GEOPHYSICS®

Semi-global visco-acoustic full waveform inversion

Journal:	*Geophysics*
Manuscript ID	GEO-2017-0773.R1
Manuscript Type:	Technical Paper
Keywords:	acoustic, anisotropy, attenuation, full-waveform inversion, inversion
Area of Expertise:	Seismic Inversion

SCHOLARONE™ Manuscripts

Semi-global visco-acoustic full waveform inversion

Nuno V. da Silva[1], Gang Yao[2], Michael Warner[1]

Running head: semi-global visco-acoustic inversion

[1]Imperial College London, Royal School of Mines building, Department of Earth Science and Engineering, Prince Consort Road, SW7 2BP, London, UK. E-mail: n.vieira-da-silva@imperial.ac.uk, nuno.vdasilva@gmail.com; m.warner@imperial.ac.uk.

[2]Department of Earth Science, Rice University, 6100 Main St., Houston, TX 77005, US. Email: g.yao@rice.edu.

ABSTRACT

Full waveform inversion deals with estimating physical properties of the Earth's subsurface by matching simulated to recorded seismic data. Intrinsic attenuation in the medium, leads to dispersion of propagating waves and absorption of energy – media with this type of rheology is not perfectly elastic. Accounting for that effect is necessary in order to simulate wave propagation in realistic geological media, leading to the need to estimate intrinsic attenuation from the seismic data. That increases the complexity of the constitutive laws leading to additional issues related to the ill-posed nature of the inverse problem. In particular, the joint estimation of several physical properties increases the null-space of the parameter-space, leading to a larger domain of ambiguity and increasing the number of different models that can equally well explain the data. Here, we introduce a method for the joint inversion of velocity and intrinsic attenuation using semi-global inversion; this combines quantum particle-swarm optimization for the estimation of the intrinsic attenuation with nested gradient-descent iterations for the estimation of the P-wave velocity. This approach takes advantage of the fact that some physical properties, and in particular the intrinsic attenuation, can be represented using a reduced basis, decreasing substantially the dimension of the search space. We demonstrate the feasibility of the method and its robustness to ambiguity with 2D synthetic examples. The 3D inversion of a field dataset for a geological medium with transversely isotropic anisotropy in velocity shows the feasibility of the method for inverting

large-scale real seismic data and improving the data fitting. The principal benefits of the semi-global multi-parameter inversion are the recovery of the intrinsic attenuation from the data and the recovery of the true undispersed infinite-frequency P-wave velocity, while mitigating ambiguity between the estimated parameters.

INTRODUCTION

Seismic full-waveform inversion (FWI) (Tarantola, 1984) is a method for estimating physical properties of the subsurface from seismic recorded data. Even though the velocity of seismic waves has been the chief estimated property in applications of FWI (Sirgue et al., 2010; da Silva et al., 2016; Routh et al., 2017; Yao and Du, 2017; Yao et al., 2018), the method can also be used for the estimation of other physical properties provided that the required property affects the seismic data in a defined way, at least in principle (Tarantola, 1988). The most common type of constitutive law used in commercial 3D FWI considers the Earth to be an acoustic medium. More recently it has become commonplace also to account for seismic anisotropy (Plessix and Cao, 2011; Warner et al., 2013; da Silva et al., 2016). Seismic anisotropy is particularly important in order to account for the different scales of heterogeneity in the medium, when compared to the signal bandwidth, as well as preferred alignment of crystals, cracks, and layers (Thomsen, 1986). Seismic anisotropy is normally indispensable for fitting seismic field data at both long and short offsets (Plessix and Cao, 2011). In addition, internal friction, crystal-defect sliding, grain-boundary processes, thermo-elastic effects, and/or fluid-filled cracks are responsible for anelastic attenuation of propagating seismic waves; elastic scattering from sub-wavelength heterogeneities can also mimic closely the effects of true anelastic loss.

Anelastic loss and sub-wavelength scattering attenuation are generally described by a parameter *quality factor*, Q. It quantifies the quantity of energy that is lost per cycle (Aki and Richards, 2002). This effect can be accounted for by correcting the amplitude and phase of the recorded signals (Xue et al., 2016, Agudo et al., 2018), prior to carrying out the inversion of data. Alternatively, intrinsic attenuation can be accounted for in the constitutive laws. Because attenuating media is frequency dependent, the dependency on the frequency can be explicitly introduced in the constitutive law when using frequency-domain simulation of seismic waves (Song et al. 1995; Liao and McMechan, 1996; 1995; Hicks and Pratt, 2001; Ribodetti and Hanyga, 2004). The frequency-dependent nature of the intrinsic attenuation translates into a convolution operation in the time domain between the anelasticity and strain tensors describing hysteresis in the medium (Aki and Richards, 2002). That is, at any given instant, the state of deformation, and consequently energy exchange in the system, depends upon all the previous states. This relation between stress and strain needs to be addressed efficiently, when carrying out numerical simulations in the time domain, in order to avoid computational burdens. Examples of methods to deal efficiently with numerical simulation in attenuating media, and it the time domain, are the use of Padé approximants (Day and Minster, 1984), the method of memory variables with the Standard Linear Solid (SLS) (Carcione et al., 1988a, 1988b; Robertsson et al., 1994), the pseudo-spectral method (Liao and McMechan, 1993), and the temporal fractional derivative with Kjartansson's Q-theory (Kjartansson, 1979; Zhu and Carcione, 2014; Zhu and Harris, 2015; Yao et al., 2017).

Song and Pratt (1995) and Hicks and Pratt (2001) reported simultaneous inversion for velocity and intrinsic attenuation in the frequency-domain FWI. Watanabe et al. (2004) introduced joint cascading inversion for the real part of the velocity using phase information, and inverting the imaginary part of the velocity using amplitude information. Rao and Wang (2015) introduced inversion of velocity and intrinsic attenuation by sequential updates of velocity and intrinsic attenuation. Bai et al. (2014) discussed time-domain visco-acoustic FWI. A particularity of using frequency-domain FWI is the fact that the parameterization for the inversion can be formulated explicitly in terms of Q for each frequency, or in terms of real and imaginary parts of the velocity or slowness (Pratt et al. 2004, Kamei and Pratt, 2013, Rao and Wang, 2015).

Song et al. (1995), Malinowski et al. (2011), Operto et al. (2015) and Plessix et al. (2016) presented examples of the application of FWI to data sets affected by intrinsic attenuation. Those previous works focus on the minimization of the objective function by updating the model parameters iteratively with the gradient of the objective function with respect to velocity and intrinsic attenuation. First-order local methods are generally preferred in order to implement feasible and efficient inversion algorithms. One of the key drawbacks of that type of method, for multi-parameter inversion, is that the estimated properties are affected by cross-talk as a result of ambiguity in the relation between perturbations in the model parameters and perturbations in the data. In other words, those jointly estimated properties are affected by a larger null-space. That has been reported when jointly inverting for velocity and intrinsic attenuation (Hak and Mulder, 2010; Malinowski et al., 2011; Kamei and Pratt, 2013; Plessix et al., 2016) with local inversion methods.

The application of global inversion methods is generally hindered by the need to perform a large number of model-space searches (Sen and Stoffa, 2013), as the size of this search space increases rapidly with the dimensionality of the model-space. Nonetheless, some physical properties can be represented by smooth models, or models that have very weak spatial variation. Models with that type of property can be represented in a space of basis functions with low dimension. This concept has been used for the estimation of smooth starting velocity models for FWI (Datta and Sen, 2016; Diouane et al., 2016), and for seismic anisotropy (Afanasiev, 2014; Debens et al., 2015a). Ji and Singh (2005) combined a genetic algorithm with the conjugate-gradient method for the estimation of the elasticity tensor coefficients in 1D.

The quality factor is in a class of physical properties that can be approximated by a reduced basis. That is, it can be sufficient to use a smooth model of Q for predicting seismic waves that have the correct kinematics and dynamics. This statement is validated later using a numerical example. Thus, global inversion of intrinsic attenuation becomes computationally feasible as the dimensionality of the search space for this parameter can be reduced.

We introduce and discuss a method for the joint estimation of seismic intrinsic attenuation and seismic velocity. This method combines conventional FWI and quantum particle-swarm optimisation (QPSO) (Sun et al., 2004a-b, Debens et al., 2015). The semi-global inversion algorithm nests local gradient-descent iterations within an outer QPSO global iteration. The velocity model is represented on a dense grid, and it is updated only at the nested gradient-descent iterations. The model of the intrinsic attenuation is represented on a sparse basis, and it is estimated only at the outer global iterations. We note that such an approach is not limited to parameterizations with velocity and intrinsic attenuation only; the wave equation can describe wave propagation in more complex rheology, and in our examples, the medium includes anisotropy of velocity.

The paper is structured as follows. First, we review the theory for the simulation of seismic waves in anisotropic visco-acoustic media. We then introduce the semi-global algorithm and we demonstrate its effectiveness with synthetic examples, discussing the accuracy of a sparse representation of intrinsic attenuation, as well as the mitigation of cross-talk between estimated parameters. Finally, we show an application of the global inversion scheme to a real data set acquired in the North Sea, demonstrating that the proposed approach is both accurate and computationally feasible for the inversion of a large-scale field seismic data, leading to improved velocity recovery and a more-complete characterization of the sub-surface.

FORWARD MODELING

Here, we consider a visco-acoustic medium with vertical transverse isotropy (VTI). The time-dependent stiffness tensor $\boldsymbol{C}(t)$ for such rheology is defined as

$$\boldsymbol{C}(t) = \begin{pmatrix} C_{11} & C_{11} & C_{13} \\ C_{11} & C_{11} & C_{13} \\ C_{13} & C_{13} & C_{33} \end{pmatrix} = \rho v_R^2 \begin{pmatrix} (1+2\varepsilon_R)f_h(t) & (1+2\varepsilon_R)f_h(t) & \sqrt{1+2\delta_R}f_n(t) \\ (1+2\varepsilon_R)f_h(t) & (1+2\varepsilon_R)f_h(t) & \sqrt{1+2\delta_R}f_n(t) \\ \sqrt{1+2\delta_R}f_n(t) & \sqrt{1+2\delta_R}f_n(t) & f_0(t) \end{pmatrix}, \quad (1)$$

where t denotes time, $K_R = \rho v_R^2$ is the relaxed bulk-modulus with density ρ and relaxed vertical velocity v_R, while ε_R and δ_R are the Thomsen's parameters (Thomsen, 1986) associated to the relaxed velocity, the relaxation mechanism is defined as (Bland, 1960).

$$f_\gamma(t) = \left[1 - \frac{1}{L}\sum_{l=1}^{L}\left(1 - \frac{\tau_\gamma^{\varepsilon l}}{\tau^{\sigma l}}\right)e^{-t/\tau^{\sigma l}}\right]H(t); \qquad \gamma = h,n,0, \quad (2)$$

where the indexes *h*, *n,* and *0* denote the horizontal, normal, and vertical components, respectively. The number of SLS's is denoted by *L*, *H*(t) is the Heaviside function, $\tau_\gamma^{\varepsilon l}$ is the strain relaxation time for each component γ, and $\tau^{\sigma l}$is the stress relaxation time. In the scope of this paper, we consider only the case when all components have the same relaxation mechanism. Hence, the strain relaxation times are the same for all components and equal to $\tau^{\varepsilon l}$. This means that in practice anisotropy exists only in velocity and that the model of intrinsic attenuation is isotropic. We choose this type of constitutive law in order to reduce the dimension of the space of inversion. The relation between the elastic and anelastic response of the medium is obtained taking $\boldsymbol{C}(t \rightarrow 0)$, yielding

$$K_U(1+2\varepsilon_U) = K_R(1+2\varepsilon_R)(1+\tau), \quad (3a)$$

$$K_U\sqrt{1+2\delta_U} = K_R\sqrt{1+2\delta_R}(1+\tau), \quad (3b)$$

$$K_U = K_R(1+\tau), \quad (3c)$$

where the subscript *U* denotes *unrelaxed* and corresponds to the elastic response of a medium (Aki and Richards, 2002). The parameter

$$\tau = \frac{\tau^{\varepsilon l}}{\tau^{\sigma l}} - 1, \tag{4}$$

is determined in a least-squares sense by fitting the response of the relaxation mechanism to a desired constant Q over the frequency range of interest. Blanch et al. (1995) and Hestholm et al. (2006) discussed comprehensively how to compute an optimal value of τ.

The Hooke's law relates the time-dependent stress tensor, $\boldsymbol{\sigma}(t)$, and the time–dependent strain-tensor, $\boldsymbol{\varepsilon}(t)$, given by

$$\boldsymbol{\sigma}(t) = \boldsymbol{C}(t) * \frac{\partial \boldsymbol{\varepsilon}}{\partial t}(t). \tag{5}$$

Cauchy's law of motion describes the dynamics of deformations,

$$\rho \frac{\partial \boldsymbol{v}}{\partial t} = \nabla \boldsymbol{\sigma} + \boldsymbol{F_v}, \tag{6}$$

where $\boldsymbol{v}$ is the particle velocity and $\boldsymbol{F_v}$ are the body forces. Combining equations 1 to 6, the convolution operator can be substituted by the introduction of memory variables (Carcione, 1988; Robertsson et al., 1994), leading to a second-order partial differential-equation system for modeling pressure waves in anisotropic media with viscosity (da Silva et al., 2018)

$$\begin{cases} \dfrac{\partial^2 p_h}{\partial t^2} = K_U\left[(1+2\varepsilon_U)\nabla_h \cdot \left(\frac{1}{\rho}\nabla_h p_h\right) + \sqrt{1+2\delta_U}\nabla_v \cdot \left(\frac{1}{\rho}\nabla_v p_n\right)\right] + \frac{1}{L}\sum_{l=1}^{L}\frac{\partial r_l}{\partial t} + s(t) \\ \dfrac{\partial^2 p_n}{\partial t^2} = K_U\left[\sqrt{1+2\delta_U}\nabla_h \cdot \left(\frac{1}{\rho}\nabla_h p_h\right) + \nabla_v \cdot \left(\frac{1}{\rho}\nabla_v p_n\right)\right] + \frac{1}{L}\sum_{l=1}^{L}\frac{\partial w_l}{\partial t} + s(t) \\ \dfrac{\partial^2 r_l}{\partial t^2} = -\frac{1}{\tau_{\sigma l}}\frac{\partial r_l}{\partial t} - K_R\frac{\tau}{\tau_{\sigma l}}\left[(1+2\varepsilon_R)\nabla_h \cdot \left(\frac{1}{\rho}\nabla_h p_h\right) + \sqrt{1+2\delta_R}\nabla_v \cdot \left(\frac{1}{\rho}\nabla_v p_n\right)\right] \\ \dfrac{\partial^2 w_l}{\partial t^2} = -\frac{1}{\tau_{\sigma l}}\frac{\partial w_l}{\partial t} - K_R\frac{\tau}{\tau_{\sigma l}}\left[\sqrt{1+2\delta_R}\nabla_h \cdot \left(\frac{1}{\rho}\nabla_h p_h\right) + \nabla_v \cdot \left(\frac{1}{\rho}\nabla_v p_n\right)\right] \end{cases}, \tag{7}$$

where $p_h = \sigma_{xx} = \sigma_{yy}$ and $p_n = \sigma_{zz}$ are the horizontal and vertical pseudo-pressure, respectively, r_l and w_l are the memory variables for the horizontal and vertical pseudo-pressure, respectively, $\nabla_h = (\partial_x, \partial_y)$ is the horizontal gradient, $\nabla_v = \partial_z$ is the vertical gradient,

K_R is the relaxed bulk modulus, and $s(t)$ is the source term. The system of equations 7 is discretized with central finite differences with second-order accuracy in time and eight-order accuracy in space. Its derivation as well as its numerical solution is discussed in more detail in da Silva et al. (2018b).

INVERSION ALGORITHM

The inversion is formulated as the minimization of the squares of the residuals constrained by a regularization term (Tarantola, 1984; Aster et al., 2012)

$$J_F(\boldsymbol{p},\boldsymbol{m}) = J(\boldsymbol{p},\boldsymbol{m}) + J_R(\boldsymbol{m}_2) = \frac{1}{2}\sum_r \|\boldsymbol{p}_r - \boldsymbol{d}_r\|_2^2 + \frac{1}{2}\lambda^2 \|\boldsymbol{L}\boldsymbol{m}_2\|_2^2, \tag{8}$$

where $\boldsymbol{d}_r$ is the time-record at the receiver r, $\boldsymbol{d} = (\boldsymbol{d}_1,\boldsymbol{d}_2,\dots\boldsymbol{d}_r,\dots)$ is the whole collection of recorded data, $\boldsymbol{p} = (\boldsymbol{p}_1,\boldsymbol{p}_2,\dots\boldsymbol{p}_r,\dots)$ is the collection of simulated data, $\boldsymbol{L}$ is a smoothing regularization operator obtained with the discretization of the Laplacian operator (Aster et al., 2012), λ is the trade-off (or regularization) parameter. The model parameters, $\boldsymbol{m} = (\boldsymbol{m}_1,\boldsymbol{m}_2)$, are the velocity, denoted with $\boldsymbol{m}_1$, and the logarithm of base 10 of the quality factor, denoted with $\boldsymbol{m}_2$. The quality factor can range over several orders of magnitude. Its parameterization with the logarithm ameliorates this issue, as it localizes the search of an optimal value. In addition, the support of the logarithm function is the set of positive real numbers. Hence, this parameterization of Q has the advantage of constraining naturally the successive updates of Q to be always positive. With the exception of an example with smoothing regularization for Q, we always take the objective function $J_F(\boldsymbol{p},\boldsymbol{m}) = J(\boldsymbol{p},\boldsymbol{m})$.

Tarantola and Vallete (1982) set the ambitious goal that the solution of the inverse problem requires a full probabilistic description of the solution space. This, however, is generally not attainable in full large-scale 3D geophysical inversion, as the computational load associated with the simulation of the data can become prohibitive if full-space searches are carried out. Consequently, the solution of large-scale geophysical inverse problems is generally estimated with local optimization methods. This class of methods relies upon *a priori* information, as for instance obtained by determining a good starting model using seismic tomography. This approach aims to place successive local iterations within the global basin of attraction. Nonetheless, convergence towards the global minimum is not guaranteed.

Local inversion methods – gradient descent

Local inversion methods rely upon the linearization of the objective function in the vicinity of the optimal solution. The search of the minimum takes place in a very small portion of the model space. A typical local inversion approach consists in carrying out successive model updates minimizing the objective function along the direction of steepest descent

$$\boldsymbol{m}_{k+1} = \boldsymbol{m}_k - \alpha_k \mathbf{H}_k^{-1} \nabla_m J_F(\boldsymbol{p},\boldsymbol{m}), \tag{9}$$

where k denotes the iteration number. Naturally, many other methods can be considered, as for example the conjugate gradient method or the L-BFGS method, to name a few (Nocedal and Wright, 2006). Herein, we always consider the steepest descent method when carrying out local iterations. The gradient of the objective function $\nabla_m J_F(\boldsymbol{p},\boldsymbol{m})$ is computed efficiently by the adjoint method (Fernández-Berdaguer, 1998, Fichtner et al., 2006). Appendix A, outlines the computation of the gradient of the objective function 8 using the adjoint-state method with a discretize-then-optimize approach. The operator $\boldsymbol{H}_k$ is a pre-conditioner of the gradient of the objective function. It approximates the action of the inverse of the Hessian over the gradient, and it is obtained taking only diagonal elements and neglecting second-order terms. See Pratt et al. (1998) for a thorough discussion on approximating the Hessian in the context of FWI. Note that for inversion solutions without regularization $J_F(\boldsymbol{p},\boldsymbol{m}) = J(\boldsymbol{p},\boldsymbol{m})$, then $\nabla_m J_F(\boldsymbol{p},\boldsymbol{m}) = \nabla_m J(\boldsymbol{p},\boldsymbol{m})$.

