

# Time domain viscoelastic full waveform inversion

Gabriel Fabien-Ouellet, Erwan Gloaguen and Bernard Giroux

INRS-ETE, 490, Rue de la Courome, Québec, QC G1K 9A9, Canada. E-mail: [Gabriel.Fabien-Ouellet@ete.inrs.ca](mailto:Gabriel.Fabien-Ouellet@ete.inrs.ca)

Accepted 2017 March 15. Received 2017 March 10; in original form 2016 August 26

## SUMMARY

Viscous attenuation can have a strong impact on seismic wave propagation, but it is rarely taken into account in full waveform inversion (FWI). When viscoelasticity is considered in time domain FWI, the displacement formulation of the wave equation is usually used instead of the popular velocity–stress formulation. However, inversion schemes rely on the adjoint equations, which are quite different for the velocity–stress formulation than for the displacement formulation. In this paper, we apply the adjoint state method to the isotropic viscoelastic wave equation in the velocity–stress formulation based on the generalized standard linear solid rheology. By applying linear transformations to the wave equation before deriving the adjoint state equations, we obtain two symmetric sets of partial differential equations for the forward and adjoint variables. The resulting sets of equations only differ by a sign change and can be solved by the same numerical implementation. We also investigate the crosstalk between parameter classes (velocity and attenuation) of the viscoelastic equation. More specifically, we show that the attenuation levels can be used to recover the quality factors of  $P$  and  $S$  waves, but that they are very sensitive to velocity errors. Finally, we present a synthetic example of viscoelastic FWI in the context of monitoring  $\text{CO}_2$  geological sequestration. We show that FWI based on our formulation can indeed recover  $P$ - and  $S$ -wave velocities and their attenuation levels when attenuation is high enough. Both changes in velocity and attenuation levels recovered with FWI can be used to track the  $\text{CO}_2$  plume during and after injection. Further studies are required to evaluate the performance of viscoelastic FWI on real data.

**Key words:** Elasticity and anelasticity; Inverse theory; Waveform inversion; Seismic attenuation; Seismic tomography.

## 1 INTRODUCTION

Full waveform inversion (FWI) is a method to recover the Earth parameters based on the solution of the wave equation (Tarantola 1988). In conventional seismic inversion, FWI is mostly used to recover the  $P$ -wave velocity in the acoustic approximation and is less often applied in the more general viscoelastic framework, despite the theory being developed from the very beginning by Tarantola (1988). Viscous attenuation, usually embodied by the  $Q$ -factor, can have a strong impact on seismic wave propagation. As shown in numerous laboratory and in-situ measurements (Barton 2007), the  $Q$ -factor correlates with many useful properties, like rock type and fluid saturation. Thus, it is a property of interest in itself.

A number of different approaches have been proposed for FWI in viscous media, most notably in the frequency domain (Song *et al.* 1995) or in the Laplace–Fourier domain (Kamei & Pratt 2013). The simplicity of modelling viscous attenuation in the frequency domain is one of its main advantages over the time-domain; one only has to define complex velocities to implement an arbitrary attenuation profile in frequency (Toksöz & Johnston 1981). In contrast, the time-domain approach usually requires solving additional

differential equations for memory variables (Carcione *et al.* 1988; Robertsson *et al.* 1994). Obtaining a desired attenuation profile in frequency is not always straightforward (Blanch *et al.* 1995). However, the time-domain approach remains useful for large 3-D models where the memory usage of the frequency approach is prohibitive, or when many frequencies are needed during inversion (Fichtner 2011).

The literature on time-domain FWI in viscous media is much more tenuous than in the frequency domain. Most authors only consider the viscoacoustic case (Liao & McMechan 1995, 1996; Causse *et al.* 1999; Bai *et al.* 2014). When viscoelasticity is considered, like in the theoretical works of Tarantola (1988) or more recently by Charara *et al.* (2000), Fichtner *et al.* (2006) and Askan *et al.* (2007), the displacement formulation of the wave equation is used. However, this formalism is not readily applicable to the velocity–stress scheme popularized by Virieux (1986) and Levander (1988), because the adjoint state equations for the velocity–stress and the displacement formulations are different (Castellanos *et al.* 2011).

The adjoint state method (see Plessix 2006 for a review) allows the computation of the misfit function derivative without having to compute the Fréchet derivative, reducing the computing burden

from one forward modelling per inversion parameter, to only two regardless of the number of parameters. It is at the heart of FWI and makes it feasible in practice. In all cases, both in time and in frequency, the adjoint state method for FWI implies the correlation of the forward propagated field and the back-propagated residuals. The details of this correlation are, however, implementation specific. Hence, the results obtained for the displacement formulation cannot be used directly with the viscoelastic equations of Robertsson *et al.* (1994) and the like. A recent study by Yang *et al.* (2016) derived the forward and adjoint equations for this formulation, which are asymmetric because the first order viscoelastic wave equation is not self-adjoint. Hence, two different systems of partial differential equations must be solved for the forward and adjoint variables.

In this paper, we present an alternative strategy to compute the gradient by the adjoint state method for the viscoelastic wave equations in the velocity–stress formulation. The purpose is to be able to recover the  $P$ -wave and  $S$ -wave velocities and the quality factor with the same modelling code to compute both the forward and adjoint wavefields. This work is divided in two main sections. First, the theory is presented. Following the methodology proposed by Fabien-Ouellet *et al.* (2016), the adjoint state method is applied to a transformed wave equation, leading to symmetric forward and adjoint sets of equations. Second, the theory is applied to a synthetic experiment, where FWI is used for monitoring  $\text{CO}_2$  injection by cross-well tomography. This experiment aims to test the feasibility of recovering both the velocities and  $Q$  factors in a reasonably realistic case.

