}, \mathbf{x}_s, \omega)$ is the Green's function for the Helmholtz equation in the background medium. The scattered field is now computed from the Born approximation

$$\begin{aligned} p_1(\boldsymbol{x}_r,\boldsymbol{x}_s,\omega) = {} & s(\omega)\int d\boldsymbol{x}\frac{\omega^2}{\rho v_0}r_{-1}G(\boldsymbol{x},\boldsymbol{x}_r,\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega) \\ & - s(\omega)\int d\boldsymbol{x}\frac{1}{2\rho}r_{-1}[G(\boldsymbol{x},\boldsymbol{x}_s,\omega)\partial_z^2 G(\boldsymbol{x},\boldsymbol{x}_r,\omega) + G(\boldsymbol{x},\boldsymbol{x}_r,\omega)\partial_z^2 G(\boldsymbol{x},\boldsymbol{x}_s,\omega)] \\ & - s(\omega)\int d\boldsymbol{x}\frac{v_0^2}{4\omega^2\rho}r_{-1}\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x}_r,\omega)\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x}_s,\omega) \\ & - s(\omega)\int d\boldsymbol{x}\frac{\omega^2}{v_0^2\rho}r_1 G(\boldsymbol{x},\boldsymbol{x}_r,\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega) \\ & - s(\omega)\int d\boldsymbol{x}\frac{1}{2\rho}r_1[G(\boldsymbol{x},\boldsymbol{x}_s,\omega)\partial_z^2 G(\boldsymbol{x},\boldsymbol{x}_r,\omega) + G(\boldsymbol{x},\boldsymbol{x}_r,\omega)\partial_z^2 G(\boldsymbol{x},\boldsymbol{x}_s,\omega)] \\ & + s(\omega)\int d\boldsymbol{x}\frac{v_0^2}{2\omega^2\rho}r_1\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x}_r,\omega)\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x}_s,\omega) \\ & - s(\omega)\int d\boldsymbol{x}\frac{v_0^2}{4\omega^2\rho}r_3\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x}_r,\omega)\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x}_s,\omega), \end{aligned} \tag{B-5}$$

where integration by parts has been used and carried out over the support of the integrands. For the computation of the derivatives of the Green's function, the asymptotic Green's functions for source and receivers is used (Sirgue, 2003):

$$\begin{aligned} G(\mathbf{x}, \mathbf{x}_s, \omega) &\propto \exp(ik_0\mathbf{p}_s \cdot \mathbf{x}), \\ G(\mathbf{x}, \mathbf{x}_r, \omega) &\propto \exp(ik_0\mathbf{p}_r \cdot \mathbf{x}), \end{aligned} \tag{B-6}$$

where the vectors $\mathbf{p}_s$ and $\mathbf{p}_r$ for a horizontal reflector are given by

$$\begin{aligned} \boldsymbol{p}_s &= (\sin(\theta_s/2), \cos(\theta_s/2)) \\ \boldsymbol{p}_r &= (-\sin(\theta_s/2), \cos(\theta_s/2)). \end{aligned} \tag{B-7}$$

Substituting the derivatives of the asymptotic Green's functions into equation B-5, and applying integration by parts, leads to the solution of the scattered field p_1 (equations 15 and 16).

REFERENCES

Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631, doi: 10.1190/1.1444361.

Alkhalifah, T., and Y. Choi, 2014, From tomography to FWI with a single objective function: Geophysics, **79**, no. 2, R55–R61, doi: 10.1190/geo2013-0291.1.

Alkhalifah, T., and R.-E. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, no. 3, R91–R101, doi: 10.1190/geo2013-0366.1.

Alkhalifah, T., S. Fomel, and B. Biondi, 2001, The space-time domain: Theory and modeling for anisotropic media: Geophysical Journal International, **144**, 105–113, doi: 10.1046/j.1365-246x.2001.00300.x.

Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566, doi: 10.1190/1.1443888.

Backus, G. E., 1962, Long-wave elastic anisotropy produced by horizontal layering: Journal of Geophysical Research, **67**, 4427–4440, doi: 10.1029/JZ067i011p04427.

Aki, K., and P.G. Richards, 2002, Quantitative seismology (2nd ed.): University Science Books.

Baek, H., H. Calandra, and L. Demanet, 2014, Velocity estimation via registration-guided least-squares inversion: Geophysics, **79**, no. 2, R79–R89, doi: 10.1190/geo2013-0146.1.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473, doi: 10.1190/1.1443880.

Chavent, G., 2010, Nonlinear least squares for inverse problems: Springer.

Cheng, X., K. Jiao, D. Sun, and D. Vigh, 2014, Multiparameter full-waveform inversion for acoustic VTI medium with surface seismic data: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, We E106 04.

Choi, Y., and T. Alkhalifah, 2013, Frequency-domain waveform inversion using the phase derivative: Geophysical Journal International, **195**, 1904–1916, doi: 10.1093/gji/ggt351.

Choi, Y., and T. Alkhalifah, 2015, Unwrapped phase inversion with an exponential damping: Geophysics, **80**, no. 5, R251–R264, doi: 10.1190/geo2014-0498.1.

Dowle, R., 2006, Solid streamer noise reduction principles: 76th Annual International Meeting, SEG, Expanded Abstracts, 85–89.

Duveneck, E., P. Milcik, P. M. Bakker, and C. Perkins, 2008, Acoustic VTI wave equations and their application for anisotropic reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2186–2189.