Semi-global inversion

On the choice of the global optimization method

This work is focused on introducing an approach for the practical estimation of intrinsic attenuation and velocity from real data, rather than discussing the computational efficiency of optimization algorithms. We point out that similar semi-global approaches could be obtained, in principle, combining different algorithms. For example, conjugate gradients or the L-BFGS method (Nocedal and Wright, 2006) could be used for the nested iterations. However, the semi-global inversion algorithm introduced herein is focused on the estimation of parameters over one or two frequency bands, and for a relatively small number of iterations. In such a case the full benefit of these methods might not be achieved, while introducing computational overhead as a consequence of increasing the number of arithmetic operations. In addition, alternative global optimization methods can also be considered. Monte Carlo methods are relatively inefficient, as they require a large number of realizations in the search space. Simulated Annealing (SA) mitigates this problem. However, it depends upon a cooling schedule. For a successful application of SA, the cooling schedule needs to be sufficiently slow in order to avoid convergence towards local minima (equivalent to the process of annealing producing a structure with defects). If that cooling schedule is too slow then the number of space searches also becomes relatively high. In addition, SA can also require model-space searches that effectively do not produce an update. Other efficient classes of metaheuristic methods include Genetic Algorithms (GA) and Particle Swarm Optimization (PSO). The former encompasses a vast domain of different approaches based mainly upon natural selection. The latter is based upon the dynamics of populations or physical systems (Kennedy and Eberhart, 1995; Kennedy and Eberhart, 2001). Each element of that population, or swarm, is called a particle. Particles move in the multi-dimensional space evolving towards an optimum position (e.g. minimum energy). The position and velocity define the states of a particle. The evolution of each particle is determined by its own current and past states, as well as, the past states of the whole swarm. The position of a particle is effectively the optimization variable. Despite the need to store a *memory* of the system, PSO algorithms are computationally very efficient, and are not memory demanding. Quantum Particle Swarm Optimization (QPSO) was introduced extending the concept of swarm optimization to systems with quantum behavior (Sun et al. 2004a-b). A key advantage of QPSO over PSO is its superior computational efficiency, as it does not need information on the velocity vector of each particle. Hence fewer parameters need to be adjusted. In quantum systems, the dynamics of particles is defined by a quantum potential. That potential sets an attractor towards a global minimum. The δ-well (Levin, 2001) is the most commonly used quantum potential in QPSO. That is also the potential used in our implementation. Note that the state of a quantum particle is defined by a potential, and the optimization variable is the position. This sets a clear difference between PSO and QPSO as in the former the state depends explicitly on the position. In QPSO, the information on the position of each particle is obtained collapsing the potential with the Monte Carlo method (see Sun et al. 2004a-b for a thorough explanation).

Debens (2015) reported higher efficiency of QPSO over GA. In addition, as one can immediately conclude, the dimension of the population (number of particles in the swarm) drives the computational cost of QPSO. However, as the state of each particle can be assessed independently, this makes QPSO very suitable for parallelization over the number of particles. That is an important feature as computing model realizations can be very intensive. Parallelization is not as trivial with SA or Monte Carlo algorithms. Hence, QPSO has significant advantages over other global optimization algorithms regarding its efficiency, and for that reason it is our method of choice. We refer to Press et al. (2003) and Sen and Stoffa (2013) for a thorough description on global optimization methods.

The semi-global inversion algorithm

Algorithm 1 describes a QPSO algorithm combined with local nested iterations of steepest descent. It is the combination of those two methods that gives the semi-global algorithm introduced herein. Effectively that algorithm can be used in any inversion application changing the function to carry out the model realizations. The position of a particle k in the swarm, at the global iteration n, is denoted $\boldsymbol{\eta}_{k,n}$. Note that effectively $\boldsymbol{\eta}_{k,n}$, represents the model of attenuation, $\boldsymbol{m}_2$, associated to a particle k at the global iteration n. One can then define the set of all the positions of the particles in the swarm at iteration n as, $\mathcal{M}_n =$

$\{\boldsymbol{\eta}_{1,n},\boldsymbol{\eta}_{2,n},\ldots,\boldsymbol{\eta}_{N_p,n}\}$. The object $\mathcal{M}_0$ defines the set of the initial position of the particles in the swarm. The elements of $\mathcal{M}_0$ must be uniformly distributed. Each position $\boldsymbol{\eta}_{k,n}$ has a respective level of energy $J_{k,n}$. The objective of the QPSO optimization is minimizing the overall energy of the system measured by an objective function. The objective function can be arbitrary. However, in the context of geophysical inversion and herein, the objective function is the L_2-norm of the data misfit, $J_{k,n}(\boldsymbol{m}_1,\boldsymbol{\eta}_{k,n}) = \frac{1}{2}\|\boldsymbol{p}(\boldsymbol{m}_1,\boldsymbol{\eta}_{k,n}) - \boldsymbol{d}\|_2^2$, where $\boldsymbol{d}$ denotes the data and $\boldsymbol{p}((\boldsymbol{m}_1,\boldsymbol{\eta}_{k,n})$ is a model realization. The set $\mathcal{J}_n = \{J_{1,n},J_{2,n},\ldots,J_{N_p,n}\}$ contains the corresponding level of misfit for each element of $\mathcal{M}_n$. The set of the best overall position for each particle is denoted $\mathcal{M}^* = \{\boldsymbol{\eta}_1^*,\boldsymbol{\eta}_2^*,\ldots,\boldsymbol{\eta}_{N_p}^*\}$. Each element of $\mathcal{M}^*$ is the position for each particle at which the lowest value of the objective function was reached, between the first and a given iteration. The corresponding set of data misfits is denoted $\mathcal{J}^* = \{J_1^*,J_2^*,\ldots,J_M^*\}$. Finally, the best overall position of the swarm $\boldsymbol{\eta}_g^*$ corresponds to the minimum of $\mathcal{M}^*$, denoted as J_g^*. Formally, $J_g^* = J_{a,b}$ with $a,b = argmin_{a \in \{1,\ldots,N_p\}, b \in \{1,\ldots,n\}}(\mathcal{J}_n)$, and $\boldsymbol{\eta}_g^* = \boldsymbol{\eta}_{a,b}^*$. The update of each particle's position depends on the diagonal matrices $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$, with random numbers uniformly distributed on their diagonals, and on $r(0,1)$, which denotes a random number following a normal distribution with zero mean and unit variance.

```
n = 0
set starting model of P-wave: v_n
set initial positions: M_n = {η_{1,n},…,η_{N_p,n}}
while min_J {J} > ε and n < N_g and f < N_f:
    for each particle k ∈ {1,…,N_p}:
        J_{k,n}, ṽ_{k,n} = modelRealizations( )
    set J_n = {J_{1,n},…,J_{N_p,n}}
    update velocity v_{n+1} = (1 / Σ_k J_{k,n}) Σ_k J_{k,n} ṽ_{k,n}
```

if $n > 0$:
    for each particle $k \in \{1,\ldots,N_p\}$:
        if $J_{k,n} > J_{k,n-1}$:
            $J_{k,n} > J_{k,n-1}$
            $\boldsymbol{\eta}_{k,n} = \boldsymbol{\eta}_{k,n-1}$
    set $\mathcal{M}^* = \{\boldsymbol{\eta}_1^*,\ldots,\boldsymbol{\eta}_{N_p}^*\}$
    select best position in the swarm $\boldsymbol{\eta}_g^*$
    for each particle $k \in \{1,\ldots,N_p\}$:
        generate $\boldsymbol{\varphi}$
        determine a local attractor $\boldsymbol{q}_k = \boldsymbol{\varphi}\boldsymbol{\eta}_k + (\boldsymbol{1} - \boldsymbol{\varphi})\boldsymbol{\eta}_g^*$
        set $\boldsymbol{l}_k = \beta|\boldsymbol{\eta}_k^* - \boldsymbol{\eta}_{k,n}|$
        generate $\boldsymbol{\phi}$
        generate random $r(0,1)$

        if $r(0,1) < 0.5$ then:
            $\boldsymbol{\eta}_{k,n+1} = \boldsymbol{q}_k - ln(\boldsymbol{\phi}^{-\boldsymbol{1}})\boldsymbol{l}_k$
        else:
            $\boldsymbol{\eta}_{k,n+1} = \boldsymbol{q}_k + ln(\boldsymbol{\phi}^{-\boldsymbol{1}})\boldsymbol{l}_k$
    $n \leftarrow n + 1$
function modelRealizations( ):
    for $l \in \{0,\ldots N_{Ln} - 1\}$:
        $\boldsymbol{m}_l = \{\boldsymbol{v}_n, \boldsymbol{\eta}_{k,n}\}$
        $J(\boldsymbol{m}_l) = \frac{1}{2}\|\boldsymbol{p}(\boldsymbol{m}_l) - \boldsymbol{d}\|_2^2$
        compute $\nabla_v J_{k,n}(\boldsymbol{m}_l)$
        $\boldsymbol{v}_{l+1} = \boldsymbol{v}_l - \alpha_l \boldsymbol{H}_l^{-1} \nabla_{\boldsymbol{v}} J_{k,n}(\boldsymbol{m}_l)$

```
    return J, v_Ln
```

Algorithm 1. The semi-global inversion algorithm combining QPSO with nested local iterations of steepest descent.

The parameter β is the contraction-expansion coefficient. Sun (2012) reports that β=0.75 is a suitable value for unimodal problems (only one global minimum and possibly several local minima). We did not test the effect of changing this parameter as it has been investigated previously, and the value pointed out yielded good convergence in all the examples in the outline.

We point out that the semi-global algorithm is described in a general fashion as it effectively finds applications with several classes of parameters (e.g. seismic anisotropy, or elastic parameters). In this scope, the position of a particle is effectively a model of Q represented over a sparse basis and parameterized with the logarithm of base 10.

Basis Functions

The higher the dimension of the search space, the higher the number of particles that is required, and the higher the number of searches that needs to be carried out. As a consequence, for a good performance of the semi-global inversion approach, reducing the dimension of the search space is crucial in order to make this method feasible in large-scale realistic applications. Formally, a distribution of a parameter $\varrho(\boldsymbol{x})$ in space can be represented over a discrete basis as

$$\varrho(\boldsymbol{x}) = \sum_{j=1}^{N} \varrho_k \chi_k(\boldsymbol{x}). \quad (10)$$

Then, the distribution of $\varrho(\boldsymbol{x})$ can be fully determined in space from the set of coefficients $\boldsymbol{\varrho} = \{\varrho_1,\ldots,\varrho_N\}$, and from a set of chosen basis function $\{\chi_k(\boldsymbol{x})\}$. The elements of $\boldsymbol{\varrho}$ are effectively coefficients of the basis functions. That means that the dimension of the set of basis functions equals that of the set of coefficients. A simple example is that of a 3D Cartesian grid discretization with equidistant grid nodes along the oriented axis. In that case a generic basis function can be defined as $\chi_k = \delta(\boldsymbol{x}_k) = \delta(x - x_k, y - y_k, z - z_k)$, where $\boldsymbol{x}_k$ is the position of the *k*-th node in the 3D space and δ is the Dirac's delta function.

The dimension of the basis functions is related to the scale of the wavelengths that are present in the distribution of a given model parameter in space. That is, the smaller the wavelength of the anomalies, the higher the dimension of the space of the basis functions has to be. Choosing the space of basis functions appropriately allows reducing the dimension of the discretization space. For example, wavelets are remarkably flexible to represent both large and small-scale anomalies with a small number of basis functions (Daubechies, 1992). That fact has been exploited for devising model-compression schemes in second-order Gauss-Newton inversion (Abubakar et al., 2012).

For large-scale global (or semi-global) inversion methods, and when the estimate of the states of the system is computationally intensive (as in the case of using finite-difference or finite-element methods), the use of global (or semi-global) inversion is feasible for up to a few tens of parameters. Thus, that requires two factors to be considered - either the parameter to be estimated has a smooth variation in space, or not having, its smooth representation yields a model-response very close to that of a full representation of all its wavelength components.

Here, we take advantage of the fact that for some classes of physical properties it is sufficient to know their long-wavelength components in order to obtain relatively accurate simulation of seismic data, or at least accurate enough to carry out inversion. Examples of physical properties that can often be represented with a low-dimensional basis are background P-wave velocity, seismic anisotropy parameters, V_p-to-V_s ratio (da Silva et al., 2018a), and P-wave intrinsic attenuation. Here, we estimate models of intrinsic attenuation in a low-dimensional space, henceforth referred to as *reduced-space* or sparse, grid, whilst the acoustic velocity is estimated in a conventional grid, henceforth referred to as the *full-space* grid. This approach requires a mapping between reduced and full-space grids. Here, we perform such mappings with B-splines (Schumaker, 2015). The mapping approach consists in estimating the values of an unknown parameter over the nodes of a sparser grid. Then, that parameter is determined in the full space grid prior to carrying out the steepest-descent iterations. Note that, we do not regularize the estimates of the model of the intrinsic attenuation in the case of semi-global inversion.

AMBIGUITY BETWEEN THE ESTIMATES OF v_p AND Q

In this example we investigate numerically the different behavior of the gradient-descent and the semi-global inversion method when dealing with ambiguity. We consider a velocity model with a positive anomaly superimposed on a background defined by a positive gradient of velocity, as depicted in Figure 1a. The Q model is depicted in Figure 2a, and it has a homogenous background with an anomaly superimposed with relatively low Q. The background of the Q model does not absorb energy. The region of the low-Q anomaly absorbs energy.

The anomalies of the P-wave velocity and Q are not correlated in space. Their spatial dimension is chosen such that both of them can be represented over a sparse grid. This allows testing the robustness of the semi-global algorithm to ambiguity between the estimates of P-wave velocity and Q.

The semi-global inversion estimates the coefficients of the basis functions for B-spline interpolation at the nodes marked with circles in Figures 2a, 2c and 2d. The semi-global inversion has three degrees of freedom. Two of the unknowns are the values of the coefficients at the red and black nodes. The third degree of freedom is assigned to all the white nodes. The black node overlaid to the velocity model in Figure 1a has the same position as that in Figure 2. If the semi-global inversion algorithm is robust to cross-talk, then the estimated coefficient of the basis function, located at the position of the black node, is close to the value of Q of the background.

We generated data for 345 shots, spaced of 24 m, and at a depth of 6 m. The temporal dependency of the source wavelet is given by a Ricker wavelet with a peak at 8 Hz. The receiver geometry is fixed with 720 receivers, spaced of 12 m, and at 12 m of depth.

Estimating parameters from data contaminated with noise is a key issue in inverse problems as a result of poorer conditioning. Hence, the existence of noise in the data is an important factor to consider when testing the robustness of an inversion algorithm. We then generated a second synthetic data set contaminating the synthetically generated data with noise. The noise is generated with a random variable following a Gaussian distribution with zero mean and unit variance. The noise is convolved with the source wavelet in order to match its spectrum with that of the data. We set a loss of 5% in the signal-to-noise ratio, when adding noise to the data.

We carried out a series of tests inverting the data with FWI and with the semi-global inversion combined with FWI. The local FWI inverts the data over four frequency bands with cut-off filters applied at 2.5, 3, 3.5, and 4 Hz. The updates are iterated twelve times in the first frequency band and six times in all the subsequent bands of frequency. The semi-global inversion is carried out at 2.5 Hz only, with six outer global iterations and two nested local gradient-descent iterations for velocity only. Conventional FWI is then carried out with the estimated P-wave and Q models at 2.5, 3, 3.5, and 4 Hz, iterating 6 times in each frequency band. The overall amount of P-wave velocity iterations is the same in both cases. FWI inverts all the data, whereas each nested local iteration of the semi-global inversion fits one sixth of the data. That means that after completing the semi-global inversion all the data has been used twice, but at a fraction of the cost. The starting model, for each parameter, for all the local inversions is the background model. The semi-global inversion generates starting models for Q randomly, setting the range between 1.2 and 5 on a logarithmic scale of basis 10.

The jointly inverted velocity and Q models, with FWI, are depicted in Figures 1b and 2b, respectively. The data used in this example is noise-free. One can observe that the algorithm did not recover the high-velocity anomaly. Even though the starting model of Q is the exact background, the final estimated model has very large errors. It diverged towards a completely different background model with low values of Q. In addition, it shows a hint of correlation with the high-velocity anomaly, as it presents an artifact matching closely the top of this velocity anomaly.

Figures 1c and 1d depict the velocity models estimated with the semi-global inversion, with noise-free and noisy data, respectively. The respective estimated *Q* models are depicted in Figures 2c and 2d. One can observe that the high-velocity anomaly is very well recovered when the data is noise-free. The existence of noise in the data affects the estimate of the velocity anomaly. However, it is clear that the algorithm updates correctly the velocity anomaly, increasing its magnitude at the correct position in space. It is also important to note that the nested local iterations of semi-global inversion utilize a reduced dataset per iteration, hence decreasing redundancy and degrading the signal-to-noise ratio. Therefore, it is expected a stronger effect of the noise over the estimates at the nested iterations. The corresponding *Q* models are estimated with a very good accuracy regarding its background, positioning of the anomaly, and their respective magnitudes. The small-scale high-*Q* anomalies in the vicinity of the true low-*Q* anomaly are artifacts generated by the interpolator. These have overall no impact, as effectively these values of *Q* are very high and they have a smooth variation yielding effectively the same response as the background. One can observe that the presence of the high-velocity anomaly did not influence the estimated *Q,* resulting from semi-global inversion. In other words, there is no evidence of ambiguity between the P-wave velocity and *Q*. Those velocity and *Q* models are then used to carry out conventional FWI from 2.5 to 4 Hz, updating for velocity only. The resulting velocity models are depicted in Figures 1e, for noise-free data, and 1f for noisy data. Those pictures show a further improvement of the inverted velocity anomaly. The corresponding inverted velocity models, estimated with FWI only, assuming that the *Q* model is known, are depicted in Figures 1g (noise-free data) and 1h (noisy data). These results show that the semi-global inversion algorithm can effectively determine a model of *Q* that is comparable with the true model of *Q*, leading to improved inversion results. The semi-global inversion clearly outperformed the joint local inversion for velocity and *Q* with conventional FWI regarding accuracy and suppression of ambiguity, due to trade-off between estimates.

A further analysis on ambiguity

The existence of ambiguity between estimates of different classes of parameters is a well-known issue. Radiation patterns, based on energy scattering, give a very good insight on its analysis. In Appendix B, we derive the radiation pattern for a visco-acoustic medium with velocity anisotropy. Figure 3, depicts the resulting radiation pattern as a function of aperture angle between source and receiver. One can immediately observe that the radiation envelopes for velocity and Q overlap the whole range of aperture angles. This has a dramatic consequence as it means that it is very difficult to determine which perturbed physical property has originated a first-order perturbation in the data. Then the inversion algorithm cannot determine unequivocally the source of the anomaly in the medium. As the radiation energy envelope of P-wave velocity and Q overlaps over the whole range of aperture angles, then trade-off between these parameters occurs for all the wavenumbers of the perturbations of these parameters, accordingly to (Wu and Toksöz, 1987)

$$\boldsymbol{k}_m = 2k_0 \cos\frac{\theta}{2}\boldsymbol{n}, \tag{11}$$

where $\boldsymbol{k}_m$ is the wavenumber of the perturbation, $\boldsymbol{n}$ is the normal to the reflector, k_0 is the wavenumber of the background, and θ is the aperture angle between source and receiver.