Throughout this paper, Einstein notation is used, where summation of variables with repeated indices is implied. Upper-case bold symbols designate matrices, lower-case bold symbols designate vectors, Latin indices are used for spatial dimensions and Greek indices are used for the elements of matrices and vectors.

## 2 THEORY

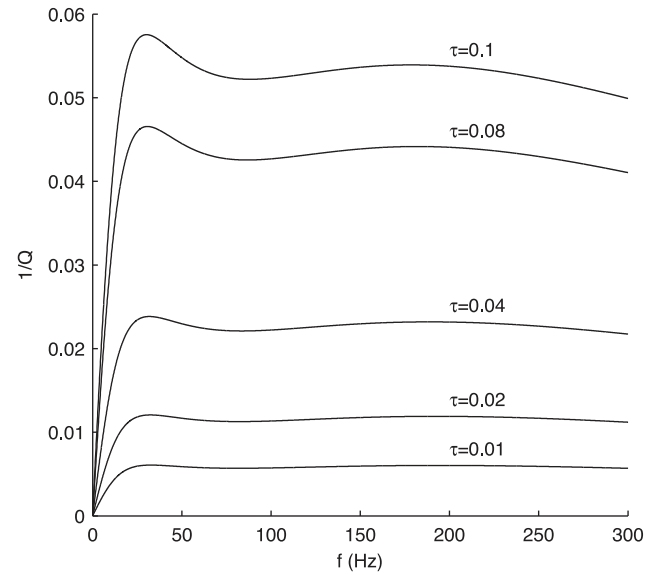
In this section, we propose a strategy to compute the misfit gradient by the adjoint state method leading to the same set of equations for the forward and adjoint variables. The section is organized as follows. First, the isotropic viscoelastic wave equation in the velocity–stress formulation is presented. Then, we perform a change of variable and apply linear transformations to this system to obtain an intermediate forward model that only involves self-adjoint and anti-self-adjoint operators. The adjoint state equations and the gradient of the misfit function are derived in this transformed system. Finally, performing the back transformations leads to our final results: a set of equations for the adjoint variable that is identical to the forward equations and the explicit expression for the misfit gradient.

### 2.1 Forward modelling

Full waveform inversion is based on the wave equation, which we also refer to as the forward model. The wave equation is taken here as the isotropic viscoelastic wave equation based on the generalized standard linear solid as described by Carcione *et al.* (1988) and Robertsson *et al.* (1994):

$$\partial_t v_i - \frac{1}{\rho} \partial_j \sigma_{ij} - s_i = 0, \quad (1a)$$

$$\partial_t \sigma_{ij} - \lambda_0 \partial_k v_k \delta_{ij} - \mu_0 (\partial_j v_i + \partial_i v_j) - r_{ijl} \delta_l - s_{ij} = 0 \quad (1b)$$



**Figure 1.** Attenuation factor in function of frequency for different attenuation levels for two Maxwell bodies ( $\tau_{\sigma l} = \frac{1}{160}, \frac{1}{1600}$  s). This shows how  $\tau$  controls the amount of attenuation without changing too much the shape of the profile.

$$\partial_i r_{ijl} + \frac{1}{\tau_{\sigma l}} [\lambda_l \partial_k v_k \delta_{ij} + \mu_l (\partial_j v_i + \partial_i v_j) + r_{ijl}] = 0, \quad (1c)$$

where  $s$  is the source term,  $v$  is the particle velocity,  $\sigma$  is the stress and  $r$  is the memory variable that implements the  $L$  Maxwell bodies, each having its own relaxation time  $\tau_{\sigma l}$ . We use the formulation corrected for the phase velocity similar to Bohlen (2002), where the Lamé parameters are corrected to obtain the proper  $P$ -wave and  $S$ -wave velocities at the reference frequency  $\omega_0$ :

$$\lambda_l = \begin{cases} M \frac{1+\tau_p}{1+\alpha\tau_p} - 2\mu \frac{1+\tau_s}{1+\alpha\tau_s} & \text{if } l = 0, \\ M \frac{\tau_p}{1+\alpha\tau_p} - 2\mu \frac{\tau_s}{1+\alpha\tau_s} & \text{if } l > 0, \end{cases} \quad (2a)$$

$$\mu_l = \begin{cases} \mu \frac{1+\tau_s}{1+\alpha\tau_s} & \text{if } l = 0, \\ \mu \frac{\tau_s}{1+\alpha\tau_s} & \text{if } l > 0, \end{cases} \quad (2b)$$

$$\alpha = \sum_{l=1}^L \frac{\omega_0^2 \tau_{\sigma l}^2}{1 + \omega_0^2 \tau_{\sigma l}^2}. \quad (2c)$$

where  $M$  is the  $P$ -wave modulus,  $\mu$  is the shear wave modulus,  $\tau_p$  and  $\tau_s$  are the  $P$ -wave and  $S$ -wave attenuation levels. In the isotropic assumption, these are the only parameters required to completely define the stiffness and quality factor tensors. The  $P$ - or  $S$ -wave quality factor is given by

$$Q = \frac{1 + \sum_{l=1}^L \frac{\omega^2 \tau_{\sigma l}^2}{1 + \omega^2 \tau_{\sigma l}^2} \tau}{\sum_{l=1}^L \frac{\omega \tau_{\sigma l}}{1 + \omega^2 \tau_{\sigma l}^2} \tau}. \quad (3)$$

In this relation, the relaxation times  $\tau_{\sigma l}$  control the shape of the  $Q$ -profile whereas the attenuation level  $\tau$  controls the strength of the attenuation. This is illustrated in Fig. 1, where the  $Q$  profile is plotted for different attenuation levels, keeping the relaxation times constant. An arbitrary attenuation frequency profile can be approximated by least-squares fitting as described by (Blanch *et al.* 1995; Emmerich & Korn 1987).