Etgen, J., 1986, High-order finite-difference reverse time migration with the 2-way non-reflecting wave equation: SEP Report 48, 133–146.

Fichtner, A., H.-P. Bunge, and H. Igel, 2006, The adjoint method in seismology. Part I: Theory: Physics of the Earth and Planetary Interiors, **157**, 86–104.

Gholami, Y., R. Brossier, S. Operto, A. Ribodetti, and J. Virieux, 2013a, Which parametrization is suitable for acoustic VTI full-waveform inversion? Part 1: Sensitivity and trade-off analysis: Geophysics, **78**, no. 2, R81–R105, doi: 10.1190/geo2012-0204.1.

Gholami, Y., R. Brossier, S. Operto, V. Prieux, A. Ribodetti, and J. Virieux, 2013b, Which parametrization is suitable for acoustic VTI full-waveform inversion? Part 2: Application to Valhall: Geophysics, **78**, no. 2, R107–R124, doi: 10.1190/geo2012-0203.1.

Grechka, V., 2009, Nonuniqueness of traveltime inversion in elliptically anisotropic media: 79th Annual International Meeting, SEG, Expanded Abstracts, 196–200.

Jupp, R., A. Ratcliffe, and R. Wombell, 2012, Application of full-waveform inversion to variable-depth streamer data: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0613.1.

Kelley, C.T., 1999, Iterative methods for optimization (Frontiers in applied mathematics): SIAM.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, no. 6, WCC177–WCC188, doi: 10.1190/1.3230502.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Proceedings of the International Conference on Inverse Scattering, Theory, and Applications, SIAM.

Lions, J. L., 1971, Optimal control of systems governed by partial differential equations: Springer-Verlag.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792.

Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: Geophysics, **78**, no., 6, R223–R233, doi: 10.1190/geo2013-0004.1.

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: Geophysical Prospecting, **62**, 1353–1375, doi: 10.1111/1365-2478.12136.

Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Métivier, and J. Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, 1040–1054, doi: 10.1190/tle32091040.1.

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/j.1365-246X.2006.02978.x.

Plessix, R.-E., and Q. Cao, 2011, A paramaterization study for seismic full waveform inversion in an acoustic vertical transversely isotropic medium: Geophysical Journal International, **185**, 539–556, doi: 10.1111/j.1365-246X.2011.04957.x.

Plessix, R.-E., G. Baeten, G. J. W. de Maag, F. ten Kroode, and Z Rujie, 2012, Full waveform inversion and distance separated simultaneous sweeping: A study with a land seismic data set: Geophysical Prospecting, **60**, 733–747, doi: 10.1111/j.1365-2478.2011.01036.x.

Podgornova, O., S. Leaney, and L. Liang, 2015, Analysis of resolution limits of VTI anisotropy with full waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1188–1192.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: Imaging compressional wave speed, density and attenuation: Geophysical Journal International, **194**, 1640–1664, doi: 10.1093/gji/ggt177.

Soubaras, R., and R. Dowle, 2010, Variable-depth streamer: A broadband marine solution: First Break, **28**, 89–96.

Sirgue, L., 2003, Inversion de la form d'onde dans le domaine frequential de donnes sismiques grands offsets: Ph.D. thesis, l'Ecole Normale Superior de Paris.

Sirgue, L., O. I. Barkved, J. P. van Gestel, O. J. Askim, and J. H. Kommendal, 2009, 3D waveform inversion in Valhall wide-azimuth OBC: 71st Annual International Conference and Exhibition, EAGE, Extended Abstracts, U038.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, doi: 10.1190/1.1649391.

Stopin, A., R.-E. Plessix, and S. Al Abri, 2014, Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman: Geophysics, **79**, no. 3, WA69–WA77, doi: 10.1190/geo2013-0323.1.

Stovas, A., 2008, Kinematically equivalent velocity distributions: Geophysics, **73**, no. 5, VE369–VE375, doi: 10.1190/1.2952683.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.

Tang, Y., S. Lee, A. Baumstein, and D. Hinkley, 2013, Tomographically enhanced full wavefield inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 1037–1041.

Thomsen, L. A., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966, doi: 10.1190/1.1442051.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667, doi: 10.1093/gji/ggt258.

Vinje, V., A. Stovas, and D. Reynaud, 2013, Preserved-traveltime smoothing: Geophysical Prospecting, **61**, 380–390, doi: 10.1111/j.1365-2478.2012.01124.x.

Wang, C., D. Yingst, J. Bai, J. Leveille, P. Farmer, and J. Brittan, 2013, Waveform inversion including well constraints, anisotropy, and attenuation: The Leading Edge, **32**, 1056–1062, doi: 10.1190/tle32091056.1.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion — FWI without cycle skipping — Theory: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, We E106 13.

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**, no. 2, R59–R80, doi: 10.1190/geo2012-0338.1.

Wu, R. S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25, doi: 10.1190/1.1442237.

Zhang, Y., H. Zhang, and G. Zhang, 2011, A stable TTI reverse time migration and its implementation: Geophysics, **76**, no. 3, WA3–WA11, doi: 10.1190/1.3554411.

Zhou, H., G. Zhang, and R. Bloor, 2006, An anisotropic wave equation for VTI media: 68th Annual International Conference and Exhibition, EAGE, Extended Abstracts, H033.