One can make the observation that the semi-global inversion algorithm decouples the interfering radiation patterns because the nested local iterations perturb for P-wave velocity only, and Q is updated at the outer iteration without perturbing any of the parameters. That is not the same as alternating successively the updates of P-wave velocity and Q. In fact, in the semi-global inversion approach outlined, the outer iteration finds a model of Q that yields the best data fitting for a given model of velocity. The same is not guaranteed when updating both parameters in an alternating fashion using local search methods. That is because the local successive updates continue to be strongly dependent on the initial guess, as well as, on the current estimate, hence introducing a bias over the successive iterations. See Watanabe et al. (2004) for a comprehensive study on updating the parameters in a cascading approach. The outer global iteration ameliorates that issue significantly as the search space is not constrained by a local search in the vicinity of the current estimate. In other words, the outer local iteration operates over a much larger search space. That justifies why the semi-global inversion converged towards models that are significantly more accurate. We can take this analysis further by actually looking at the relation between the perturbation in the model parameters and the perturbation in the objective function. Visco-acoustic media are dispersive hence we carry out this analysis taking an objective function in the frequency domain

$$J(\boldsymbol{m},\omega) = \frac{1}{2}\sum_s \int d\boldsymbol{x}\delta(\boldsymbol{x}-\boldsymbol{x}_{r,s})[u(\boldsymbol{m},\boldsymbol{x},\omega)-d(\boldsymbol{x},\omega)][u(\boldsymbol{m},\boldsymbol{x},\omega)-d(\boldsymbol{x},\omega)]^*, \qquad (12)$$

where, the sum is carried out over the number of sources, $\boldsymbol{x}_{r,s}$ is the vector of the receiver coordinates for a source s, $d(\boldsymbol{x},\omega)$ are the observations, $u(\boldsymbol{m},\boldsymbol{x},\omega)$ are the synthetic data, $\boldsymbol{m} = (v_U,Q)$ are the model parameters. The Dirac delta function $\delta(\boldsymbol{x}-\boldsymbol{x}_{r,s})$ under the sign of integration discretizes the functions denoting the observations and the synthetic data.

The relation between the perturbations in the model parameters and the perturbations in the objective is formally given by

$$\delta J(\boldsymbol{m},\omega) = \frac{\partial J}{\partial \boldsymbol{m}} \cdot \delta\boldsymbol{m}, \qquad (13)$$

and the gradient of the objective function expressed in equation 12 is explicitly given by

$$\frac{\partial J}{\partial \boldsymbol{m}} = \Re\mathrm{e}\sum_s \int d\boldsymbol{x}\frac{\partial u}{\partial \boldsymbol{m}}\delta d(\boldsymbol{x}_{r,s},\omega), \qquad (14)$$

where $\delta d(\boldsymbol{x}_{r,s},\omega) = [u(\boldsymbol{m},\boldsymbol{x},\omega)-d(\boldsymbol{x},\omega)]^*\delta(\boldsymbol{x}-\boldsymbol{x}_{r,s})$. Substituting equation 14 into equation 13 yields

$$\delta J(\boldsymbol{m},\omega) = \Re\mathrm{e}\sum_s \int d\boldsymbol{x}\delta d\frac{\partial u}{\partial \boldsymbol{m}} \cdot \delta\boldsymbol{m}, \qquad (15)$$

relating perturbations in the data, due to changes in the model, with perturbations in the objective function. As the parameters for which we are inverting are encapsulated into the anelasic moduli, $C(\omega,v_U,Q)$, then we use the chain rule yielding

$$\delta J(\boldsymbol{m},\omega) = \Re\mathrm{e}\sum_s \int d\boldsymbol{x}\delta d\frac{\partial u}{\partial C}\frac{\partial C}{\partial \boldsymbol{m}} \cdot \delta\boldsymbol{m} = \Re\mathrm{e}\sum_s \int d\boldsymbol{x}\delta d\frac{\partial u}{\partial C}\left(\frac{\partial C}{\partial v_U}\delta v_U + \frac{\partial C}{\partial Q}\delta Q\right), \qquad (16a)$$

and,

$$\delta C(\boldsymbol{m},\omega) = \frac{\partial C}{\partial v_U}\delta v_U + \frac{\partial C}{\partial Q}\delta Q. \qquad (16b)$$

Expressions 16a and 16b give the explicit relation between perturbations in the model parameters and perturbations in the objective function. One can observe that there is no control over the possible different combinations of δv_U and δQ that yield the same net δC. In that case δJ remains unchanged. In the trivial case $\delta C = 0$, the condition $\nabla_{\boldsymbol{m}} C \cdot \delta \boldsymbol{m} = 0$ defines level sets in the space of the model parameters for which the physical realization does not change the objective function. In other words, different model parameters can yield the same model response. Hence, the existence of ambiguity when estimating simultaneously different classes of parameters with gradient-type methods is inevitable. The semi-global inversion method updates only the velocity model. Then, expression 16a becomes

$$\delta J(\boldsymbol{m},\omega) = \Re\mathrm{e} \sum_s \int d\boldsymbol{x} \delta d \frac{\partial u}{\partial C} \frac{\partial C}{\partial \boldsymbol{m}} \cdot \delta \boldsymbol{m} = \Re\mathrm{e} \sum_s \int d\boldsymbol{x} \delta d \frac{\partial \boldsymbol{u}}{\partial C} \frac{\partial C}{\partial v_U} \delta v_U. \qquad (17)$$

One can see that, in this case, the perturbations in the objective function are exclusively related with perturbations in P-wave velocity. Hence, there is no ambiguity introduced from one parameter update into another. In this case, perturbations in Q cannot influence perturbations in P-wave velocity as they do not contribute to the minimization of the objective function while updating velocity. Then the gradient of the objective function is free of trade-off between the perturbations of the different parameters. It is important to note that this is not the same as stating that the estimates of one parameter do not influence the other. The estimates of Q at the outer global iteration are influenced by the quality of the estimate of P-wave velocity. Conversely, the estimate of P-wave velocity at the nested iterations is influenced by the estimates of Q. In addition, inverse problems are inherently ill posed, i.e. there is no guarantee of unique solution. The key advantage of the semi-global inversion is that ambiguities are not being introduced by jointly estimating the model perturbations, hence deflating a significant region of the null-space.

As a last remark, in the case of parameterizing Q with the logarithm of base 10, equation 16b becomes, $\delta C(\boldsymbol{m},\omega) = \frac{\partial C}{\partial v_U} \delta v_U + \frac{\partial C}{\partial Q} \frac{\partial Q}{\partial (\log_{10} Q)} \delta(\log_{10} Q)$. Hence, the same conclusions regarding the trade-off between estimates hold.

SYNTHETIC EXAMPLES

Comparison of shot records obtained with true and background *Q*

We show with a numerical example, that in the case of Q, it is sufficient at least in some practical settings, to represent the model with the correct long wavelength, thus demonstrating the relevance and feasibility of the method outlined in this paper. We first generate an acoustic shot record for different combinations of velocity and intrinsic attenuation. The velocity model is the Marmousi model (Figure 4a), and synthetic shot gathers are generated without Q, and with the Q models depicted in Figures 4b and 4d. The intrinsic attenuation model depicted in Figure 4d is the long-wavelength component of the model of Q shown in Figure 4b. The model of velocity anisotropy is kept fixed in all the synthetic examples. The Thomsen's parameters, ε and δ, defining the model of anisotropy are depicted in Figure 5a-b, respectively. The receiver geometry is fixed with a 12 m horizontal receiver separation, 720 receivers in total, at 12 m depth. The source is placed at 6 m depth and at 1440 m from the leftmost boundary of the model. The temporal history of the source function is given by a Ricker wavelet with a peak frequency at 8 Hz. A free-surface boundary condition is imposed at the top of the model, and absorbing boundaries (Berenger, 1994; Yao et al., 2018) are used at the lateral and bottom boundaries.

Figure 6a, depicts a shot gather generated without intrinsic attenuation. Figures 6a and 6b depict shot gathers generated with the intrinsic attenuation models shown in Figures 4b and 4d, respectively. Comparing the highlighted events, the effect of intrinsic attenuation in the synthetic data can be seen clearly by observing the stronger amplitude of the events in the shot record that is not affected by intrinsic attenuation (Figure 6a). Figure 6d depicts the difference between the shot gathers in Figures 6b and 6c. One can see that the difference between those shot gathers is residual. In order to compare these data more closely we compare two traces from each one of the shot gathers. We selected trace number 300, with a relatively short offset (Figure 7a) and trace number 700 with a larger offset (Figure 7c). One can see that the traces generated with attenuating models are different from the trace generated without intrinsic attenuation in the model. In addition, one cannot detect any noticeable difference between the traces generated with the different models of intrinsic attenuation. As the intrinsic attenuation affects the frequency response of the seismic data, it is important to carry out a comparison between the spectra of these traces. Figure 7b compares the spectrum of the traces in Figure 7a, and Figure 7d compares the spectrum of the traces in Figure 7c. One can see that the spectrum of the data generated with intrinsic attenuation is different from the spectrum of the data generated without intrinsic attenuation, in both cases. In addition, one can also make the observation that the spectrum remained unchanged despite ignoring the high-wavenumber component of the model of intrinsic attenuation, at both selected source-receiver offsets. We emphasize that the background model of Q is correct in both cases. This is a clear demonstration of a potential ambiguity in the model space, as two different models have a very close response. In the presence of noise this issue becomes more critical. One can then conclude that data generated, and possibly recorded, with similar settings (geometry and bandwidth) to those in this example, may not contain enough information in order to reconstruct the high-wavenumbers of the model of intrinsic attenuation. This example also shows that the high-wavenumbers in the true Q model can be largely ignored when estimating P-wave velocity, provided that the background model of Q is sufficiently accurate.

Inversion with the true and background *Q* models

To test the hypothesis formulated in the previous sentence, we carried out the inversion for velocity only using FWI generating a full data set using the Marmousi model (Figure 4a) and the model of Q with long- and short- wavelength perturbations (Figure 4b). The synthetic data is generated for 345 shots spaced 24 m along the horizontal and placed at 6 m depth. Both the receiver geometry and temporal dependency of the source wavelet are the same as in the previous example.

The inversion is carried out using the starting velocity model depicted in Figure 4c, and the model of intrinsic attenuation is left unchanged throughout the inversion. Thus only the velocity model is updated. The inversion is set to run in blocks of six iterations per frequency band. After completing each block of iterations, the frequency band is widened 1 Hz. The first frequency band is selected applying a cut-off filter at 3 Hz. Then, the process is repeated up to a band of frequencies limited at 16 Hz. This means that we run 84 iterations in total. Inversions are carried out leaving *Q* unchanged throughout the inversion, and using only its long-wavelength component as depicted in Figure 4d. The corresponding inverted velocity models are depicted in Figures 8a and 8b, respectively. By inspection of the two figures, one can conclude that the differences between the two inverted models are negligible. That result demonstrates that P-wave velocity can be estimated accurately inverting surface seismic data if the background *Q* model is known. The latter can be represented in a sparse basis. Then a sparse representation of *Q* is relevant, with the proviso that the long wavelengths of the perturbations of *Q* are appropriately represented in that reduced basis.

Conventional FWI versus semi-global inversion

Inversion with fixed Q

In this example, the medium is visco-acoustic and it has velocity anisotropy. The data are generated using the Marmousi model (Figure 4a) and the model of *Q* in Figure 4b. We use those models in all the following synthetic examples. The model of seismic anisotropy is defined by the Thomsen's parameters depicted in Figures 5a-b. The source-receiver configuration is the same as that used in the previous section.

Conventional FWI is carried out over the generated synthetic dataset using the same setting as that in the example in the previous section. First, it is assumed that the model of *Q* is known, leading to the inverted velocity model in Figure 9a. As expected, the inversion recovers most parts of the structures both quantitatively and qualitatively. When the intrinsic attenuation is ignored throughout the inversion, the inverted velocity model shows strong errors as shown in

Figure 9b. The effect of energy absorption in the upper left region of the model due to the presence of an anomaly with small values of Q is clear. The effect of the error in the model of intrinsic attenuation (as it is disregarded) becomes especially visible bellow 1.2 km of depth, where the high velocity sharp structures become completely smeared (highlighted with white arrows). In addition, the average estimated velocity is lower than that of the true velocity model. That is most visible in the central and deeper part of the model. The error in the macro-velocity model impedes the correct positioning and reconstruction of the high wavenumber anomalies of P-wave velocity. Those inaccuracies happen because in the presence of intrinsic attenuation, propagating seismic waves suffer from dispersion, meaning that the lower-frequency components of the signal propagate at a slower rate than that of the higher-frequency components (Aki and Richards, 2002). This implies that the overall envelope of energy, or group energy, propagates more slowly. Because the generated synthetic data are affected by intrinsic attenuation, when inverting this dataset without accounting for intrinsic attenuation will drive the estimated velocity to be less than the true one in order to match correctly the travel-time and recorded phases at the receiver positions.

Joint local inversion without regularization

A third example involves jointly inverting for velocity and intrinsic attenuation. The starting model for Q is homogeneous and with $Q = 10^5$. This inversion example does not include the regularization term. Hence, the inversion is carried out minimizing only the data-misfit term. In this case, the inverted velocity model, depicted in Figure 9c, shows short wavelength artifacts and a very poor reconstruction especially in the central region of the model. In addition, some of the anomalies are smeared, as in the previous example. The corresponding inverted model of Q is shown in Figure 10a. One can observe that the inverted Q (with conventional FWI) contains both short and long wavelength perturbations. The background model of the inverted model is on average close to the background of the true model. That is, the algorithm estimated a model with a background that quantitatively approached the true values of Q. Nonetheless, this estimate of the background of Q is still inaccurate and does not have the same distribution in space as that of the true model (Figure 4b). In addition, the inverted model of *Q* shows short-wavelength anomalies, which do not exist in the true model. These artifacts clearly correlate with interfaces of the anomalies in the true models of velocity and *Q*, and they result significantly from a trade-off between the estimation of velocity and *Q*.

The trade-off between the estimated parameters results from the fact that anomalies of velocity and *Q* scatter energy with a similar dependency on direction of the radiation envelope, as pointed out earlier. Then, it is very difficult to determine which perturbed physical property has originated a perturbation in the data. That statement is supported by the radiation patterns of scattered energy (figure 3).

Joint local inversion with regularization

As demonstrated, the correct estimation of the long wavelengths of the model of *Q* is sufficient for a relatively accurate reconstruction of the P-wave velocity model. In addition, the previous example shows that joint local estimation of velocity and intrinsic attenuation is not suitable for carrying out this type of inversion, as a consequence of the strong ambiguity in the estimated models. However, in this case all the wavelength components of the *Q* model are estimated throughout the inversion. Hence, a question that is naturally raised is what happens if the local inversion penalizes the short-wavelength components of the model of intrinsic attenuation. Smoothing the update of *Q* with a local scheme is conceptually equivalent to the effect of utilizing a *reduced basis*. Malinowski et al. (2011) and Kamei and Pratt (2013) reported smoothing regularization over *Q*. The former applied a smoothing operator over the gradient for the model parameter(s), whereas the latter regularized the update of *Q* with a smoothing penalty term.

As pointed out earlier, the joint estimation of P-wave velocity and *Q* suffers from ambiguity in the whole range of wavenumbers. Hence, smoothing the successive updates of *Q* can at best eliminate ambiguities with high wavenumber. For the low wavenumbers this issue is significantly ameliorated, as long as the long wavelengths of the starting velocity model are estimated correctly. This condition is generally met, as it is necessary in order to carry out FWI successfully.

As in the previous example, the starting model of Q is homogeneous and with $Q = 10^5$. The inverted velocity model, depicted in Figure 9d, shows a relatively good reconstruction of both long- and short-wavelength anomalies. However, high-wavenumber artifacts are still observed. Even though the inverted velocity model improved in comparison to the one in Figure 9c, its reconstruction is not as good as the one obtained when the true model of Q is known (Figure 9a). The corresponding inverted model of Q is shown in Figure 10b. One can observe the clear effect of the regularization term as only the long wavelengths of Q are estimated. The values of the inverted model of Q are on average close to the background model (Figure 4d). However, qualitatively one can observe that the models have very different structure. The inaccuracy of the estimated model of Q introduces traveltime errors and errors in the interfering pattern of the modeled wavefield. This introduces overall errors in the simulated recorded phases that will be translated into errors in the estimated velocity model. Hence, one can conclude that the smoothing regularization term is effective in eliminating high-wavenumber artifacts in the inverted Q model, and mitigating some of the high-wavenumber artifacts in the velocity model resulting from a trade-off between estimates (comparing Figures 9c and 9d). However, it is not effective in estimating a sufficiently accurate background Q model. As a result the jointly inverted velocity model is also inaccurate.

Semi-global inversion

The semi-global inversion scheme jointly updates velocity and Q at 3 Hz, using a swarm of 20 particles, performing 21 global iterations each with 3 nested local iterations for velocity only. Only each sixth shot is utilized in a nested local iteration. The whole set of data was used upon completing the semi-global inversion. The starting velocity model is the same as that used in conventional FWI (Figure 4c). The starting models for Q are generated randomly setting a range of variation between 10 and 1000 for this parameter. The search space of Q is parameterized with its logarithm of base 10. The values of Q are estimated at the position of the nodes (in white) overlaid to the true Q model (Figure 10f), except at the ones at the top row. These are forced to match the value of Q in the seawater. The resulting inverted velocity and Q models are depicted in Figures 9e and 10c, respectively. The Q model estimated with the semi-global inversion is very close to the background of the true model. Effectively the semi-global inversion method estimated an accurate background Q model, concerning its structure in space and the magnitude of the recovered anomalies. Those inverted models are used to carry out conventional FWI inversion subsequently. Local FWI starts at 4 Hz and the frequency band is widened 1 Hz after the completion of blocks of 6 iterations up to 16 Hz.

The final inverted model of velocity is depicted in Figure 9f. One can observe that this model is very similar to the inverted velocity model when the true model of Q is used to invert for velocity only (Figure 9a). There is no evidence of existing trade-off between the estimated velocity and the estimated intrinsic attenuation. It is important to note that errors in the estimated models of Q affect the estimates of the velocity model, as one can conclude from the previous examples. There is no evidence of existence of these errors affecting the estimate of the velocity model (Figure 9f), as this model is comparable to that when the true model of intrinsic attenuation is known (Figure 9a).

Projection of the gradient of Q onto a sparse basis

As we demonstrated above, the regularized inversion example does not lead to an accurate update of the long wavelengths of Q. On the other hand, the semi-global inversion algorithm was very effective in estimating a Q model that is very close to the true model. However, one could argue that such discrepancy is because the regularization term behaves much differently than estimating the components of the long wavelengths in a sparse basis. In this example we test that hypothesis by projecting the updates of Q over the same sparse basis as the one used in the semi-global inversion algorithm. See Appendix C for details on the projection on to a sparse-basis.

The inversion is carried out jointly updating the velocity and the intrinsic attenuation over the same bandwidth utilized in the previous local inversion examples. The number of local gradient descent iterations is also the same. The key difference in this example is the projection of the gradient of Q onto the sparse basis. That projection is carried out at each local iteration, after computing the gradient of Q. The projected gradient is then used to compute the update of Q. The starting model of Q is homogeneous with $Q = 10^5$. Figure 9g, depicts the resulting inverted velocity model. One can see that the main features are reconstructed. However, this model shows high-wavenumber artifacts. In addition, the velocity anomalies in the central part of the model are smeared. The inverted model of velocity is comparable to that obtained regularizing the update of Q with a smoothing constraint. The corresponding model of inverted Q is depicted in Figure 10d. The updates of Q converged towards a model that is almost homogeneous and it does not show any of the anomalies present in the background model. However, the values of Q are very low, therefore in a sense the local inversion updates approximated an attenuating medium, which is the main feature of the true model of Q. We carried out several inversion tests with different starting Q models leading to similar results.

Inversion of data affected with noise

The noise is generated with a random variable with Gaussian distribution with zero mean and unit variance. The spectrum of the noise is then matched to that of that data, convolving the noise with the source wavelet. Finally we contaminated the synthetically generated data with the generated noise setting a loss of 5 % in the signal to noise ratio. One should notice that FWI carried out subsequently to the semi-global inversion, relies upon estimates of velocity and Q models that are affected by noise. Hence, in this section we compare our results against the effect of noise when the Q model is known *a priori*.

We first inverted for velocity only with the same number of iterations and frequency bands as used in the previous examples. The model of Q is the true model and it unchanged throughout the iterations. The resulting inverted velocity model is depicted in Figure 9h. One can identify some degradation of the reconstructed velocity model when it is compared with the noise-free data example (figure 9a). More noticeable is the less accurate reconstruction of the high-velocity layer highlighted with the white arrow, and the small-scale structures highlighted with the black ellipse. Nonetheless, the inverted model is still overall relatively accurate.