The viscoelastic wave equation as expressed in eq. (1) is not self-adjoint. Therefore, applying the adjoint state method to it would lead to different set of equations for the forward and adjoint wave-fields (see Yang *et al.* 2016). We instead transform eq. (1) by using the integration in time of the memory variables  $\partial_t R_{ijl} = r_{ijl}$  and by transferring the modified memory variables from eq. (1b) to eq. (1a). The resulting set of equations is:

$$\partial_t v_i - \frac{1}{\rho} \partial_j \sigma'_{ij} - \frac{1}{\rho} \partial_j R_{ijl} \delta_l = 0, \quad (4a)$$

$$\partial_t \sigma'_{ij} - \lambda_0 \partial_k v_k \delta_{ij} - \mu_0 (\partial_j v_i + \partial_i v_j) = 0 \quad (4b)$$

$$-\tau_{\sigma l} \partial_{tt} R_{ijl} - \lambda_l \partial_k v_k \delta_{ij} - \mu_l (\partial_j v_i + \partial_i v_j) - \partial_t R_{ijl} = 0, \quad (4c)$$

with  $\sigma'_{ij} = \sigma_{ij} - R_{ijl} \delta_l$ . Eqs (1) and (4) are equivalent and share the same solution. However, the latter can more easily be made symmetric using linear transformations. It is also more convenient for the remainder of this development to use matrix notation. We first define the state vector:

$$\begin{aligned} \boldsymbol{\phi} = & (v_x, v_y, v_z, \sigma'_{xx}, \sigma'_{yy}, \sigma'_{zz}, \sigma'_{xy}, \sigma'_{xz}, \sigma'_{yz}, \\ & R_{xx1}, R_{yy1}, R_{zz1}, R_{xy1}, R_{xz1}, R_{yz1}, \dots, \\ & R_{xxL}, R_{yyL}, R_{zzL}, R_{xyL}, R_{xzL}, R_{yzL})^T. \end{aligned} \quad (5)$$

We can then express eq. (4) in matrix notation:

$$F(\boldsymbol{\phi}; \mathbf{m}) = \mathbf{A}\boldsymbol{\phi} + \mathbf{B}\boldsymbol{\phi} - \mathbf{C}\boldsymbol{\phi} - \mathbf{s} = 0, \quad (6)$$

where  $F(\boldsymbol{\phi}; \mathbf{m})$  is the forward operator that gives the seismic wave-field given the model parameters  $\mathbf{m}$ , which can be any combination of the isotropic viscoelastic variables, for example  $\mathbf{m} = (\rho, M, \mu, \tau_p, \tau_s)^T$ . The operators in eq. (6) are given by

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \mathbf{0}_9 & & \\ & -\mathbf{I}_{6L} & \\ & & \tau_{\sigma l} \partial_{tt} \end{bmatrix}, \\ \mathbf{B} &= \begin{bmatrix} \mathbf{I}_9 & & \\ & -\mathbf{I}_{6L} & \\ & & \partial_t \end{bmatrix}, \\ \mathbf{C} &= \begin{bmatrix} \mathbf{0}_3 & \frac{1}{\rho} \mathbf{D}^T & \dots & \frac{1}{\rho} \mathbf{D}^T \\ \Psi_0 \mathbf{D} & & & \\ \vdots & & \mathbf{0}_{6+6L} & \\ \Psi_L \mathbf{D} & & & \end{bmatrix}, \end{aligned} \quad (7)$$

with the derivative matrix:

$$\mathbf{D}^T = \begin{bmatrix} \partial_x & 0 & 0 & \partial_y & \partial_x & 0 \\ 0 & \partial_y & 0 & \partial_z & 0 & \partial_x \\ 0 & 0 & \partial_z & 0 & \partial_z & \partial_y \end{bmatrix}, \quad (8)$$

and the stiffness matrices:

$$\Psi_l = \begin{bmatrix} \lambda_l + 2\mu_l & \lambda_l & \lambda_l & 0 & 0 & 0 \\ \lambda_l & \lambda_l + 2\mu_l & \lambda_l & 0 & 0 & 0 \\ \lambda_l & \lambda_l & \lambda_l + 2\mu_l & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu_l & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu_l & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu_l \end{bmatrix}, \quad (9)$$

The symbols  $\mathbf{0}_n$  and  $\mathbf{I}_n$  are respectively the  $n \times n$  zero and identity matrices. Operators  $\mathbf{A}$  and  $\mathbf{B}$  involve derivatives with respect to time and operator  $\mathbf{C}$  involves derivatives with respect to space.

## 2.2 Transformation to self-adjoint operators

The goal of full waveform inversion is to estimate the viscoelastic parameters of the ground  $\mathbf{m} = (\rho, M, \mu, \tau_p, \tau_s)$  based on some records of the ground motion  $\mathbf{d}_i$ , usually in the form of particle velocities or pressure. This is performed by the minimization of a cost function, usually taken as the least-squares misfit of the raw seismic traces:

$$J(\boldsymbol{\phi}; \mathbf{m}) = \frac{1}{2} (\mathcal{S}(\boldsymbol{\phi}_\alpha) - \mathbf{d}_\alpha)^T \mathbf{W} (\mathcal{S}(\boldsymbol{\phi}_\alpha) - \mathbf{d}_\alpha), \quad (10)$$

where  $\mathcal{S}$  is the sampling operator that samples the continuous wave-field at discrete times for each source and receiver location and  $\mathbf{W}$  is the data weighting matrix.