We then carried out the semi-global inversion with subsequent FWI. We used the same setting as that used previously in the noise-free examples. The final inverted velocity and Q models are depicted in Figures 9i and 10e, respectively. First, one can observe that the quality of the inverted velocity model is comparable to that of the model inverted when the true model of Q is known (figure 9h). That is highlighted with the white arrow and the black ellipse. The inverted model of Q is less accurate than that estimated with noise-free data. This is most visible as the low Q anomaly in the upper-left region has a longer extension than that of the true model. In addition, the overall values of Q are smaller both in the upper-right region and bottom-left region, than those estimated with noise free data, and than those of the true model. However, overall the model is still very accurate, and the final estimate of the velocity model is comparable to that when the true model of Q is known. Therefore there are no inaccuracies related to Q being introduced in the estimates of P-wave velocity.

REAL-DATA EXAMPLE

The field data set was acquired in shallow water in the Norwegian North Sea. This field is a gas-condensate discovery and the reservoir is a fractured chalk formation. The reservoir consists of fractured chalk sitting at the crest of an anticline. This reservoir sits in a seismically obscured region due to the presence of a gas hosted in an overlying formation of interbedded sandstone and silt (Granli et al., 1999; Warner et al., 2013).

A full-azimuth, 3D, ocean-bottom cable survey, was carried out over this field. The cables were 6 km long with inline spacing of 25 m and the cross-line spacing between adjacent cables was 300 m. Figure 11, shows the acquisition geometry with respect to the overlying gas cloud and a well drilled in the area. The shooting was carried out orthogonally to the receiver cables, with 75 m cross-track and 25 m along-track separation, representing a total of 96000 sources for 5760 four-component receivers, covering an area of 180 km^2 approximately. The dark-blue lines, labeled IL for inline, and XL for cross-line, represent the position of selected sections for plotting the velocity and *Q* models in the following examples.

Setting the inversion framework

Low-frequency content in the data is crucial for the successful application of FWI (Bunks et al., 1995; Sirgue and Pratt, 2004). However, real data is affected by noise thus limiting the region of the spectrum that can be used and consequently imposing a restriction on the lowest frequency that can be used in the inversion. The band of frequencies that can be used in the signal is determined from the signal-to-noise ratio. The starting frequency is determined from phase variation within common-receiver gathers by identifying the lowest frequency with coherent signal. For this dataset, it was determined that the starting frequency is 3 Hz.

The data was acquired with four components, from which three components are geophones, which measure particle velocity, and one component is a hydrophone, which measures pressure. In this visco-acoustic inversion, we invert only the hydrophone data. Source and receiver ghosts, the source bubble, surface and internal multiples, pre-critical and post-critical reflections, and wide-angle refractions are all included in the inversion.

Determining a wavelet that simulates the seismic data with significant accuracy over the frequency range of interest is important for inverting data with FWI. The source signature was derived from the direct arrival, and the source and receiver ghosts, and water-bottom multiples were deconvolved. This source wavelet has to be free of ghosts and multiples since the modeling of seismic data includes an explicit free surface, which will re-generate all these arrivals. Warner et al. (2013) gives a comprehensive discussion on the pre-processing of this dataset.

FWI is computationally intensive and it scales with the dimensions of the grid and the number of time steps with $O(n^4)$, where n is the number of cells. The number of shots to be simulated then multiplies this computational cost, representing a large computational load if the volume of data to be inverted is large. This burden is decreased dramatically by considering reciprocity, this is, interchanging receivers with sources and vice-versa. Thus, the number of effective shots in the simulation of data is reduced to 1440 sources (as only the hydrophone components are being used) at each iteration with 96,000 reciprocal receivers. Because the dataset covers the area very densely, a sparse subset of the shots can be used, which further reduces the computational cost (van Leeuwen and Herrmann, 2012). This sparse subset of reciprocal sources is different at each iteration. However, all the data are used after completing each block of iterations for each selected frequency band.

A good starting model has to account accurately for the kinematics of wave propagation and guarantee that the misfit between the recorded and simulated data is within half a cycle. Early tests carried out to determine a starting velocity model demonstrated that it is not possible to match both short and long offsets using an isotropic model. This indicated that seismic anisotropy has to be incorporated. In fact, the existence of anisotropy would be expected from the geological characteristics of the area, due to the fine layering of silt and sand in the overlying gas-cloud region, as well as potentially in the fractured chalk in the reservoir.

Figure 12 shows the starting velocity model. This model is built from reflection tomography by picking residual moveout on the pre-stack time-migrated volume. The parameters of anisotropy ε and δ (Figure 13a-b) are estimated by matching the stacked time-migrated gathers to well information and by matching longer-offset moveout. The starting velocity model matches the kinematics of the real data, presenting an anomaly with low velocity at the center of the model matching the region of the gas cloud. This low-velocity anomaly was inserted in the model to match prior geological information from a sonic log. This sonic log (overlaid in Figure 20) was acquired in a vertical well that crosses the section along the edge of the gas cloud, in the direction orthogonal to that section, down to the reservoir, at the crest of the anticline, and crossing the deeper carbonates with higher velocity.

A close correlation can be observed between the sonic log and the variation of the velocity profile along the vertical, where the velocity of the sonic log decreases in the region of the overlaying gas cloud and an increase in the velocity correlating with the crest of the carbonates structure. A comprehensive description of this seismic dataset pre-processing, as well as the starting velocity- and anisotropy-model building is outlined in Warner et al. (2013).

In this example, inversion is carried out up to 6.5 Hz. The models of anisotropy are smooth and match the kinematics of the waves only, rather than the dynamics. These approximations for velocity anisotropy are valid in the selected frequency range where the dynamics plays a lesser role than the kinematics. The density is derived from Gardners's law. The presence of gas in the overlying interbedded silt and sand creates a low-velocity anomaly, and is responsible for strong intrinsic attenuation of the seismic energy that propagates through this region. This has an impact in the inversion of the seismic data and resulting inverted models if not taken into account.

The inversion was carried out for six frequency bands with a cut-off low-pass filter applied at 3.0, 3.5, 4.1, 4.8, 5.6, and 6.5 Hz. The order of the shots is randomized prior to running the inversion to avoid generating coherent interference patterns. The seismic anisotropy is left fixed throughout the inversion.

Inversion without intrinsic attenuation - conventional FWI without *Q*

FWI without Q is carried out for velocity only, the medium is assumed to be VTI, and intrinsic attenuation is not considered in the constitutive relation. The inversion is carried out iterating 18 times for each frequency band, up to 5.6 Hz, and using every 18th shot. In the last frequency band, the number of iterations is 36, and every 18th shot is used. This means that the overall number of local gradient descent iterations is 126, and that all the data have been used upon completing each block of iterations in each frequency band. Each trace is thus used just seven times during the inversion.

Semi-global inversion followed by conventional FWI with *Q*

The semi-global inversion of the seismic data assumes a medium with anisotropy and intrinsic attenuation. In a first stage, the inversion updates both the velocity and Q with the outlined semi-global inversion algorithm. Then, conventional FWI is used inverting for velocity with local gradient-descent updates, while keeping the Q model fixed. The seismic anisotropy is not updated in any of the stages. The same starting velocity model is used and a population of 12 particles with random models is generated. The random values of Q in the starting model range between 10 and 2000. The values of Q are estimated in the sediments and carbonates regions only, over the grid nodes of the sparse basis as depicted in Figure 12. However, we do not estimate a different value of Q for each node, as this would represent an insurmountable computational cost. Instead we use the structure of the starting velocity model as a prior for defining three distinct regions. We then group the nodes lying within each one of these regions, and the same value of Q is assigned to all nodes within that same group. We consider three principal distinct regions in the model. The first is the gas-cloud (black nodes), the second is the carbonates formation (grey nodes) and the third is the background (white nodes). As in the case of the synthetic example, Q is not estimated at the node positions in the first row. At these positions we set Q fixed and equal to that of the seawater ($Q = 10^5$). Effectively the dimension of the sparse basis is three, i.e. only three values of Q are estimated. Semi-global inversion is carried out with six semi-global iterations at 3.0 Hz and six semi-global iterations at 3.5 Hz, each with three nested local iterations. This means that upon completing the semi-global inversion, the algorithm iterated the velocity model 36 times with 18 iterations at both 3.0 and 3.5 Hz. Thus, the velocity model was updated as many times as in the case of the conventional FWI after completing two bands of frequency. We also further reduced the amount of data used in each local nested iteration, using only every 36th shot. This means that all the data has been used after completing the 36 semi-global iterations. This procedure did not produce artifacts in the estimated velocity model with the semi-global inversion, because the survey is very dense, the inversion frequency band in the signal is relatively low, and the shot positions are randomized prior to inversion. After completing the semi-global iterations, we switched to the setting used in the previous example conventional FWI, holding fixed the *Q* model estimated with the semi-global inversion. The FWI inversion starts at 4 Hz and runs up to 6.5 Hz.

Figures 14a-b, show inline sections of the inverted velocity models using the conventional FWI, without *Q*, at 3.5 and 6.5 Hz, respectively. One can note that at 3.5 Hz the inversion has changed the starting model significantly. Most visibly is the sharpening of the gas cloud and channels, as well as, significant updates of the velocity in the region of the anticline. At 6.5 Hz the gas cloud becomes larger and the velocity of the structure at the bottom decreases. The inversion clearly is forcing an estimate of velocity that is relatively more homogenous, when compared to that at 3.5 Hz. There is significantly less energy propagating closer to the edges, which implies that the estimates of velocity are less constrained towards the edges.

Figure 14c shows the velocity model after carrying out the semi-global inversion up to 3.5 Hz. As pointed out earlier the overall number of nested local iterations is the same as that when carrying out FWI at 3.5 Hz. The resolution of the velocity model depicted in Figure 14c is the same as that of the velocity model shown in Figure 14a. The velocity model depicted in Figure 14c is used as a starting model to carry out conventional FWI, with *Q*, starting at 4 Hz and up to 6.5 Hz. Figure 14d, depicts the final inverted velocity model resulting from semi-global inversion followed by conventional FWI with *Q*. Comparing Figures 14b and 14d, one can identify that both inversion approaches lead to comparable velocity models. Both show a sharpening of the gas cloud and of the lateral channels through which gas has travelled. In both cases, the structure at the top of the anticline, corresponding to the reservoir of fractured chalk, is resolved. An evident difference between the resulting inverted velocity models is the fact that the gas cloud shows higher heterogeneity and is sharper when intrinsic attenuation is considered in the inversion. In particular, the bottom region of the gas cloud is now clearer, and is almost detached from the main gas cloud. We show the corresponding velocity models inverted with FWI and the semi-global method at 6.5 Hz, along the cross-line direction, in Figures 15a and 15b, respectively.

The overall magnitude of the P-wave velocity estimated without taking intrinsic attenuation into account (Figures 14b and 15a) is higher than that estimated when intrinsic attenuation is taken into account (Figures 14d and 15b). That is because when considering intrinsic attenuation in the modeling the group velocity is slower, thus the velocity has to increase in order to match the recorded phases in the data, as these are invariant in the two examples, with and without intrinsic attenuation. This is especially evident in the formation of carbonates bellow 3 km of depth, where the velocity is clearly higher when considering intrinsic attenuation (comparing Figures 14b and 14d with Figures 15a and 15b).

The resulting model of intrinsic attenuation along the selected cross-line is depicted in Figure 16. The estimated model of Q is a macro-model since it is represented over a sparse support. The inverted model of Q shows three main regions. In the central part of the model, is estimated an anomaly that matches the position of the gas cloud with a low value of Q, thus correlating with the physical properties of this region. The presence of gas is responsible for strong intrinsic attenuation and absorption of energy, thus demonstrating a clear matching between the estimated values of Q from the real data and the physical phenomena that occurs in this region. The other two regions are formed by sediments ($Q \approx 200$) and by the carbonates ($Q \approx 1000$) in the deeper part. The estimated values of Q in these two regions are also in agreement with the geological conditions as the carbonates form a large macroscopic region that is relatively homogenous, whereas the sediments are less consolidated than the carbonates. Thus, it is expected that the formations with sediments exert stronger intrinsic attenuation than the carbonates. However, the model of Q does not have enough resolution to capture the fractured chalk at the top of the crest where stronger intrinsic attenuation should in principle occur due to the presence of these fractures.

Figure 17 compares the objective function when carrying out FWI only, and semi-global inversion prior to FWI. The values of the objective function of the semi-global inversion are compared at each third iteration of FWI, as the value of the objective function is obtained after completing three nested local iterations. It is important to note that, in this example, the semi-global inversion uses less data per iteration than conventional FWI. Then the values of the objective function over the first thirty-six iterations are not directly comparable. One can observe that carrying out FWI after the semi-global inversion method, and accounting for Q yields an improved data fitting.

Figure 18, compares the data misfit using the semi-global inversion and conventional FWI after completing the inversion, for each 75th trace in a receiver-gather (position of the red star in Figure 11). The black bars represent the data fitting when inverting for velocity only and disregarding Q. The grey bars represent the resulting data fitting when inverting for velocity and Q utilizing the semi-global inversion scheme. The range of misfit in the histogram is normalized by the maximum overall value of misfit. One can see a consistent improvement of the data fitting for each trace in the receiver gather, as a result of our semi-global inversion method. This resulting improvement in the data misfit results principally from computing waveforms with improved amplitude as a result of considering attenuation.

The difference between traces computed with and without the estimated Q model is mainly noticeable when comparing traces individually. Figure 19, compares traces for the same receiver gather. The traces are selected at the positions denoted with red circles and labeled *a* to *d* in Figure 11. The relative shot position is that of the red star. The amplitude of each trace has been scaled in order to facilitate the comparison. The traces on the left are real data, the traces at the center are computed considering the inverted Q (with the corresponding velocity model), and the traces on the right are computed disregarding Q, utilizing the velocity model resulting from conventional FWI. One can observe how the waveforms computed with the estimated visco-acoustic model became closer to those of the real data. Hence, the semi-global inversion algorithm determined a model of the subsurface with a response closer to that recorded in the real data.

We further justify our result with a sonic log. Figure 20, compares the vertical profiles of the velocity extracted at the location of the well (denoted by the black circle in Figure 11) with a sonic log. One can see an overall good agreement between the two estimated models. The difference between the inverted models is not significant down to about 2 km of depth corresponding to the bottom of the gas cloud. With increasing depth, the effect of correcting the inversion with intrinsic attenuation becomes more noticeable as the errors in the modeling accumulated when not accounting for intrinsic attenuation, and the difference is more evident, demonstrating an overall better agreement between the estimated velocity model using the combined semi-global scheme and the sonic log, when compared to the inverted velocity model not accounting for intrinsic attenuation. The estimate of the velocity in the layer with carbonates is also significantly more accurate when considering intrinsic attenuation. As mentioned above, the velocity model estimated with the semi-global inversion scheme shows a second smaller region with gas separated from the main overlying gas deposit. This is also validated with the sonic log, showing that the vertical profile of the velocity has a better match with respect to the sonic log just below 2500 m of depth, where this separation occurs.

COMPUTATIONAL ASPECTS

In this section, we compare the computational cost of the semi-global inversion, $\mathcal{O}(G)$, against that of conventional FWI, $\mathcal{O}(L)$. This comparison, as outlined herein, is valid whenever the nested local iterations of the semi-global algorithm have the same computational complexity as that of the FWI implementation, regardless of the order of the local minimization algorithm (first- or second-order).

The overwhelming computational load of local minimization algorithms, in large-scale applications, comes from the solution of the forward problem. In this case, it comes from the computation of wavefields. These wavefields are computed several times per local iteration, as the adjoint-state method requires the solution of both the state-variable (wavefields) and of the adjoint-variable over a grid. The dimension of these grids can range between several hundred thousand, in 2D, and several million or thousand of million in typical large-scale 3D applications. In addition, a large amount of memory needs to be allocated in order to store those wavefiels, as well as the physical parameters, and respective updates. Any other operations and memory storage are insignificant, when compared to propagating the wavefields over a grid.

On the other hand, the key operations carried out at each QPSO iteration are generating a set of random numbers, carrying out the update of the position of each particle, keeping track of the best position for each particle, and keeping track of the evolution of the cost function. The computational load of those operations is detrimental when compared to that of carrying out adjoint-state operations. Hence, the computational cost of a local nested iteration is the key factor driving that of the semi-global inversion algorithm. Table 1, compares the computational cost of the conventional FWI with that of the semi-global inversion method. One can immediately observe that the number of particles, N_p, is the differentiating factor in driving the computational cost of the semi-global inversion, when compared to conventional FWI. The total number of local iterations N_L', is determined by the number of local nested iterations, N_n, the number of outer global iterations, N_g, and the number of frequency bands N_f'. If the number of global and nested global iterations remain constant over each frequency band, then the total number of local iterations carried out by the semi-global inversion algorithm is simply $N_L' = N_g N_n N_f'$.

Then, we can get the relative computational costs of the examples outlined. The semi-global synthetic data example runs a total of $N_L' = 63$ local iterations distributed over 21 global iterations and one frequency band, and each iteration utilizes one sixth of the data, that is, $N_S' = N_S/6$. The relative cost of the semi-global inversion is then $\mathcal{O}(G) = N_p N_L N_S 63/504$, or $\mathcal{O}(G) = 2.5\mathcal{O}(L)$, showing that the computational cost of the semi-global inversion method, is competitive to that of typical applications of FWI.

In the 3D real-data inversion example, the local inversion completed a total of $N_L = 126$ local iterations utilizing N_s shots per iteration. The semi-global inversion carried out a total of $N_L' = 36$ iterations and utilizing $N_S' = N_S/2$, per iteration, and $N_p = 12$. Then, the relative cost of the semi-global inversion method is $\mathcal{O}(G) = N_L N_S \times (12 \times 36)/(126 \times 2)$ or $\mathcal{O}(G) \approx 1.7\mathcal{O}(L)$. One can see that the computational cost of the semi-global inversion is very comparable to that of conventional FWI over a typical band of frequencies. It is important to note that FWI was carried out in a relatively narrow band of frequency. If the FWI were upscaled up to 10 Hz, this relation would become even more favorable for the semi-global inversion method, approaching a relation close to $\mathcal{O}(G) \sim \mathcal{O}(L)$.

CONCLUSION

We introduced a semi-global inversion method for the joint estimation of P-wave velocity and intrinsic attenuation. The method relies upon the use of local iterations updating velocity only, nested within a loop of global iterations for the estimation of quality factor (Q) on a reduced basis. We have demonstrated the feasibility of using a reduced basis when estimating Q with a numerical example, showing that a good estimate of the macro-model of intrinsic attenuation suffices for the estimation of a velocity model from seismic data affected by intrinsic attenuation. The accuracy of that estimated velocity model is of the order of that when the true intrinsic attenuation model is known. Synthetic examples demonstrated that unlike pure gradient-descent methods, semi-global inversion does not suffer from noticeable cross-talk between estimated parameters, and it works in complex rheology in the presence of anisotropy. In addition, we also demonstrated that smoothing regularization combined with local optimization, or updating Q over a sparse basis with a local descent method is not suitable for estimating the long wavelengths of the model of Q. Hence, the outer global update of intrinsic attenuation clearly yields improved joint estimates of intrinsic attenuation and P-wave velocity, outperforming any of the approaches outlined based exclusively on local search methods.

A real-data example demonstrated the feasibility of the method for the inversion of data recorded over large 3D surveys. This example demonstrated that the semi-global algorithm estimated a model from the real data that matches the known geological conditions of the area, and improved the match between the estimated velocity model and a sonic log acquired in a well drilled in the area. In addition, the waveforms of the predicted data match more closely those of the recorded data. This demonstrates that the semi-global inversion method outlined here improved estimation of the models of the subsurface – both for intrinsic attenuation and for intrinsic attenuation-corrected P-wave velocity.

ACKNOWLEDGEMENTS

The authors are grateful to the comments and insights of assistant editor Deyan Draganov, associate editor Partha Routh, Rie Kamei, Madhav Vyas and an anonymous reviewer who helped to improve the outline of this paper significantly. The authors are also thankful to the sponsors of the FULLWAVE Game Changer research consortium, under the IFT research program.