As the name implies, the adjoint state method requires an adjoint state, defined by the adjoint state vector:

$$\begin{aligned} \boldsymbol{\psi} = & (\overleftarrow{v}_x, \overleftarrow{v}_y, \overleftarrow{v}_z, \overleftarrow{\sigma}'_{xx}, \overleftarrow{\sigma}'_{yy}, \overleftarrow{\sigma}'_{zz}, \overleftarrow{\sigma}'_{xy}, \overleftarrow{\sigma}'_{xz}, \overleftarrow{\sigma}'_{yz}, \\ & \overleftarrow{R}_{xx1}, \overleftarrow{R}_{yy1}, \overleftarrow{R}_{zz1}, \overleftarrow{R}_{xy1}, \overleftarrow{R}_{xz1}, \overleftarrow{R}_{yz1}, \dots, \\ & \overleftarrow{R}_{xxL}, \overleftarrow{R}_{yyL}, \overleftarrow{R}_{zzL}, \overleftarrow{R}_{xyL}, \overleftarrow{R}_{xzL}, \overleftarrow{R}_{yzL})^T. \end{aligned} \quad (11)$$

The overhead arrow designates the adjoint of a seismic variable. Both the state vector and the adjoint state vector are defined over real spaces  $\mathbf{U}$  and  $\mathbf{U}^*$ , with the following scalar product:

$$\langle \mathbf{a}(\mathbf{x}, t), \mathbf{b}(\mathbf{x}, t) \rangle_{\mathbf{U}} = \int_T \int_X a_\alpha(\mathbf{x}, t) b_\alpha(\mathbf{x}, t) d\mathbf{x} dt. \quad (12)$$

The scalar product contains two distinct parts: a summation over the  $9 + 6L$  discrete dimensions of the seismic variables and integration over the continuous time and space dimensions. In this context, the adjoint  $\mathbf{A}^*$  of a linear operator  $\mathbf{A}$  is defined as:

$$\langle \mathbf{A}\mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{A}^*\mathbf{b} \rangle. \quad (13)$$

An operator is said to be self-adjoint when  $\mathbf{A} = \mathbf{A}^*$ . From the definition of the scalar product, this implies two different conditions: over the discrete dimensions, the operator must be symmetric  $A_{\alpha\beta} = A_{\beta\alpha}$  and over the time and space dimensions, the application of the operator must commute with respect to derivation. Self-adjoint operators are of particular interest because the derivation of the adjoint state equations become trivial when they are involved and because they give rise to symmetric forward and adjoint equations. Hence, a logical approach is to first transform the forward equations into a form involving only self-adjoint operators.

A first step to transform the operators  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  into self-adjoint operators is to apply a linear transformation to symmetrize them. For that purpose, we use transform matrices similar to those employed by Castellanos *et al.* (2011) for the elastic wave equation and explicitly defined Fabien-Ouellet *et al.* (2016) in the viscoelastic case. Given the transform matrices:

$$\mathbf{T} = \begin{bmatrix} \mathbf{I}_3 & & & \\ & \mathbf{R} & & \\ & & \ddots & \\ & & & \mathbf{R} \end{bmatrix}, \quad (14a)$$

$$\boldsymbol{\Lambda} = \begin{bmatrix} \rho \mathbf{I}_3 & & & \\ & \boldsymbol{\Lambda}_0 & & \\ & & \ddots & \\ & & & \boldsymbol{\Lambda}_L \end{bmatrix}, \quad (14b)$$

with the eigenvectors and eigenvalues of the stiffness tensor:

$$\mathbf{R} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{2}(\sqrt{3}+1) & \frac{1}{2}(\sqrt{3}-1) & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{2}(\sqrt{3}-1) & -\frac{1}{2}(\sqrt{3}+1) & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (15a)$$

$$\mathbf{\Lambda}_l = \begin{bmatrix} \frac{1}{3\lambda_l+2\mu_l} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2\mu_l} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2\mu_l} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\mu_l} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\mu_l} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\mu_l} \end{bmatrix}, \quad (15b)$$

symmetric operators are obtained by performing a change of variables  $\phi' = \mathbf{T}\phi$  and multiplying eq. (6) by  $\mathbf{\Lambda}T$ , giving:

$$F'(\phi; \mathbf{m}) = \mathbf{A}'\phi' + \mathbf{B}'\phi' - \mathbf{C}'\phi' - s' = 0. \quad (16)$$

The transformed operators are now symmetric:

$$\begin{aligned} \mathbf{A}' &= \mathbf{\Lambda}T\mathbf{A}T = - \begin{bmatrix} \mathbf{0}_9 & & & & & \\ & \mathbf{\Lambda}_1\tau_{\sigma 1} & & & & \\ & & \ddots & & & \\ & & & & & \\ & & & & & \mathbf{\Lambda}_L\tau_{\sigma L} \end{bmatrix} \partial_{tt}, \\ \mathbf{B}' &= \mathbf{\Lambda}T\mathbf{B}T = \begin{bmatrix} \rho\mathbf{I}_3 & & & & & \\ & \mathbf{\Lambda}_0 & & & & \\ & & -\mathbf{\Lambda}_1\tau_{\sigma 1} & & & \\ & & & \ddots & & \\ & & & & & -\mathbf{\Lambda}_L\tau_{\sigma L} \end{bmatrix} \partial_t, \\ \mathbf{C}' &= \mathbf{\Lambda}T\mathbf{C}T = \begin{bmatrix} \mathbf{0}_3 & \mathbf{D}^T\mathbf{R} & \dots & \mathbf{D}^T\mathbf{R} \\ \mathbf{R}\mathbf{D} & & & \\ \vdots & & \mathbf{0}_{6+6L} & \\ \mathbf{R}\mathbf{D} & & & \end{bmatrix}. \end{aligned} \quad (17)$$

Let's mention that the same transformation can be easily performed for the 2-D case. Note that  $\Psi_l = \mathbf{R}\mathbf{\Lambda}^{-1}\mathbf{R}$ . Hence, our transformation, which was first presented by Fabien-Ouellet *et al.* (2016), is equivalent the approach of Yang *et al.* (2016), that is, we multiply the wave equation by the inverse of the stiffness tensor, or the compliance.

The second criterion for the operator to be self-adjoint is commutation of the partial derivative in space and time. To show this commutation, we use the high-dimensional derivation by part formula

$$\int_{\Omega} \partial_i ab \, d\Omega = \int_{\Gamma} ab \hat{n}_i \, d\Gamma - \int_{\Omega} a \partial_i b \, d\Omega, \quad (18)$$

where  $\Omega$  is the space  $\mathbb{R}^4$  containing the time and space dimensions,  $\Gamma$  is the surface containing the boundary in space and time and  $\hat{n}_i$  is the  $i$ th component of the unit vector normal to  $\Gamma$ . Hence, for partial derivatives to commute, the integration along the boundary must

vanish. This is achieved by using suitable boundary conditions, for example:

- (i)  $\phi|_{t=0} = \partial_t \phi|_{t=0} = 0$  (quiescent past),
- (ii)  $\psi|_{t=T} = \partial_t \psi|_{t=T} = 0$  (quiescent future),
- (iii)  $\phi|_{x \rightarrow \infty} = 0$  (vanishing boundaries).

With those conditions, the transformed operators have the following properties:

$$\begin{aligned} \langle \mathbf{A}'\psi', \phi' \rangle &= \langle \psi', \mathbf{A}'\phi' \rangle, \\ -\langle \mathbf{B}'\psi', \phi' \rangle &= \langle \psi', \mathbf{B}'\phi' \rangle, \\ -\langle \mathbf{C}'\psi', \phi' \rangle &= \langle \psi', \mathbf{C}'\phi' \rangle, \end{aligned} \quad (19)$$

that is,  $\mathbf{A}'$  is self-adjoint and  $\mathbf{B}'$  and  $\mathbf{C}'$  are anti-self-adjoint.

### 2.3 Adjoint state equations

The derivation of the adjoint state equations is trivial once self-adjointness of the operators has been proven. First, we define the Lagrangian of the modified system as:

$$\begin{aligned} \mathcal{L}(\tilde{\phi}', \tilde{\psi}'; \mathbf{m}) &= J(\tilde{\phi}'; \mathbf{m}) \\ &\quad - \langle \tilde{\psi}', \mathbf{A}'\tilde{\phi}' + \mathbf{B}'\tilde{\phi}' - \mathbf{C}'\tilde{\phi}' - s' \rangle. \end{aligned} \quad (20)$$

where  $\tilde{\phi}'$  is any realization (physical or not) of the state vector and  $\tilde{\psi}'$  is any realization of the adjoint state. Using the (anti) self-adjointness of the transformed operators, we can write:

$$\begin{aligned} \mathcal{L}(\tilde{\phi}', \tilde{\psi}'; \mathbf{m}) &= J(\tilde{\phi}'; \mathbf{m}) \\ &\quad - \langle \mathbf{A}'\tilde{\psi}' - \mathbf{B}'\tilde{\psi}' + \mathbf{C}'\tilde{\psi}', \tilde{\phi}' \rangle + \langle \tilde{\psi}', s' \rangle. \end{aligned} \quad (21)$$

Equating to zero the derivative of eq. (21) with respect to  $\tilde{\phi}'$  gives the adjoint state equations for the modified forward model:

$$\frac{\partial J}{\partial \phi'} - \mathbf{A}'\psi' + \mathbf{B}'\psi' - \mathbf{C}'\psi' = 0. \quad (22)$$

Performing the back transformation, we obtain the adjoint state equation for the original forward model:

$$\overleftarrow{F}(\psi, \phi; \mathbf{m}) = \mathbf{A}\psi - \mathbf{B}\psi + \mathbf{C}\psi - \mathbf{T}\mathbf{\Lambda}^{-1}\mathbf{T} \frac{\partial J}{\partial \phi} = 0. \quad (23)$$

This last equation is similar to the usual back propagation equation obtained by Tarantola (1988) for the displacement formulation. It also has the same form as the original forward model. To see this more clearly, we can return to the standard notation, with the traditional memory variables. Making the change of variable  $t' = T - t$ , we obtain

$$\partial_t \bar{v}_i + \frac{1}{\rho} \partial_j \bar{\sigma}_{ij} - \frac{1}{\rho} \frac{\partial J}{\partial v_i} = 0, \quad (24a)$$

$$\begin{aligned} \partial_t \bar{\sigma}_{ij} + \lambda_0 \partial_k \bar{v}_k \delta_{ij} + \mu_0 (\partial_j \bar{v}_i + \partial_i \bar{v}_j) + \bar{r}_{ijl} \delta_l \\ - \lambda_0 \frac{\partial J}{\partial \sigma_{kk}} \delta_{ij} - (1 + \delta_{ij}) \mu_0 \frac{\partial J}{\partial \sigma_{ij}} = 0 \end{aligned} \quad (24b)$$

$$\partial_t \bar{r}_{ijl} + \frac{1}{\tau_{\sigma l}} [\lambda_l \partial_k \bar{v}_k \delta_{ij} + \mu_l (\partial_j \bar{v}_i + \partial_i \bar{v}_j) + \bar{r}_{ijl}] = 0. \quad (24c)$$

This set of equations is very similar to the forward model of eq. (1), the only difference being the sign of the spatial derivatives and the source terms. Hence, it can easily be implemented with the same forward modelling code. Furthermore, the finite difference solution is not affected by the sign changes and both forward and adjoint modelling share the same stability criteria. Our adjoint equations are different from (Tarantola 1988; Tromp *et al.* 2004) that are

given in the displacement formulation, or from (Yang *et al.* 2016) that results from the direct application of the adjoint state method to the original forward model, giving asymmetric forward and adjoint equations. However, all those formulations are equivalent, as they all share the same forward and adjoint second order wave equation for particle displacements:

$$\begin{aligned} \rho \partial_{tt} u_i - \lambda_0 \partial_i \partial_j u_j - \mu_0 \partial_j (\partial_j u_i + \partial_i u_j) - \partial_j R_{ijl}^l \delta_l &= 0, \\ \tau_i \partial_t R_{ijl} + R_{ijl} + \lambda_i \partial_k u_k \delta_{ij} + \mu_i (\partial_j u_i + \partial_i u_j) &= 0. \end{aligned} \quad (25)$$