APPENDIX A – COMPUTATION OF THE GRADIENT OF THE OBJECTIVE FUNCTION WITH THE ADJOINT-STATE METHOD

The gradient of the objective function 8 with respect to the model parameters is given by

$$\nabla_{\boldsymbol{m}} J_F(\boldsymbol{p},\boldsymbol{m}) = \nabla_{\boldsymbol{m}} J(\boldsymbol{p},\boldsymbol{m}) + \lambda^2 \nabla_{\boldsymbol{m}_2} J_R(\boldsymbol{m}_2). \tag{A1}$$

The gradient of the regularizing term is straightforward to compute and is given by

$$\nabla_{\boldsymbol{m}_2} J_R(\boldsymbol{Q}) = \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{m}_2. \tag{A2}$$

In the case of non-regularized solutions, equation A1 becomes

$$\nabla_{\boldsymbol{m}} J_F(\boldsymbol{p},\boldsymbol{m}) = \nabla_{\boldsymbol{m}} J(\boldsymbol{p},\boldsymbol{m}). \tag{A3}$$

The gradient of the data misfit term, $\nabla_{\boldsymbol{m}} J(\boldsymbol{p},\boldsymbol{m})$, is computed in a discretize-then-optimize fashion by introducing the Lagrangian functional

$$\mathcal{L}(\boldsymbol{p},\boldsymbol{\lambda},\boldsymbol{m}) = J(\boldsymbol{p}) + \langle \boldsymbol{\lambda}^T, \boldsymbol{A}(\boldsymbol{m})\boldsymbol{p} - \boldsymbol{s} \rangle_S, \tag{A4}$$

where $\boldsymbol{p}$, $\boldsymbol{m}$, and $\boldsymbol{\lambda}$ are the discretized state variables (pseudo-pressure in this case), the model parameters and the adjoint variables, respectively, $\boldsymbol{A}$ is the discretized forward-modeling operator. The angle brackets denote an inner product defined over the space of source functions S, and T in super-script denotes transpose. When the constraint is satisfied $\mathcal{L}(\boldsymbol{p}(\boldsymbol{m}),\boldsymbol{\lambda},\boldsymbol{m}) = J(\boldsymbol{p}(\boldsymbol{m}))$, and $\nabla_{\boldsymbol{m}}\mathcal{L}(\boldsymbol{p}(\boldsymbol{m}),\boldsymbol{\lambda},\boldsymbol{m}) = \nabla_{\boldsymbol{m}}J(\boldsymbol{p}(\boldsymbol{m}))$. Finally, forcing the condition $\nabla_{\boldsymbol{p}}\mathcal{L}(\boldsymbol{p},\boldsymbol{\lambda},\boldsymbol{m}) = \boldsymbol{0}$, yields

$$\begin{cases} \boldsymbol{A}\boldsymbol{p} = \boldsymbol{s} \\ \boldsymbol{A}^T\boldsymbol{\lambda} = -\nabla_{\boldsymbol{p}}J(\boldsymbol{p},\boldsymbol{d}) \\ \quad \nabla_m J(\boldsymbol{p},\boldsymbol{\lambda},\boldsymbol{m}) = \langle \boldsymbol{\lambda}^T, [\nabla_{\boldsymbol{m}}\boldsymbol{A}]\boldsymbol{p} \rangle_S \end{cases} . \tag{A5}$$

The first equation is the discrete forward-modeling operator, the second equation determines the adjoint field, and the third is the decision equation determining the update of the model parameters. The mathematical operations described by equation A5 are carried out at each gradient-descent iteration.

APPENDIX B – RADIATION PATTERN IN A VISCO-ACOUSTIC ANISOTROPIC MEDIUM

The derivation of the radiation patterns for each one of the parameters is more conveniently carried out in the frequency domain. In the frequency domain the constitute law in equation 1 is given by

$$\boldsymbol{\sigma}(\omega) = i\omega\boldsymbol{C}(\omega)\boldsymbol{\varepsilon}(\omega), \tag{B1}$$

where ω is the angular frequency. The Cauchy's law of motion is given in the frequency domain by

$$i\omega\rho\boldsymbol{v}(\omega) = \nabla\boldsymbol{\sigma}(\omega) + \boldsymbol{F}_{\boldsymbol{v}}(\omega). \tag{B2}$$

Note that all the fields and physical properties are implicitly dependent on the position in space. However, we omit that explicit dependency in the notation for a matter or simplification.

Setting $p_h(\omega) = \sigma_{xx}(\omega) = \sigma_{yy}(\omega)$ and $p_n(\omega) = \sigma_{zz}(\omega)$, substituting B1 into B2 and using the constitutive law 1 leads to the system of equations

$$\begin{cases} -\frac{\omega^2}{\rho v_R^2} p_h - i\omega\tilde{f}(1+\varepsilon_R)\nabla_H \cdot \left(\frac{1}{\rho}\nabla_H p_h\right) - i\omega\tilde{f}\sqrt{1+2\delta_R}\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial p_n}{\partial z}\right) = s(\omega) \\ -\frac{\omega^2}{\rho v_R^2} p_n - i\omega\tilde{f}\sqrt{1+2\delta_R}\nabla_H \cdot \left(\frac{1}{\rho}\nabla_H p_h\right) - i\omega\tilde{f}\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial p_n}{\partial z}\right) = s'(\omega) \end{cases}, \quad \text{(B3)}$$

where $\tilde{f}$ is the Fourier transform of the relaxation function. The system of equations B3 is transformed into

$$\begin{cases} -\frac{\omega^2}{\rho v_R^2} p - i\omega\tilde{f}(1+2\delta_R)\nabla_H \cdot \left(\frac{1}{\rho}\nabla_H (p+q)\right) - i\omega\tilde{f}\sqrt{1+2\delta_R}\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial}{\partial z}\frac{p}{\sqrt{1+2\delta_R}}\right) = s(\omega) \\ -\frac{\omega^2}{\rho v_R^2} q - i2\omega\tilde{f}(\varepsilon_R - \delta_R)\nabla_H \cdot \left(\frac{1}{\rho}\nabla_H (p+q)\right) = 0 \end{cases},$$

(B4)

first setting $p_n \leftarrow p_n/\sqrt{1+2\delta_R}$ and then defining $p = p_n$ and $q = p_h - p_n$. This equation has the same structure as the one used for the analysis of radiation patterns in Alkhalifah and Plessix (2014), apart from the frequency response of the medium, introduced herein, given by the association of the frequency response of the relaxation mechanism and the relaxed physical properties. Hence we can now take the same rationale for deriving the radiation patterns for an attenuating medium.

For the computation of the frequency response of the relaxation function we consider a relaxation mechanism with only one body. In the time domain the stress-relaxation function 2 can be written in terms of the quality factor as (Fichtner, 2011)

$$f(t) = \left(1 + \frac{1}{KQ}e^{-t/\tau^{\sigma}}\right)H(t), \quad \text{(B5)}$$

where K is a constant (see Fichtner (2011) for more details on the expression for K). Expression B5 has Fourier transform

$$\tilde{f} = \frac{1}{i\omega}\left(1 + \frac{\gamma}{KQ}\right), \tag{B6}$$

where $\gamma = i\omega\tau^{\sigma}/(1 + i\omega\tau^{\sigma})$ is constant as $\tau^{\sigma} = 1/\omega_{\sigma}$ (Liu, 1976; Blanch et al., 1995).

In this analysis we only consider a medium with Vertical Transverse Isotropy (VTI) in the velocity and with an isotropic model of intrinsic attenuation. The analysis is carried out using the Born approximation, decomposing fields and physical properties into background and perturbed components

$$v = v_0(1 + r_v), \tag{B7a}$$

$$Q = Q_0\left(1 - r_Q\left(1 + \frac{KQ_0}{\gamma}\right)\right), \tag{B7b}$$

$$\varepsilon = \varepsilon_0 + r_{\varepsilon}, \tag{B7c}$$

$$\delta = \delta_0 + r_{\delta}, \tag{B7d}$$

The background medium is assumed to be attenuating and isotropic thus $\varepsilon_0 = \delta_0 = 0$, and it is also assumed that the density of the medium is constant $\nabla\rho = 0$. The reason for the factor $1 + KQ_0/\gamma$ appearing in expression B7b will become self-evident along the derivation of the radiation pattern for Q. We do not derive the radiation pattern for density, as it is irrelevant for the current discussion. In addition, the fact that the density is assumed to be constant does not affect the analysis of the other radiation patterns, as long as, the density is an independent parameter and it is not coupled with any other (e.g. velocity), as in the case of considering a parameterization with impedance for example.

First, we substitute expression B7b into equation B6 and use a Maclaurin series for $1/Q$, yielding

$$i\omega\tilde{f} \approx \left(1+\frac{\gamma}{KQ_0}\right)(1+r_Q). \tag{B8}$$

Second we introduce the identity

$$\bar{v}_0^2 = v_0^2\left(1+\frac{\gamma}{KQ_0}\right). \tag{B9}$$

Defining the pseudo-stresses in background and perturbed components $p = p_0 + p_1$ and $q = q_0 + q_1$, the second equation of system B4 becomes,

$$-\frac{\omega^2}{\rho\bar{v}_0^2}(1-2r_v)(q_0+q_1) - 2(1+r_Q)(r_\varepsilon - r_\delta)\frac{1}{\rho}\nabla_H^2(p_0+p_1+q_0+q_1) = 0. \tag{B10}$$

Eliminating the second-order terms and equation background and perturbed terms, leads to,

$$\begin{cases} q_0 = 0 \\ \frac{\omega^2}{\bar{v}_0^2}q_1 = -2(r_\varepsilon - r_\delta)\nabla_H^2 p_0 \end{cases}. \tag{B11}$$

One can observe that the wavefield q_1 is not affected by perturbations in intrinsic attenuation. Using the same rationale over the first equation of system B4 leads to the wave equation for the background medium

$$-\frac{\omega^2}{\rho\bar{v}_0^2}p_0 - \frac{1}{\rho}\nabla^2 p_0 = s, \tag{B12}$$

and the wave equation for the perturbed field p_1 is

$$-\frac{\omega^2}{\rho\bar{v}_0^2}p_1 - \frac{1}{\rho}\nabla^2 p_1 = -\frac{\omega^2}{\rho\bar{v}_0^2}2r_v p_0 + r_Q\frac{1}{\rho}\nabla^2 p_0 - \frac{2\bar{v}_0^2}{\rho\omega^2}\nabla_H^2\left(r_\varepsilon\nabla_H^2 p_0\right) + \frac{2\bar{v}_0^2}{\rho\omega^2}\nabla_H^2\left(r_\delta\nabla_H^2 p_0\right) + \frac{2}{\rho}r_\delta\nabla_H^2 p_0 - \frac{1}{\rho}\left(\partial_z^2 r_\delta p_0 - r_\delta\partial_z^2 p_0\right). \tag{B13}$$

The variable q_1 is eliminated substituting the second expression of equation B11 where appropriate. The background and perturbed pseudo-pressures are computed from the Green's function for the background medium, $G(\boldsymbol{x},\boldsymbol{x}',\omega)$, defined as,

$$-\frac{\omega^2}{\rho\bar{v}_0^2}G(\boldsymbol{x},\boldsymbol{x}',\omega)-\nabla^2 G(\boldsymbol{x},\boldsymbol{x}',\omega)=\delta(\boldsymbol{x}-\boldsymbol{x}'). \quad \text{(B14)}$$

For a source term $s = s(\omega)\delta(\boldsymbol{x}-\boldsymbol{x_s})$ the wavefield in the background medium is given by,

$$p_0(\boldsymbol{x}) = G(\boldsymbol{x},\boldsymbol{x}_s,\omega)s(\omega). \quad \text{(B15)}$$

The Green's function for a homogeneous background with sources at $\boldsymbol{x_s}$ and virtual sources at $\boldsymbol{x_r}$ (the receiver position) is approximated asymptotically as

$$G(\boldsymbol{x},\boldsymbol{x_s},\omega) \propto e^{i\boldsymbol{k_s}\cdot\boldsymbol{x}}, \quad \text{(B16)}$$

and

$$G(\boldsymbol{x_r},\boldsymbol{x},\omega) \propto e^{i\boldsymbol{k_r}\cdot\boldsymbol{x}}, \quad \text{(B17)}$$

respectively, with

$$\boldsymbol{k}_s = \frac{\omega}{\bar{v}_0}(p_{sh},p_{sz}) = \frac{\omega}{\bar{v}_0}(\sin(\theta/2),\cos(\theta/2)), \quad \text{(B18)}$$

and,

$$\boldsymbol{k}_r = \frac{\omega}{\bar{v}_0}(p_{rh},p_{rz}) = \frac{\omega}{\bar{v}_0}(-\sin(\theta/2),\cos(\theta/2)), \quad \text{(B19)}$$

where θ defines the aperture between source and receiver. Then the perturbed wavefield p_1 is determined from

$$p_1(\boldsymbol{x_r},\boldsymbol{x}_s,\omega) = -s(\omega)\omega^2\int d\boldsymbol{x}\frac{G(\boldsymbol{x_r},\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x_s},\omega)}{\rho\bar{v}_0^2}2r_v + s(\omega)\int d\boldsymbol{x}G(\boldsymbol{x_r},\boldsymbol{x},\omega)\frac{r_Q}{\rho}\nabla^2 G(\boldsymbol{x},\boldsymbol{x_s},\omega) - s(\omega)\frac{2\bar{v}_0^2}{\rho\omega^2}\int d\boldsymbol{x}G(\boldsymbol{x_r},\boldsymbol{x},\omega)\nabla_H^2\left(r_\varepsilon\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x_s},\omega)\right) + s(\omega)\frac{2\bar{v}_0^2}{\rho\omega^2}\int d\boldsymbol{x}G(\boldsymbol{x_r},\boldsymbol{x},\omega)\nabla_H^2\left(r_\delta\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x_s},\omega)\right) + s(\omega)\int d\boldsymbol{x}\frac{2r_\delta}{\rho}G(\boldsymbol{x_r},\boldsymbol{x},\omega)\nabla_H^2 G(\boldsymbol{x},\boldsymbol{x_s},\omega) - s(\omega)\int d\boldsymbol{x}\frac{1}{\rho}G(\boldsymbol{x_r},\boldsymbol{x},\omega)\left(\partial_z^2 r_\delta G(\boldsymbol{x},\boldsymbol{x_s},\omega) - r_\delta\partial_z^2 G(\boldsymbol{x},\boldsymbol{x_s},\omega)\right). \quad \text{(B20)}$$

Computing the expressions of the derivatives of the Green's functions, substituting those where appropriate, and integrating by parts expression B20 yields,

$$p_1(\boldsymbol{x_r},\boldsymbol{x}_s,\omega) = -s(\omega)\omega^2\int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}2r_v - s(\omega)\omega^2\int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}r_Q - s(\omega)\omega^2 \int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}r_\varepsilon 2p_{sh}^2p_{rh}^2 + s(\omega)\omega^2\int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}r_\delta 2p_{sh}^2p_{rh}^2 - s(\omega)\omega^2 \int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}r_\delta 2p_{sh}^2 + s(\omega)\omega^2\int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}r_\delta\left(p_{rz}^2 - p_{sz}^2\right). \tag{B21}$$

The last term in expression B21 is eliminated as $p_{sz}=p_{rz}$ for the given Green's functions, which do not take into account tilted anisotropy. Finally expression B21 is written in a compact form

$$p_1(\boldsymbol{x_r},\boldsymbol{x}_s,\omega) = -s(\omega)\omega^2\int d\boldsymbol{x}\frac{G(\boldsymbol{x}_r,\boldsymbol{x},\omega)G(\boldsymbol{x},\boldsymbol{x}_s,\omega)}{\rho\bar{v}_0^2}\boldsymbol{r}(\boldsymbol{x})\cdot\boldsymbol{\vartheta}(\boldsymbol{x}), \tag{B22}$$

with,

$$\boldsymbol{r} = (r_v, r_Q, r_\varepsilon, r_\delta), \tag{B23}$$

and,

$$\boldsymbol{\vartheta} = \left(2,1,2p_{sh}^2p_{rh}^2, 2p_{sh}^2 - 2p_{sh}^2p_{rh}^2\right) = \left(2,1,2\sin^4\frac{\theta}{2}, 2\sin^2\frac{\theta}{2}\cos^2\frac{\theta}{2}\right). \tag{B24}$$

When $Q_0 \rightarrow \infty$ the second term disappears and the radiation pattern in B24 becomes that of an acoustic VTI medium without intrinsic attenuation. The radiation pattern, $\boldsymbol{\vartheta}$, is plotted in Figure 3. The same approach can be used to derive radiation patterns for other parameterizations and other anisotropy relations, including the model of intrinsic attenuation.

APPENDIX C – PROJECTION ONTO A SPARSE BASIS

In this appendix we outline the approach for projecting the gradient of the objective function with respect to Q onto a sparse basis. From a general point of view as it can be applied to any other examples requiring a projection of a vector over a different basis. This includes any quantity that takes a discrete form (discrete gradient or discrete physical property).

Let a vector $\boldsymbol{u}'$ be defined over a sparse basis, with dimension *N*, and a vector $\boldsymbol{u}$ defined over a full basis, with dimension *M*, and $N < M$. We can assume that there is a linear mapping between these two vectors defined as

$$\boldsymbol{R}\boldsymbol{u}' = \boldsymbol{u}, \tag{C1}$$

where $\boldsymbol{R}$ is a linear operator with dimensions $M \times N$. Projecting a vector $\boldsymbol{u}$ defined over the full basis onto a vector $\boldsymbol{u}'$ defined over the sparse basis requires the solution of the inverse of the linear system C1. This system is underdetermined and its least-squares solution is (Golub and van Loan, 2012)

$$\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{u}' = \boldsymbol{R}^T\boldsymbol{u}. \tag{C2}$$

In our particular case, the gradient of the objective function with respect to *Q* is estimated in the full basis, as it results from applying the adjoint-state method to the wave operator defined over that basis. The gradient with respect to *Q*, computed with the adjoint-state method, is effectively $\boldsymbol{u}$ in equation C2. The gradient with respect to *Q* projected onto the sparse basis is denoted by $\boldsymbol{u}'$. The solution of equation C2 is obtained at each non-linear gradient descent iteration. The operator $\boldsymbol{R}$ depends on how the mapping between the sparse and full basis is defined. In the case of splines, it is sparse and its entries are defined by the coefficients of the chosen spline. This approach can be utilized with any other suitable mapping between a sparse and a full basis.

REFERENCES

Abubakar, A., T.M. Habashy, Y. Lin, and M. Li, 2012, A model-compression scheme for nonlinear electromagnetic inversions: Geophysics, 77(5), E379-E389.

Abubakar, A., G. Pan, M. Li, L. Zhang, T.M. Habashy, and P. van den Berg, 2011, Three-dimensional seismic full-waveform inversion using the finite-difference contrast source inversion method: Geophysical Prospecting, 59, 874-88.

Aki, K., and G. Richards, 2002, Quantitative Seismology (2nd Edition): University Science Books.

Alkhalifah, T., and R-É. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, 79(3) R91-R101.

Aster, R., B. Borchers, and C.H. Thurber, 2012, Parameter Estimation and Inverse Problems (2nd Edition): Academic Press. 376 pp.

Afanasiev, M.V., R.G. Pratt, R. Kamei, and G. McDowell, 2014, Waveform-based simulated annealing of crosshole transmission data: a semi-global method for estimating seismic anisotropy: Geophysical Journal International, 199 (3): 1586-1607.

Agudo, Ò.C., N.V. da Silva, M. Warner, T. Kalinicheva, and J. Morgan, 2018, Addressing viscous effects in acoustic full-waveform inversion: Geophysics, 83(6), R611-R628.

Bai, J., D. Yingst, R. Bloor, and J. Leveille, 2014, Viscoacoustic waveform inversion of velocity structures in the time domain: Geophysics, 79(3), R103-R119.

Blanch, J.O., J.O. Robertsson, and W.W. Symes, 1995, Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique: Geophysics, 60(1), 176 184.

Bland D.R., 1960, Theory of linear viscoelasticity (International series and monographs on pure and applied mathematics): Pergamon Press.