Finally, the misfit gradient for the viscoelastic parameters is obtained by the derivative of the Lagrangian with respect to the inversion parameters. After the back transformation and some manipulations, we get:

$$\frac{\partial J}{\partial m_\alpha} = \left\langle \boldsymbol{\psi}, \mathbf{T} \frac{\partial \Lambda}{\partial m_\alpha} \mathbf{T} (\mathbf{A}\boldsymbol{\phi} + \mathbf{B}\boldsymbol{\phi} - \mathbf{s}) \right\rangle, \quad (26a)$$

$$= \left\langle \boldsymbol{\psi}, \mathbf{T} \frac{\partial \Lambda}{\partial m_\alpha} \mathbf{T} \mathbf{C}\boldsymbol{\phi} \right\rangle, \quad (26b)$$

where the last equality holds because of the wave equation (eq. 6). The misfit gradient can thus be obtained by the cross-correlation of the forward and adjoint wavefields. This cross-correlation can be computed in time or in the frequency domain using Parseval's theorem:

$$\begin{aligned} \frac{\partial J}{\partial m_\alpha} &= \int_T \int_X \boldsymbol{\psi} \mathbf{T} \frac{\partial \Lambda}{\partial m_\alpha} \mathbf{T} (\mathbf{A}\boldsymbol{\phi} + \mathbf{B}\boldsymbol{\phi} - \mathbf{s}) \, dx \, dt, \\ &= \int_\omega \int_X \mathcal{F}(\boldsymbol{\psi}) \mathbf{T} \frac{\partial \Lambda}{\partial m_\alpha} \mathbf{T} \mathcal{F}(\mathbf{A}\boldsymbol{\phi} + \mathbf{B}\boldsymbol{\phi} - \mathbf{s}) \, dx \, d\omega. \end{aligned} \quad (27)$$

Note that for finite difference, the material parameters are defined at discrete positions,  $m_\alpha(\mathbf{x}) \rightarrow m_\alpha \delta(\mathbf{x} - \mathbf{x}_\alpha)$ , and only the time integration is required. The explicit expressions of the misfit gradient with respect to  $\rho$ ,  $M$ ,  $\mu$ ,  $\tau_p$  and  $\tau_s$  are given in the appendix. The gradient for a different parametrization can be obtained using the chain differentiation rule.

### 3 CROSSTALK BETWEEN PARAMETERS

As stated in several studies (see Kamei & Pratt 2013), the crosstalk between the real and imaginary part of the velocities can be quite problematic and can cause strong anomalies in the recovered  $Q$  model. In fact, this is a general feature of multi-parameter FWI, where different classes of variables must be recovered during inversion. Operto *et al.* (2013) highlights different strategies to mitigate the effect of crosstalk between parameter types: the choice of a suitable parametrization, the use of the Hessian, a data-driven methodology where different parts of the data are targeted for different parameters and a model driven methodology where the dominant parameters are inverted before the less influential parameters.

The viscoelastic wave equation depends on a set of at least five parameters. An obvious choice is the density  $\rho$ , the  $P$ -wave velocity  $V_p$ , the  $S$ -wave velocity  $V_s$  and their respective attenuation levels  $\tau_p$  and  $\tau_s$ . As mentioned previously, the attenuation level controls the magnitude of the attenuation. Consequently,  $\tau_p$  and  $\tau_s$  should be a good choice of parameters to infer  $Q$  from FWI. On the other hand, the relaxation times, which control mostly the frequency dependency of  $Q$ , are not well constrained and are considered fixed in this study. We do not investigate further the choice of parametrization, but a complete study on the coupling of the parameter classes should be the focus of future work. The role of the Hessian should also be studied in more details.

**Table 1.** Summary of parameter pairs used to evaluate crosstalk.

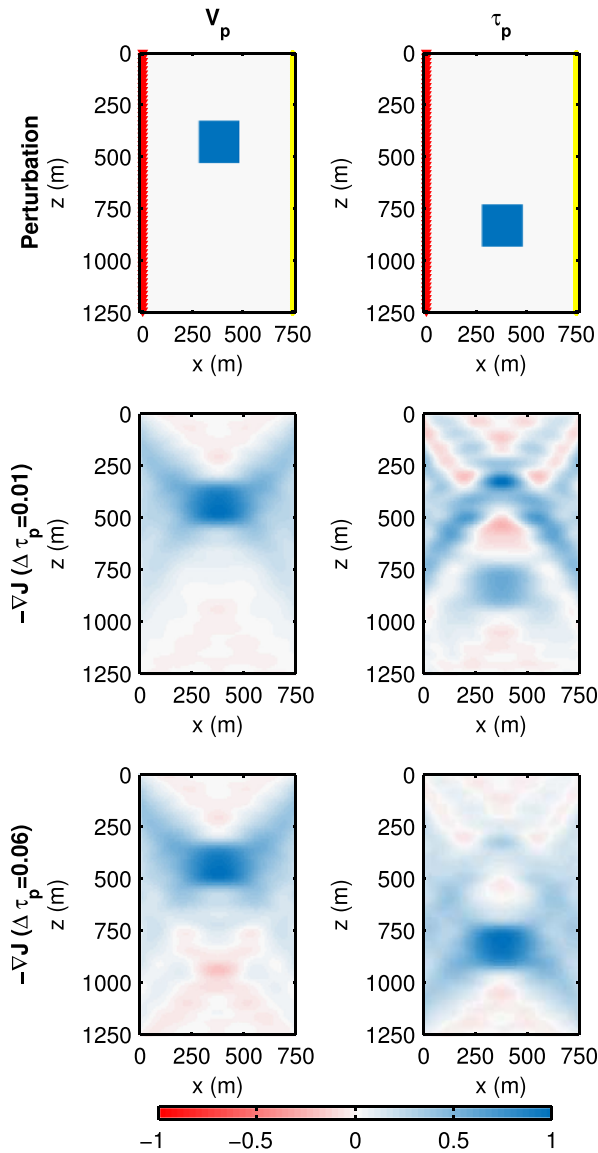
Figure	Source type	Parameter 1	Parameter 2
2	explosive	$\Delta V_p = 1.4$ per cent	$\Delta \tau_p = (50, 300)$ per cent
3	z-force	$\Delta V_s = 2.5$ per cent	$\Delta \tau_s = (50, 300)$ per cent
4	z-force	$\Delta V_p = (1.4, 8)$ per cent	$\Delta V_s = 2.5$ per cent
5	explosive	$\Delta V_p = 1.4$ per cent	$\Delta V_s = (2.5, 5)$ per cent

For the chosen set of parameters ( $\rho$ ,  $V_p$ ,  $V_s$ ,  $\tau_p$ ,  $\tau_s$ ), we investigate the crosstalk between parameter classes in the setting of a crosswell experiment. The choice of a crosswell setting is driven by the next section, where a synthetic crosswell FWI is presented. In the following, we use a source with the shape of a Ricker wavelet with a dominant frequency of 25 Hz. Source points are placed every 20 m along the left side of the model and geophones are placed on the right side at the same interval. Two source types are considered: an explosive source and a z oriented force. The first emits principally  $P$  waves and the latter emits principally  $S$  waves. To investigate the crosstalk between parameters, we introduce two perturbations for two different parameter classes at two different locations over a constant model ( $V_p = 3500$  m s<sup>-1</sup>,  $V_s = 2000$  m s<sup>-1</sup>,  $\rho = 2000$  kg m<sup>-3</sup>,  $\tau_p = 0.02$ ,  $\tau_s = 0.02$ ). We then calculate the gradient. For two uncoupled parameters, their gradient should be maximum around the location of the respective perturbation. On the contrary, a coupled parameter will suffer from crosstalk and its gradient won't point at the perturbation location. A summary of the following experiments is presented at Table 1.

In a first crosswell experiment, two perturbations,  $\Delta \tau_p$  and  $\Delta V_p$ , are introduced. We use an explosive source. The perturbation  $\Delta V_p$  is set at 50 m s<sup>-1</sup> (1.4 per cent) and two magnitudes are considered for  $\Delta \tau_p$ : 0.01 and 0.06 (50 and 300 per cent). The gradients for both  $V_p$  and  $\tau_p$  are computed for the two perturbation magnitudes and shown in Fig. 2. For the small  $\tau_p$  perturbation, the effect of the  $\tau_p$  anomaly is severely contaminated by the crosstalk from the velocity perturbation. On the other hand, the velocity gradient is relatively unaffected by the  $\tau_p$  perturbation. When the attenuation perturbation has a stronger magnitude, the effect of crosstalk on the gradient of  $\tau_p$  decreases but is still present. For the large  $\tau_p$  perturbation, the velocity gradient begins to show the effect of crosstalk, but it is still relatively small. This example shows that  $V_p$  is the dominant parameter over  $\tau_p$  and illustrates how a relatively small velocity error (1.4 per cent) could significantly disrupt the gradient update for the attenuation level.

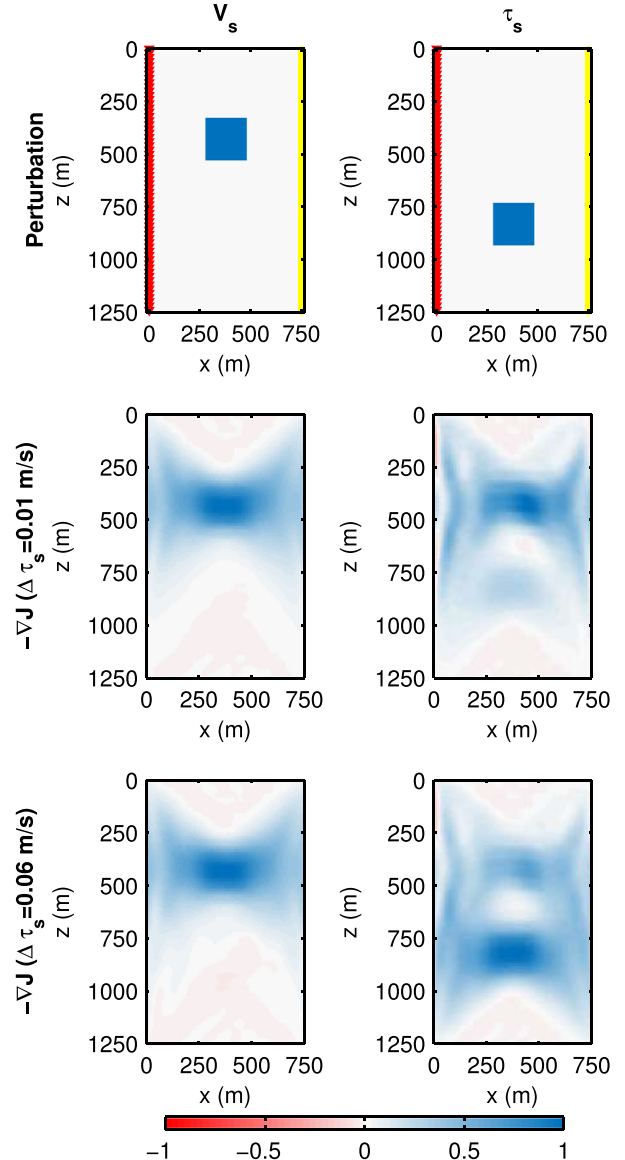
The results for the same experiment between  $\tau_s$  and  $V_s$  is shown in Fig. 3. In this case, we used a z-oriented force source. The perturbation  $\Delta V_s$  was set to 50 m s<sup>-1</sup> (2.5 per cent) and to 0.02 and 0.06 (50 and 300 per cent) for  $\Delta \tau_s$ . The same behaviour can be observed:  $V_s$  is very mildly affected by crosstalk whereas  $\tau_s$  is strongly affected. For the small  $\tau_s$  perturbation, the gradient is stronger at the  $V_s$  perturbation location and is strongly correlated to the  $V_s$  gradient. For the large perturbation, the correction is stronger at the  $\tau_s$  location, even though a non-negligible component is still present at the  $V_s$  perturbation location.