Bunks, C., M. F.M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, 60(5), 1457-1473.

da Silva, N V., A. Ratcliffe, V. Vinje, and G. Conroy, 2016, A new parameter set for anisotropic multiparameter full-waveform inversion and application to a North Sea data set: Geophysics, 81(4), U25-U38.

da Silva, N.V. and G. Yao, and M. Warner, 2018a, Semi-global inversion of V_p to V_s ratio for elastic wavefield inversion: Inverse Problems, 34, 115011, 21pp.

da Silva, N.V., G. Yao, and M. Warner, 2018b, Wave modeling in visco-acoustic media with transverse isotropy (TI): Geophysics, accepted.

Daubechies, I., 1992, Ten lectures on wavelets, 2nd ed.: SIAM, 377 pp.

Datta, D., and M.K. Sen, 2016, Estimating a starting model for full-waveform inversion using a global optimization method: Geophysics, 81(4), R211-R223.

Debens, H., 2015, Three-Dimensional Anisotropic Full-Waveform Inversion: PhD Thesis, Imperial College London.

Debens, H.A., M. Warner, A. Umpleby, and N.V. da Silva, 2015, Global anisotropic 3D FWI: SEG Technical Program Expanded Abstracts 2015, 1193-1197. https://doi.org/10.1190/segam2015-5921944.1

Diaz, J., and P. Joly, 2006, A time domain analysis of PML models in acoustics: Computer Methods in Applied Mechanics and Engineering, 195, no. 29-32, 3820–3853.

Diouane, Y. S. Gratton, X. Vasseur, L.N. Vicente, and H. Calandra, 2016, A parallel evolution strategy for an earth imaging problem in geophysics: Optimization and Engineering, 17, 3-26.

Farquharson, C.G., and D. Oldenburg, 2004, A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems: Geophysical Journal International, 156(3), 411-425.

Fernández-Berdaguer, E., 1998, Parameter Estimation in Acoustic Media Using the Adjoint Method: SIAM Journal on Control and Optimization, 36(4), 1315-1330.

Fichtner, A., 2011, Full seismic waveform modelling and inversion (Advances in geophysical and environmental mechanics and mathematics): Springer, 364pp.

Fichtner, A., H.-P. Bunge, and H. Igel, 2006, The adjoint method in seismology: Physics of the Earth and Planetary Interiors, Volume 157(1), 86-104.

Golub, G., and C.F. van Loan, 2012, Matrix computations (John Hopkins Studies in the mathematical sciences), 4th edition, John Hopkins University Press, 784 pp.

Hestholm, S., S. Ketcham, R. Greenfield, M. Moran, and G. McMechan, 2006, Quick and accurate Q parameterization in viscoelastic wave modeling: Geophysics, 71(5), T147-T150.

Hicks, G., and R.G. Pratt, 2001, Reflecting waveform inversion using local descent methods: estimating intrinsic attenuation and velocity over a gas-sand deposit: Geophysics 66(2), 598–612.

Kamei, R., and R.G. Pratt, 2013, Inversion strategies for viscoacoustic waveform inversion: Geophysical Journal International, 194(2), 859–884.

Kennedy, J., and R.G. Eberhart, 1995, Particle swarm optimization: Proceedings of IEEE International Conference on Neural Networks, 1942-1948.

Kennedy, J., and R.C. Eberhart, 2001, Swarm Intelligence (The Morgan Kaufmann Series in Artificial Intelligence): Morgan Kaufmann Publishers, 512 pp.

Levin, F.S., 2001, An introduction to quantum theory: Cambridge University Press, 808 pp.

Liao, Q., and G.A. McMechan, 1993, 2-D pseudo-spectral viscoacoustic modeling in a distributed-memory multi-processor computer: Bulletin of the Seismological Society of America, 83(5), 1345-1354.

Liao, Q., and G.A. McMechan, 1996, Multifrequency viscoacoustic modeling and inversion: Geophysics, 1996, 61(5), 1371-1378.

Malinowski, M., S. Operto, and A. Ribodetti, 2011, High-resolution seismic intrinsic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion: Geophysical Journal International, 186(3), 1179-1204.

Nocedal, J., and S.J. Wright, 2006, Numerical optimization (Springer series in operations research and financial engineering), 2nd edition: Springer, 688 pp.

Operto S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain multi-parameter full-waveform inversion of ocean bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, 202(2), 1362-1391.

Plessix, R.-E., and Q. Cao, 2011, A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium: Geophysical Journal

International, 185(1), 539-556.

Plessix, R.E., A., Stopin, H. Kuehl, V. Goh, and K. Overgaag, 2016, Visco-acoustic Full Waveform Inversion: 78th EAGE Conference and Exhibition 2016, SRS2 04.

Pratt, R.G., F. Hou, K. Bauer, and M.H. Weber, 2004, Waveform tomography images of velocity and inelastic attenuation from the Mallik 2002 Crosshole Seismic Surveys, Scientific Results from the Mallik 2002 Gas Hydrate Production Research Well Program, Mackenzie Delta, North Territories, Canada, (ed.) S.R. Dallimore and T.S. Collet: Geological Survey of Canada, Bulletin, 585, 14 pp.

Pratt, R.G., C. Shin, and G.J. Hicks, 1998, Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion: Geophysical Journal International, 133(2), 341-362.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, 2007, Numerical recipes - The art of scientific computing, 3rd edition: Cambridge University Press, 1256 pp.

Routh, P., R. Neelamani, R. Lu, S. Lazaratos, H. Braaksma, S. Hughes, R. Saltzer, J. Stewart, K. Naidu, H. Averill, V. Gottumukkula, P. Homonko, J. Reilly, and D. Leslie, 2017, Impact of high-resolution FWI in the Western Black Sea: revealing overburden and reservoir complexity: The Leading Edge, 36(1), 60-66.

Rao, Y., and Y. Wang, Y., Seismic intrinsic attenuation in fractured media: Journal of Geophysics and Engineering, 12(1), 26–32.

Ribodetti A., and A. Hanyga, 2004, Some effects of the memory kernel singularity on wave propagation and inversion in viscoelastic media – II inversion: Geophysical Journal International, 158(2), 426-442.

Schumaker, L.L., 2015, Spline Functions: computational methods: SIAM, 412 pp.

Sen, M., and P. Stoffa, 2013, Global Optimization Methods in Geophysical Inversion: Cambridge University Press, 302 pp.

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

Sirgue, L., and R.G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, 69(1), 231-248.

Song, Z.-M., P.M. Williamson, and R.G. Pratt, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part II - Inversion method, synthetic experiments and real-data results: Geophysics 60(3), 796-809.

Sun, J., B. Feng, and W. Xu, 2004a, Particle swarm optimization with particles having quantum behavior: Congress on Evolutionary Computation '04: Institute of Electrical and Electronics Engineers, 325–331.

Sun, J., Feng, B., and W. Xu, 2004b, A global search strategy of quantum-behaved particle swarm optimization: IEEE Conference on Cybernetics and Intelligent Systems, 111-116, 2004.

Sun, J., W. Fang, X. Wu, V. Palade, and W. Xu, 2012, Quantum-behaved particle swarm optimization: Analysis of individual particle behavior and parameter selection: Evolutionary

Computation, 20(3), 349–393.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, 49(8), 1259-1266.

Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms including elasticity and intrinsic attenuation: Pure and Applied Geophysics, 128, 365-399.

Tarantola, A., and B. Valette, 1982, Inverse Problems = Quest for Information: Journal of Geophysics, 50, 159–170.

van Leeuwen, T., and F.J. Herrmann, 2012, Fast waveform inversion without source encoding: Geophysical Prospecting, 61, 10-19.

van den Berg, P.M., A. Abubakar, and J.T. Fokkema, 2003, Multiplicative regularization for contrast profile inversion: Radio Science, 38(2), 8022-31.

Watanabe, T., K.T. Nihei, S. Nakagawa, and L.R. Myer, 2004, Viscoacoustic wave form inversion of transmission data for velocity and intrinsic attenuation: Journal of the Acoustical Society of America, 115(6), 3059-3067.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Štekl, L. Guasch, L. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, 78(2), R59-R80.

Wu, R. S., and M.N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, 52(1), 11–25.

Yao, G., N.V. da Silva, M. Warner, T. Kalinicheva, 2018, Separation of migration and tomography models of full waveform inversion in the plane wave domain: Journal of Geophysical Research: Solid Earth, 123(2), 1486-1501.

Yao, G., N.V. da Silva, and D. Wu, 2018, An effective absorbing layer for the boundary condition in acoustic seismic wave simulation: Journal of Geophysics and Engineering, 495-511.

Yao, G., and D. Wu, 2017, Reflection full waveform inversion: Science China Earth Sciences, 60(10), 1783-1794.

Yao, J., T. Zhu, F. Hussain, and D.J. Kouri, 2017, Locally solving fractional Laplacian viscoacoustic wave equation using Hermite distributed approximating functional method: Geophysics, 82(2), T59-T67.

Zhu, T., and J.M. Carcione, 2014, Theory and modelling of constant-Q P- and S-waves using fractional spatial derivatives: Geophysical Journal International, 196(3), 1787-1795.

Zhu, T. and J.M. Harris, 2015, Improved seismic image by Q-compensated reverse time migration: Application to crosswell field data, west Texas: Geophysics, 80(2), B61-B67.

FIGURES

Figure 1. a) true velocity model, b) inverted model with FWI jointly updating Q (noise-free data), c) model estimated with semi-global inversion (noise-free data), d) model estimated with semi-global inversion (noisy data), e) model estimated with FWI after semi-global inversion (noise-free data), f) model estimated with FWI after semi-global inversion (noisy data), g) model estimated with FWI assuming Q is known (noise-free data), and h) model estimated with FWI assuming Q is known (noisy data).

Figure 2. a) true *Q* model, b) model inverted with FWI jointly updating velocity (noise-free data), c) model inverted with the semi-global inversion (noise-free data), d) model inverted with semi-global inversion (noisy data).

Figure 3. Radiation pattern, $\boldsymbol{\vartheta}$, (equation B24) of a visco-acoustic medium with vertical transverse isotropy in the velocity.

Figure 4. a) Marmousi velocity model, b) model of *Q*, c) starting model of velocity for carrying out inversion, and d) background model of *Q*.

Figure 5. Models of Thomsen's parameters a) epsilon and b) delta.

Figure 6. Shot gather computed a) without intrinsic attenuation, b) with the *Q* model in Figure 4b, and c) with the intrinsic attenuation model in Figure 4d, and d) the difference between the shot gathers in b) and c).

Figure 7. a) Comparison between the trace number 300 for each shot gather in Figures 6a-c, b) comparison between the frequency spectrum of each trace in a, c) comparison between the trace number 700 for each shot gather in Figures 6a-c, and d) comparison between the frequency spectrum of each trace in c). The data in the dashed red line is generated with the model of intrinsic attenuation in Figure 4b (respective to the shot gather in Figure 4b). The data in the black line is generated with the model of intrinsic attenuation in Figure 4d (respective to the shot gather in Figure 6c).

Figure 8. Inverted velocity models: a) using and holding fixed the model of *Q* in Figure 4b and, b) using and holding fixed the model of *Q* in Figure 4d, throughout the iterations.

Figure 9. Inverted velocity models: a) with conventional FWI assuming the true model of *Q* (in Figure 4d) is know, b) with conventional FWI assuming the medium does not have intrinsic attenuation (the white arrows point out where inaccuracies in the inverted model are most visible), c) jointly updating velocity and *Q* with conventional FWI, and not regularizing *Q*, d) jointly updating velocity and *Q* with conventional FWI, and regularizing *Q*, e) after the semi-global inversion, f) carrying out FWI for velocity only, after the semi-global inversion, g) jointly updating velocity and *Q* with FWI, projecting the gradient of *Q* onto the spline basis, h) FWI inversion of noisy data using the true *Q* model, i) FWI of noisy data for velocity only, after the semi-global inversion, and j) true model of velocity (same as depicted in Figure 4a).

Figure 10. Inverted models of *Q*: a) jointly updating velocity and *Q* with conventional FWI and without smoothing regularization of *Q*, b) jointly updating velocity and *Q* with conventional FWI and with smoothing regularization of *Q*, c) inverted *Q* model after 21 semi-global iterations at 3 Hz, d) jointly updating velocity and *Q*, projecting the gradient of *Q* onto the spline basis, e) inverted model with noisy data, and f) background model of *Q* (same as model depicted in Figure 4d). The white dots represent the position of the nodes for B-spline interpolation.

Figure 11. Survey geometry overlaid with the starting P-wave velocity model. The grey dots represent the shot positions and the vertical black lines represent the position of the cables with receivers. The black circle represents the position of a well drilled in the area. The vertical and horizontal dark-blue lines are a selected inline (labeled IL) and a selected cross-line (labeled XL), respectively, for plotting vertical slices of the models. The red star and labeled red circles represent selected shot and receiver positions for displaying and comparing the data.

Figure 12. Starting velocity model for the inversion of the North-Sea dataset along the inline direction. The white, grey and black dots represent the position of the nodes of the sparse basis for estimating Q.

Figure 13. Models for the Thomsen's parameters a) $\boldsymbol{\delta}$ and b) $\boldsymbol{\varepsilon}$ along the inline direction. The models have vertical variation only.

Figure 14. Inverted velocity models along the inline direction: a) utilizing conventional FWI without Q at 3.5 Hz, and after 36 local iterations, b) utilizing conventional FWI without Q, after completing the inversion at 6.5 Hz, c) utilizing semi-global inversion at 3.5 Hz corresponding to 36 local iterations, and d) after completing conventional FWI at 6.5 Hz with Q, starting at 4Hz from the velocity model in c).

Figure 15. Inverted velocity models at 6.5 Hz along the cross-line direction using, a) conventional FWI without Q, and b) the semi-global inversion and FWI.

Figure 16. Inverted model of Q along the inline direction with the sparse-grid nodes overlaid.

Figure 17. Progression of the objective function with iteration number for FWI without Q (grey line), semi-global iterations (black dots), and FWI with Q after semi-global inversion (black line).

Figure 18. Plot of the L2 norm of the data misfit between real data and synthetic data generated with the velocity model resulting from conventional FWI (black bars), the synthetic data generated with Q model and respective velocity model (grey bars).

Figure 19. For all plots: the trace on the left is real recorded data, at the center the traces are synthetically generated with the inverted velocity and Q models resulting from combining the semi-global inversion with FWI, and on the right the traces are synthetically generated with the inverted velocity model resulting from conventional FWI without Q. The labels a), b), c) and d) match the labeled virtual receiver positions in Figure 11.

Figure 20. Comparison between a sonic log recorded in a well in the area, the vertical profile of velocity along the location of the well inverted using the conventional FWI (without accounting for Q; represented by the dotted line), and the vertical profile of velocity at the well location inverted with the semi-global method and FWI (taking into account Q; represented by the solid black line).

TABLES

Table 1. Comparison between the computational cost of FWI and that of the semi-global inversion method; $\boldsymbol{N_L}$ and $\boldsymbol{N'_L}$ are the overall number of local and nested local iterations, respectively.

	FWI	**Semi-global inversion**
Number of local iterations	N_L	N'_L
Number of shots per iteration	N_S	N'_S
Number of particles	-	N_p
Computational cost	$\mathcal{O}(L) = N_L N_S$	$\mathcal{O}(G) = N_p N'_L N'_S$

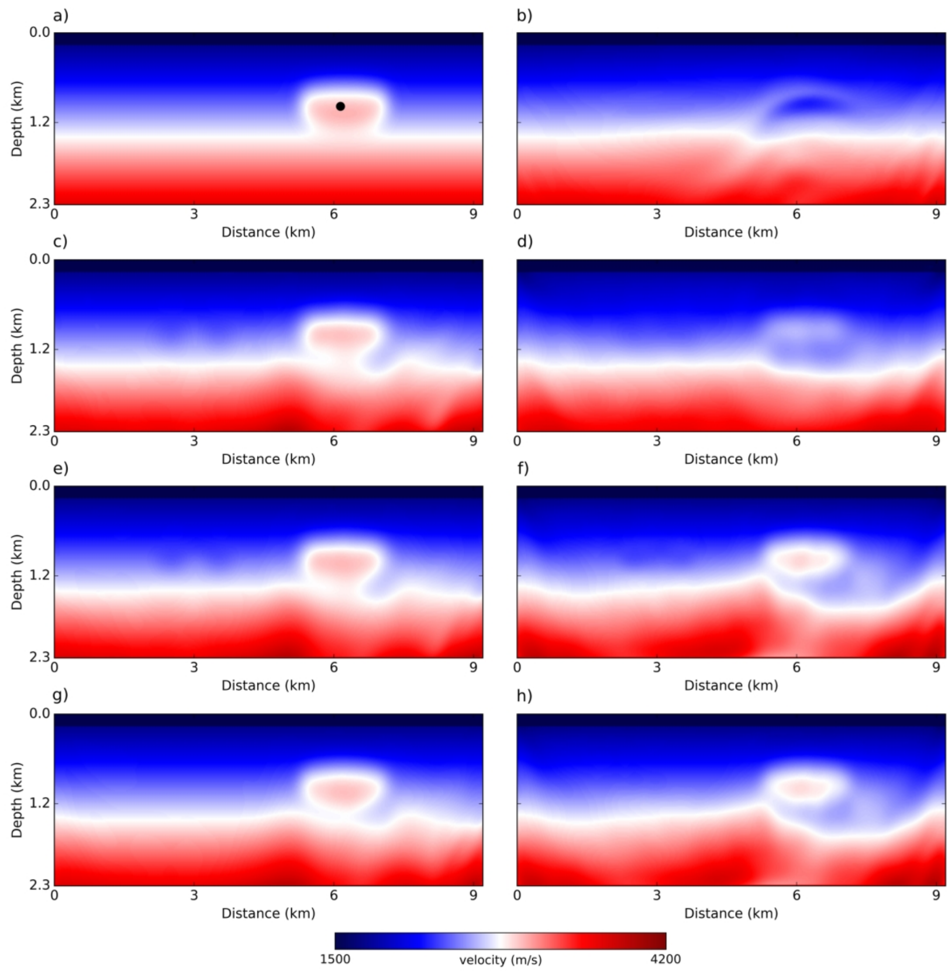

Figure 1. a) true velocity model, b) inverted model with FWI jointly updating Q (noise-free data), c) model estimated with semi-global inversion (noise-free data), d) model estimated with semi-global inversion (noisy data), e) model estimated with FWI after semi-global inversion (noise-free data), f) model estimated with FWI after semi-global inversion (noisy data), g) model estimated with FWI assuming Q is known (noise-free data), and h) model estimated with FWI assuming Q is known (noisy data).

174x178mm (300 x 300 DPI)

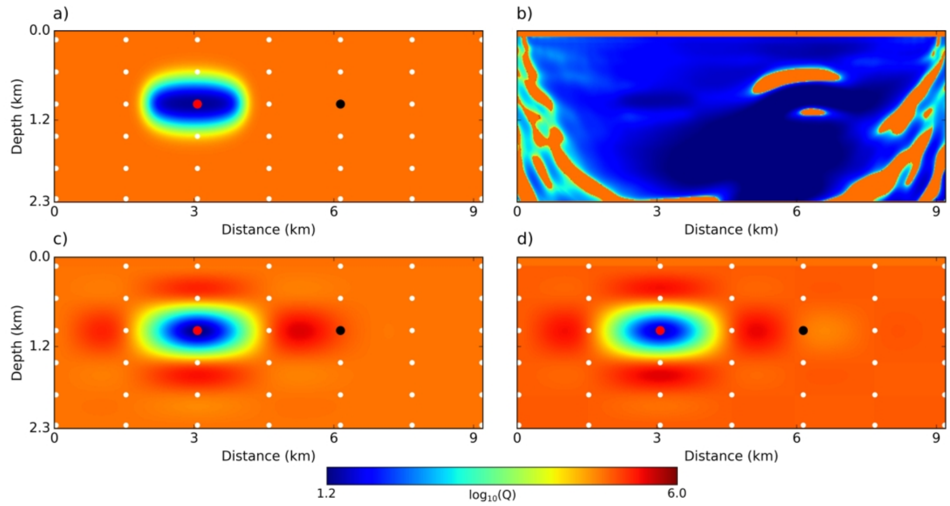

Figure 2. a) true Q model, b) model inverted with FWI jointly updating velocity (noise-free data), c) model inverted with the semi-global inversion (noise-free data), d) model inverted with semi-global inversion (noisy data).

174x94mm (300 x 300 DPI)

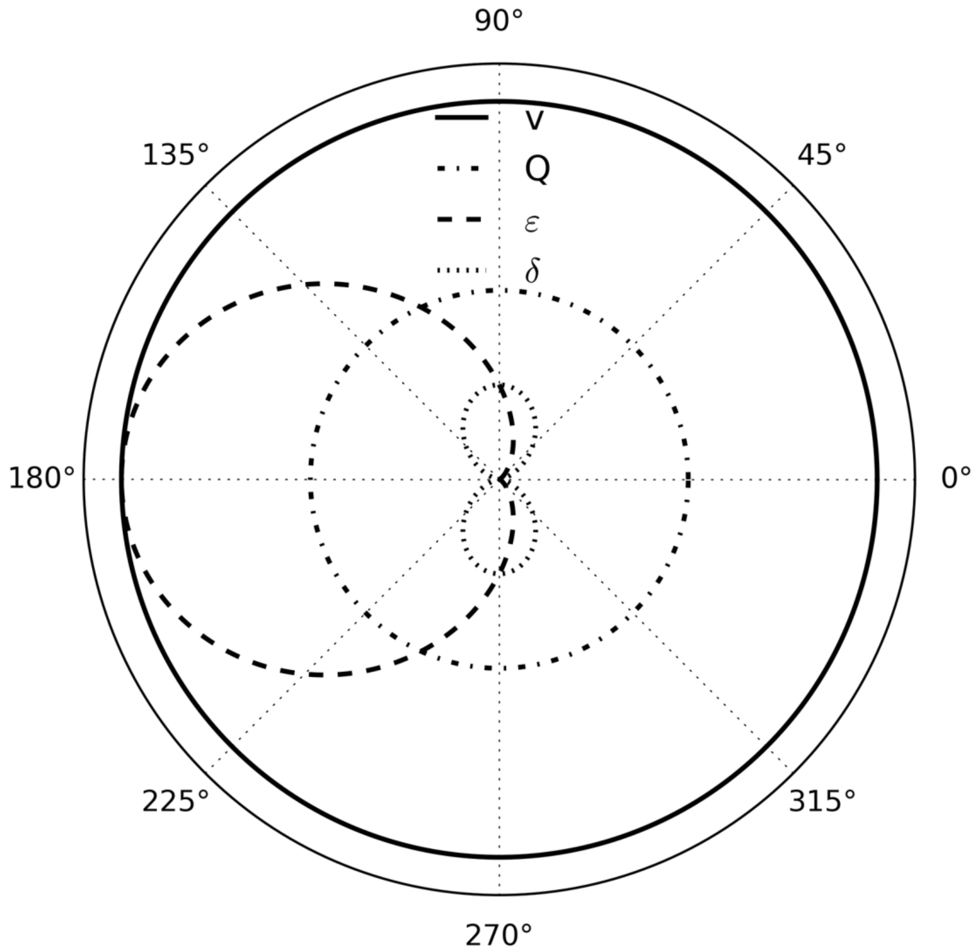

Figure 3. Radiation pattern, ϑ, (equation B24) of a visco-acoustic medium with vertical transverse isotropy in the velocity.

146x143mm (300 x 300 DPI)

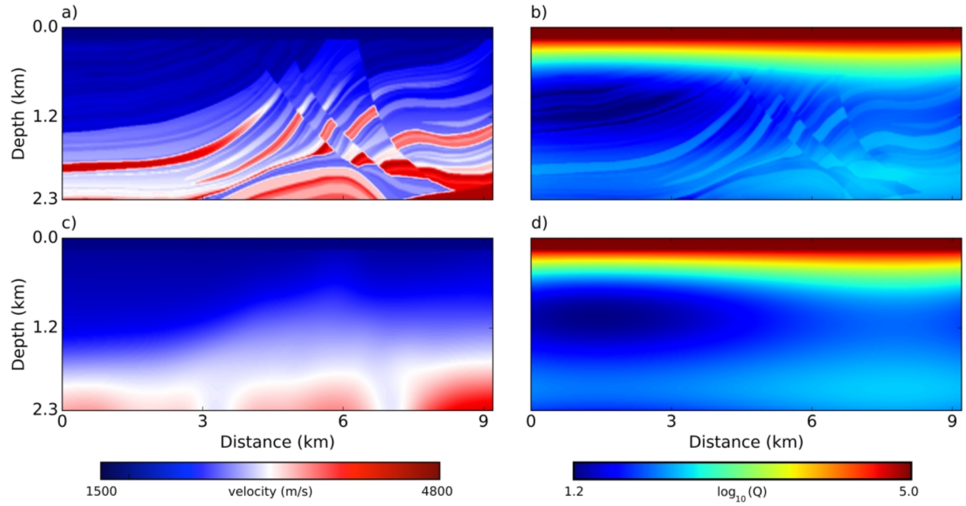

Figure 4. a) Marmousi velocity model, b) model of Q, c) starting model of velocity for carrying out inversion, and d) background model of Q.

170x92mm (300 x 300 DPI)

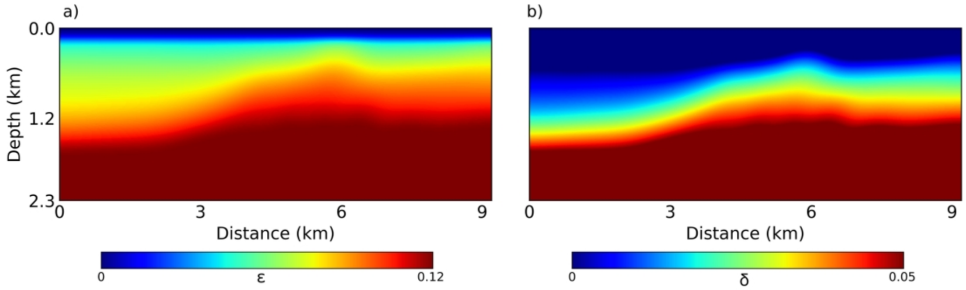

Figure 5. Models of Thomsen's parameters a) epsilon and b) delta.

174x55mm (300 x 300 DPI)

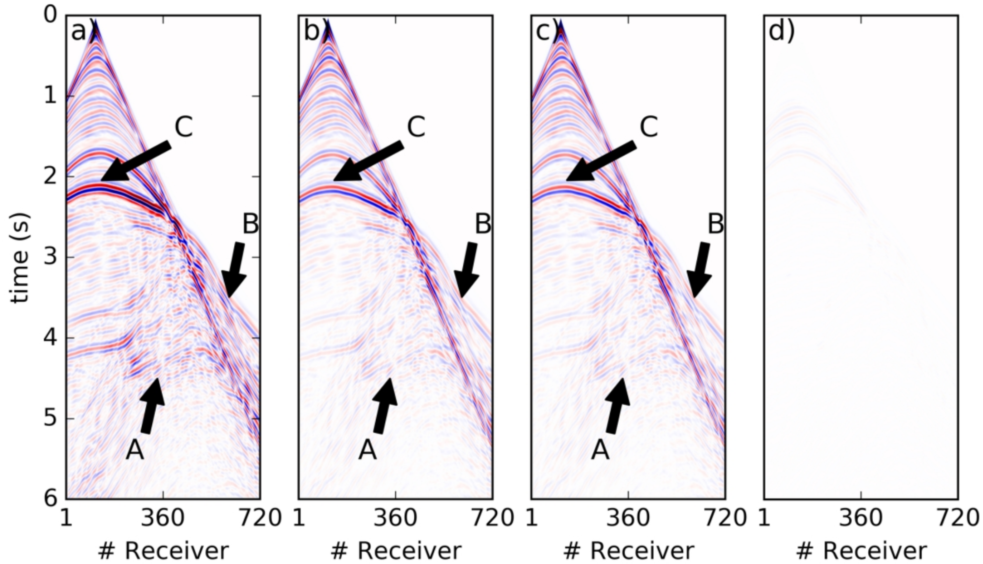

Figure 6. Shot gather computed a) without intrinsic attenuation, b) with the Q model in Figure 4b, and c) with the intrinsic attenuation model in Figure 4d, and d) the difference between the shot gathers in b) and c).

176x104mm (300 x 300 DPI)

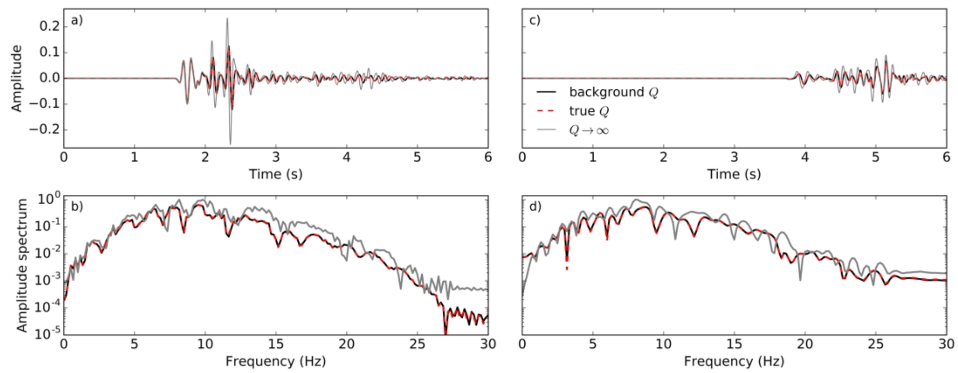

Figure 7. a) Comparison between the trace number 300 for each shot gather in Figures 6a-c, b) comparison between the frequency spectrum of each trace in a, c) comparison between the trace number 700 for each shot gather in Figures 6a-c, and d) comparison between the frequency spectrum of each trace in c). The data in the dashed red line is generated with the model of intrinsic attenuation in Figure 4b (respective to the shot gather in Figure 4b). The data in the black line is generated with the model of intrinsic attenuation in Figure 4d (respective to the shot gather in Figure 6c).

197x82mm (300 x 300 DPI)

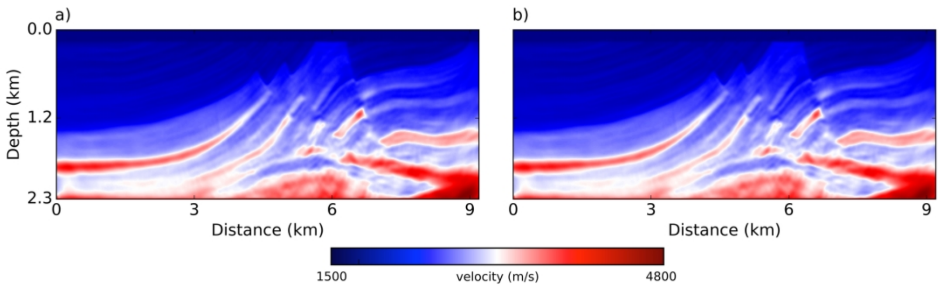

Figure 8. Inverted velocity models: a) using and holding fixed the model of Q in Figure 4b and, b) using and holding fixed the model of Q in Figure 4d, throughout the iterations.

174x55mm (300 x 300 DPI)

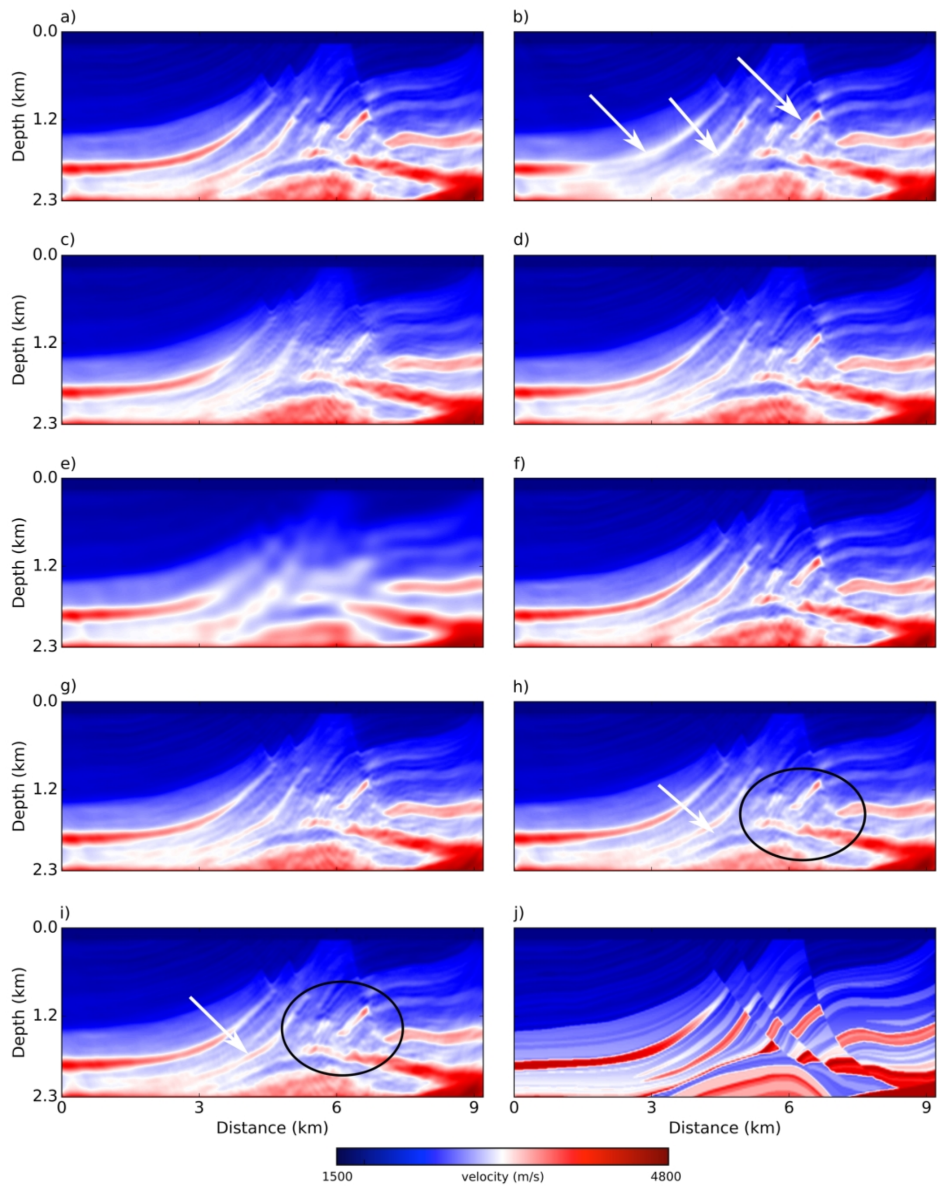

Figure 9. Inverted velocity models: a) with conventional FWI assuming the true model of Q (in Figure 4d) is know, b) with conventional FWI assuming the medium does not have intrinsic attenuation (the white arrows point out where inaccuracies in the inverted model are most visible), c) jointly updating velocity and Q with conventional FWI, and not regularizing Q, d) jointly updating velocity and Q with conventional FWI, and regularizing Q, e) after the semi-global inversion, f) carrying out FWI for velocity only, after the semi-global inversion, g) jointly updating velocity and Q with FWI, projecting the gradient of Q onto the spline basis, h) FWI inversion of noisy data using the true Q model, i) FWI of noisy data for velocity only, after the semi-global inversion, and j) true model of velocity (same as depicted in Figure 4a).

174x209mm (300 x 300 DPI)

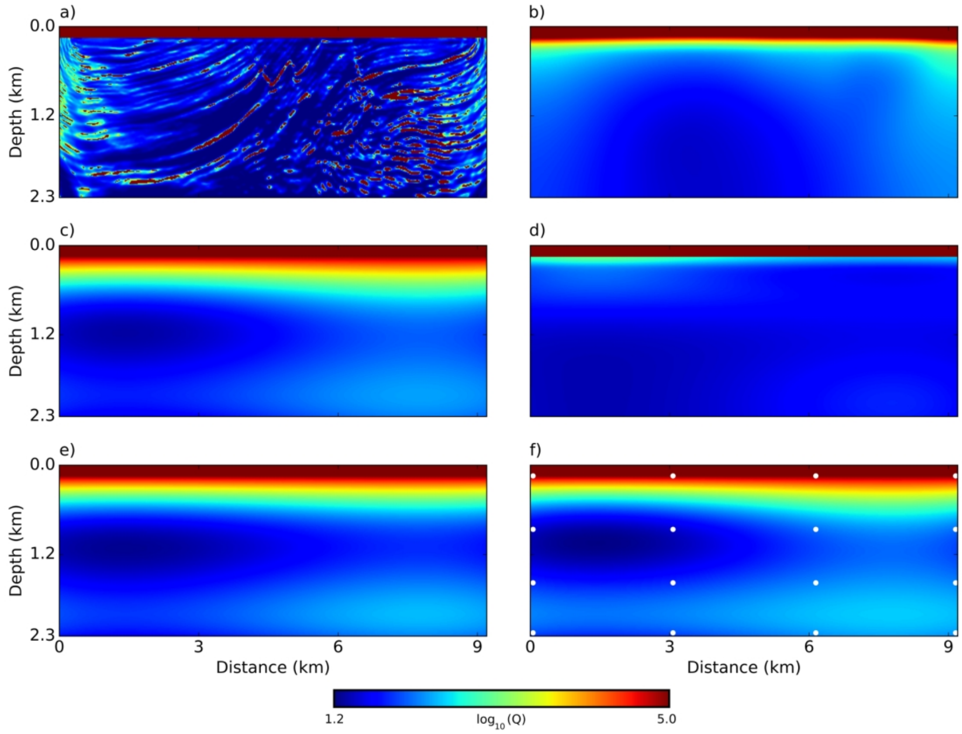

Figure 10. Inverted models of Q: a) jointly updating velocity and Q with conventional FWI and without smoothing regularization of Q, b) jointly updating velocity and Q with conventional FWI and with smoothing regularization of Q, c) inverted Q model after 21 semi-global iterations at 3 Hz, d) jointly updating velocity and Q, projecting the gradient of Q onto the spline basis, e) inverted model with noisy data, and f) background model of Q (same as model depicted in Figure 4d). The white dots represent the position of the nodes for B-spline interpolation.

172x131mm (300 x 300 DPI)

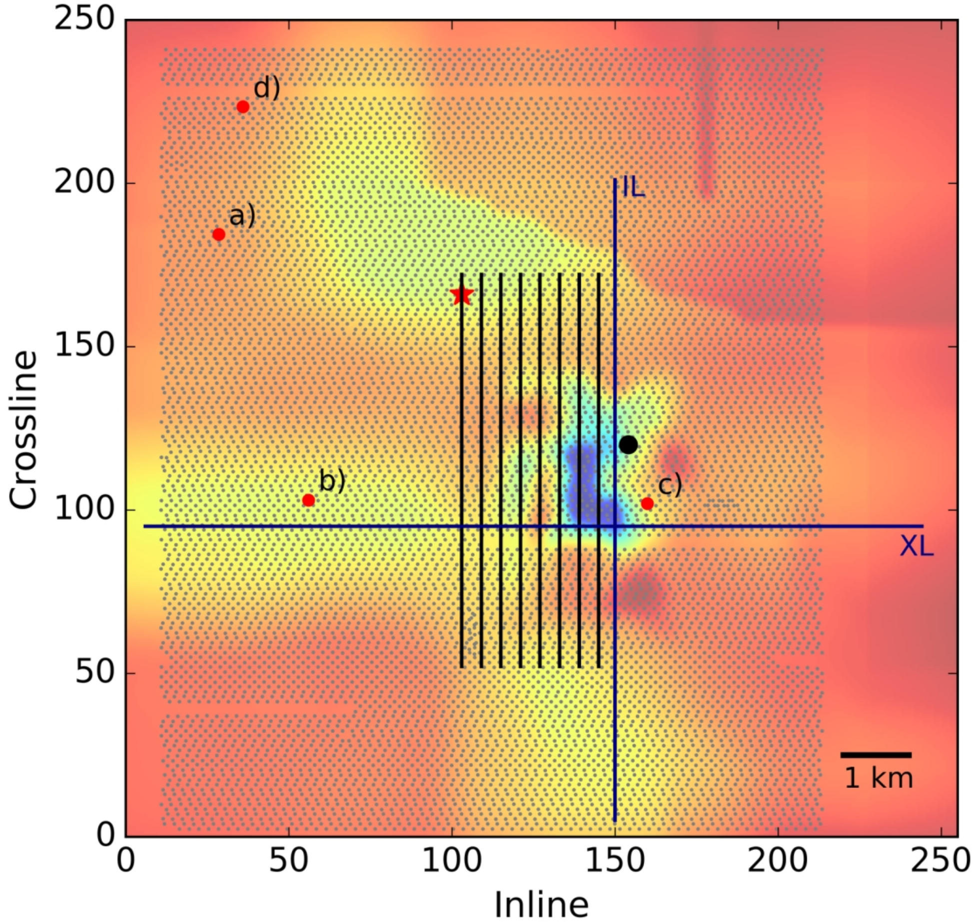

Figure 11. Survey geometry overlaid with the starting P-wave velocity model. The grey dots represent the shot positions and the vertical black lines represent the position of the cables with receivers. The black circle represents the position of a well drilled in the area. The vertical and horizontal dark-blue lines are a selected inline (labeled IL) and a selected cross-line (labeled XL), respectively, for plotting vertical slices of the models. The red star and labeled red circles represent selected shot and receiver positions for displaying and comparing the data.

149x142mm (300 x 300 DPI)

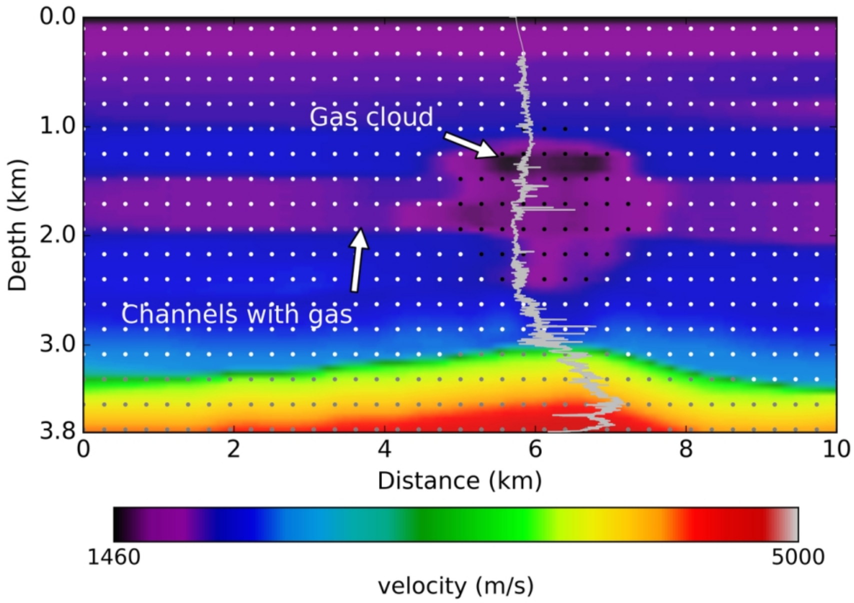

Figure 12. Starting velocity model for the inversion of the North-Sea dataset along the inline direction. The white, grey and black dots represent the position of the nodes of the sparse basis for estimating Q.

129x86mm (300 x 300 DPI)

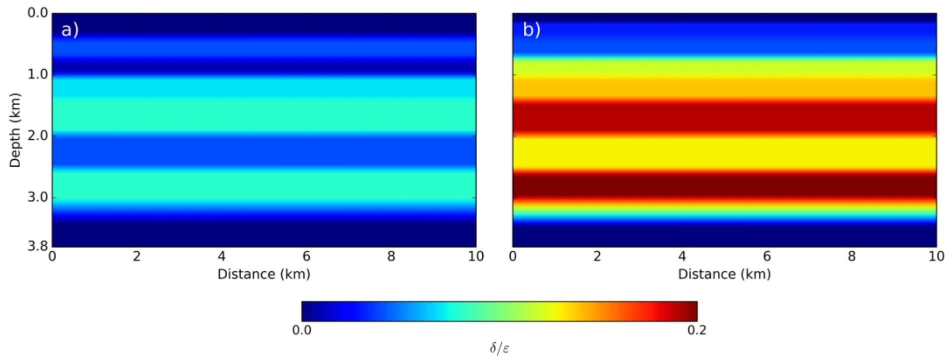

Figure 13. Models for the Thomsen's parameters a) δ and b) ε along the inline direction. The models have vertical variation only.

230x89mm (300 x 300 DPI)

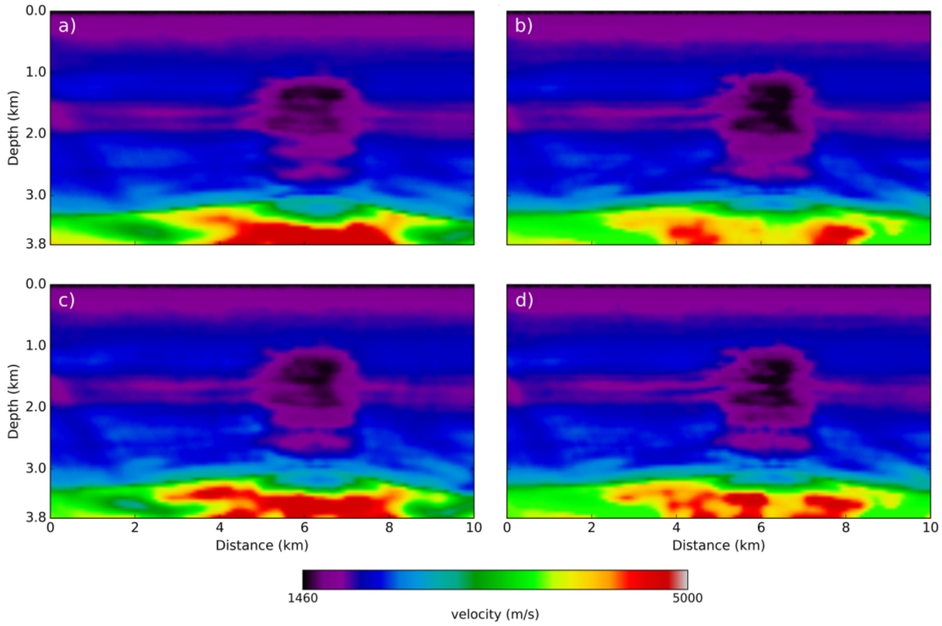

Figure 14. Inverted velocity models along the inline direction: a) utilizing conventional FWI without Q at 3.5 Hz, and after 36 local iterations, b) utilizing conventional FWI without Q, after completing the inversion at 6.5 Hz, c) utilizing semi-global inversion at 3.5 Hz corresponding to 36 local iterations, and d) after completing conventional FWI at 6.5 Hz with Q, starting at 4Hz from the velocity model in c).

230x159mm (300 x 300 DPI)

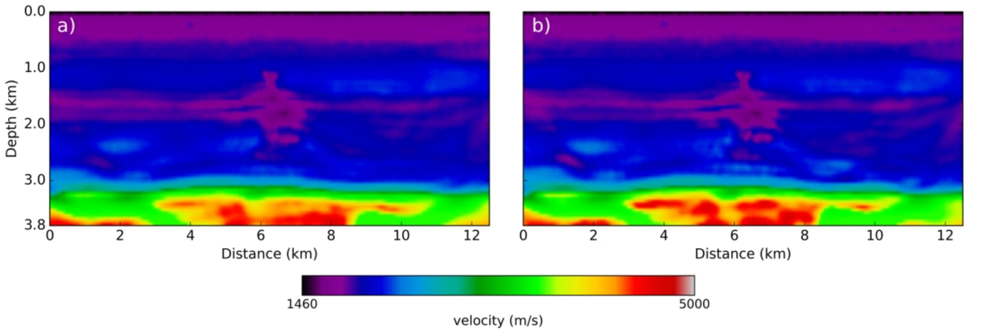

Figure 15. Inverted velocity models at 6.5 Hz along the cross-line direction using, a) conventional FWI without Q, and b) the semi-global inversion and FWI.

225x89mm (300 x 300 DPI)

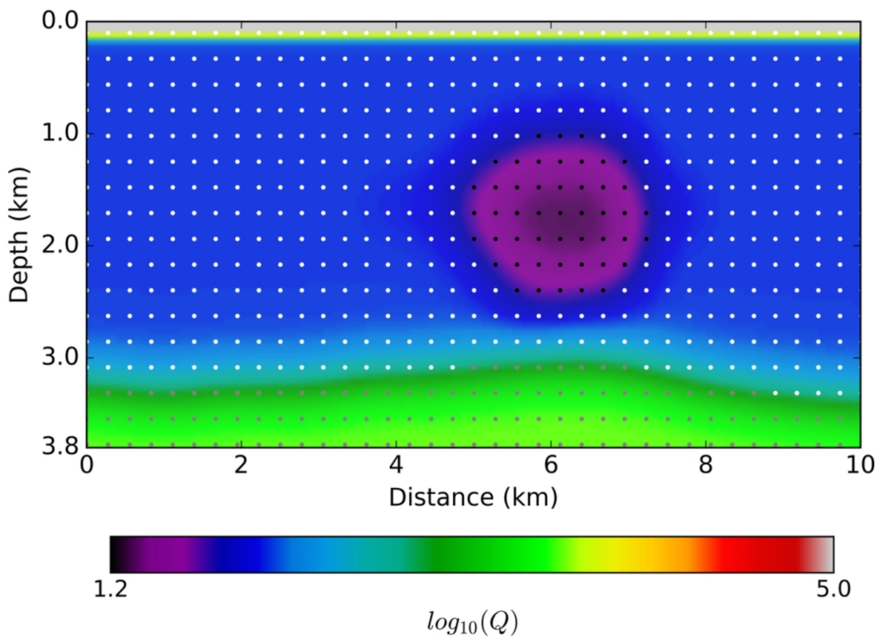

Figure 16. Inverted model of Q along the inline direction with the sparse-grid nodes overlaid.

126x87mm (300 x 300 DPI)

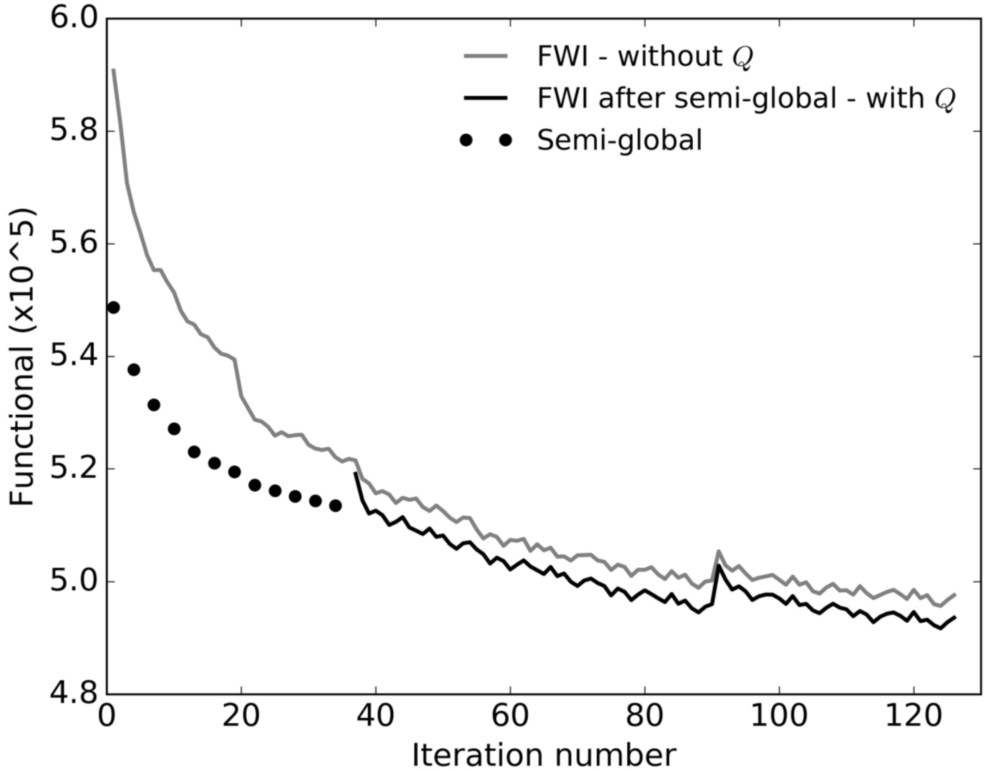

Figure 17. Progression of the objective function with iteration number for FWI without Q (grey line), semi-global iterations (black dots), and FWI with Q after semi-global inversion (black line).

180x143mm (300 x 300 DPI)

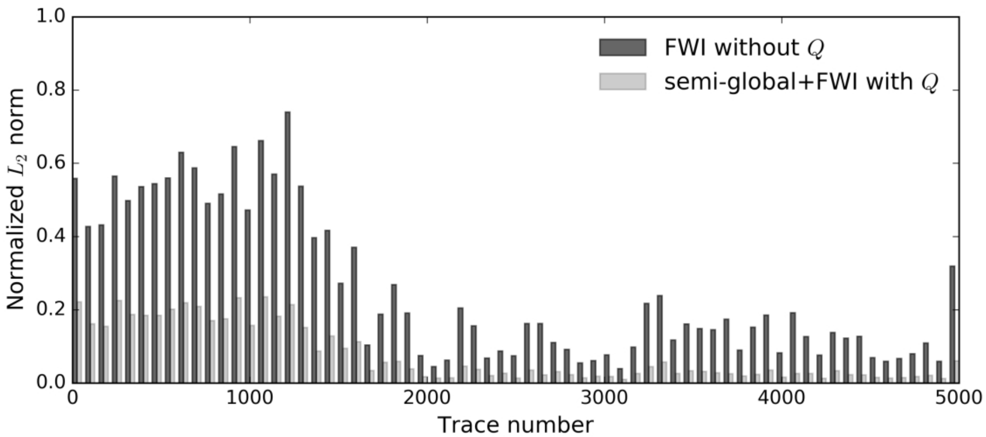

Figure 18. Plot of the L2 norm of the data misfit between real data and synthetic data generated with the velocity model resulting from conventional FWI (black bars), the synthetic data generated with Q model and respective velocity model (grey bars).

221x100mm (300 x 300 DPI)

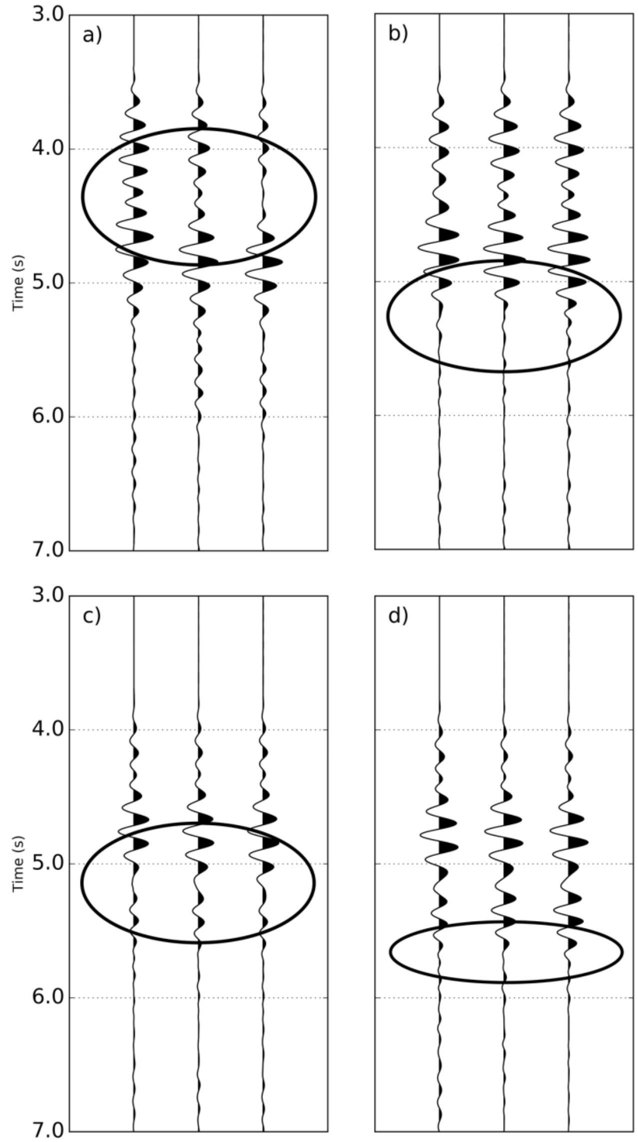

Figure 19. For all plots: the trace on the left is real recorded data, at the center the traces are synthetically generated with the inverted velocity and Q models resulting from combining the semi-global inversion with FWI, and on the right the traces are synthetically generated with the inverted velocity model resulting from conventional FWI without Q. The labels a), b), c) and d) match the labeled virtual receiver positions in Figure 11.

118x203mm (300 x 300 DPI)

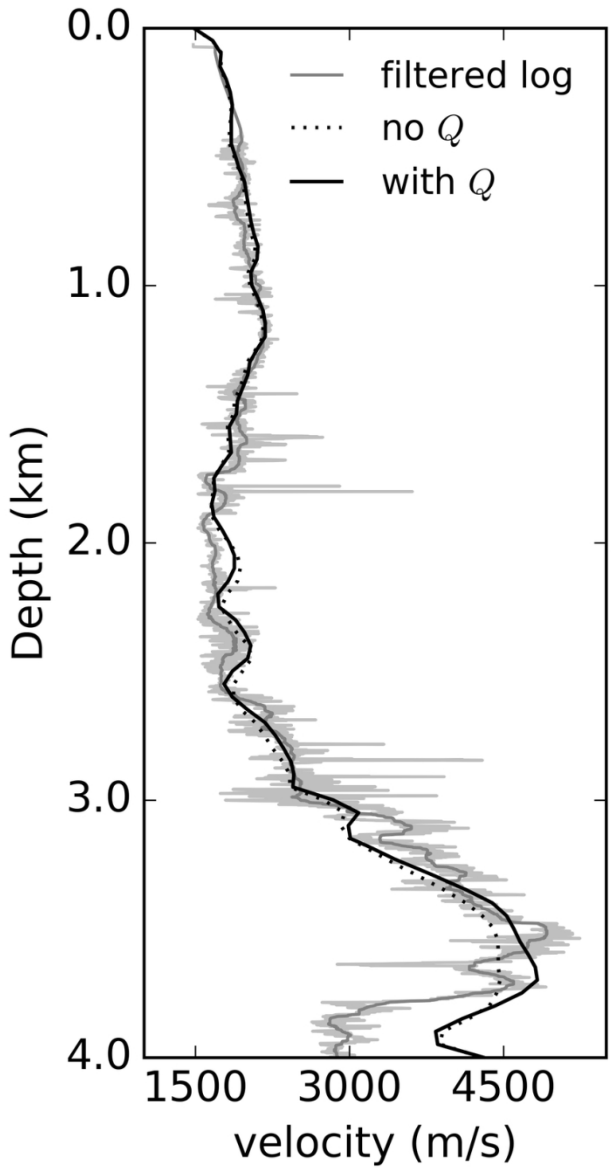

Figure 20. Comparison between a sonic log recorded in a well in the area, the vertical profile of velocity along the location of the well inverted using the conventional FWI (without accounting for Q; represented by the dotted line), and the vertical profile of velocity at the well location inverted with the semi-global method and FWI (taking into account Q; represented by the solid black line).

75x142mm (300 x 300 DPI)

DATA AND MATERIALS AVAILABILITY

Data associated with this research are confidential and cannot be released.