Crosstalk also exists between  $V_p$  and  $V_s$ . Two experiments were conducted with  $V_p$  and  $V_s$  perturbations: one for the z oriented force and one for an explosive force. The results are shown respectively in Figs 4 and 5. For the z-oriented source,  $\Delta V_s$  is kept fixed at 50 m s<sup>-1</sup> (2.5 per cent) with a small and a large  $\Delta V_p$  of 50 and 200 m s<sup>-1</sup> (1.4 and 8 per cent) respectively. For this source,  $V_s$  is the dominant parameter: its gradient is affected by neither the large nor the small  $V_p$  perturbation. On the contrary, the  $V_p$  gradient is affected by the



**Figure 2.** Effect of crosstalk between  $V_p$  (left column) and  $\tau_p$  (right column) for an explosive source. For a constant  $V_p$  perturbation, the  $\tau_p$  gradient is far more affected by crosstalk for a small  $\Delta\tau_p$  of 0.01 (50 per cent) (middle row) than for a large perturbation of 0.06 (300 per cent). Sources are represented by the red line and receivers by the yellow line.

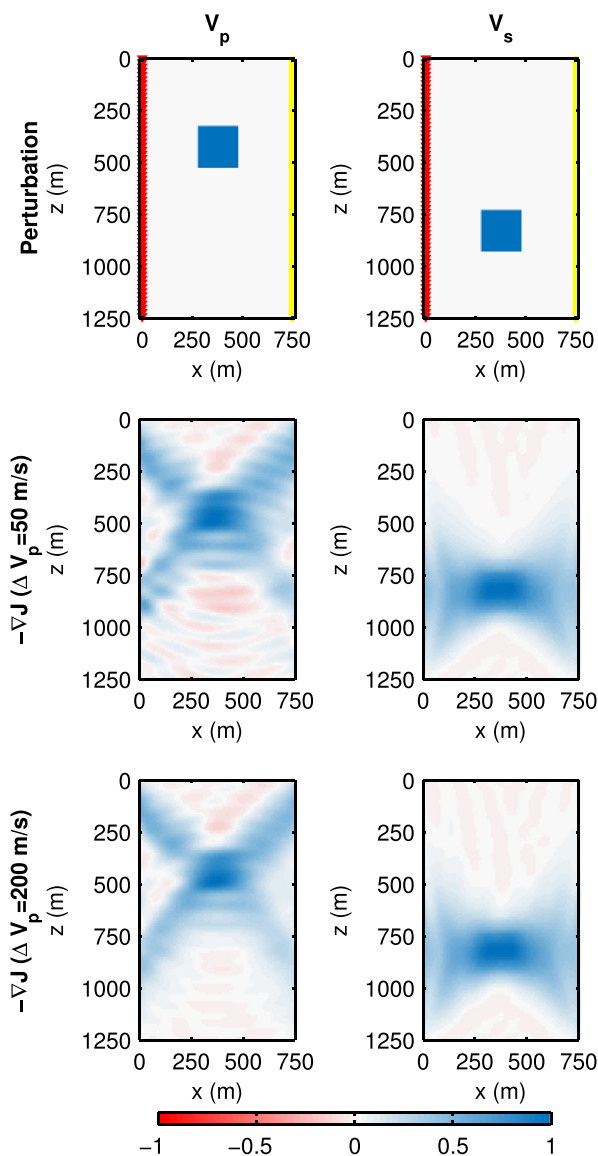
$V_s$  perturbation, with diminishing effect for larger  $V_p$  perturbations. Note that the crosstalk in this case is not as coherent as the crosstalk between velocities and the attenuation levels. For the explosive source,  $\Delta V_p$  is kept fixed at  $50 \text{ m s}^{-1}$  (1.4 per cent) and two sizes are considered for  $\Delta V_s$ :  $50 \text{ m s}^{-1}$  and  $100 \text{ m s}^{-1}$  (2.5 and 5 per cent). The results shown in Fig. 5 indicate that  $V_p$  is the dominant parameter for an explosive source, although it shows signs of crosstalk for the large  $V_s$  perturbation. The gradient for  $V_s$  is not coherent for neither perturbation sizes. These results can be explained by the radiation pattern of both sources: the explosive source produces only  $P$  waves and  $S$  waves are created only through mode conversion. This is the reason why the  $V_s$  gradient is not coherent for this source. For a  $z$ -oriented source,  $S$  waves are dominant for small angles from the horizontal and  $P$ -wave becomes dominant principally around  $\pm 45^\circ$ . Because both wave types are produced by this source, both  $V_p$  and  $V_s$



**Figure 3.** Effect of crosstalk between  $V_s$  (left column) and  $\tau_s$  (right column) for a  $z$ -force source. For a constant  $V_s$  perturbation, the  $\tau_s$  gradient is more affected by crosstalk for a small  $\Delta\tau_s$  of 0.01 (50 per cent) (middle row) than for a large perturbation of 0.06 (300 per cent). Sources are represented by the red line and receivers by the yellow line.

have coherent gradients pointing more or less in the right direction. However, because of the acquisition configuration of the crosswell experiment, small angles of propagation are favoured and  $S$  waves are much more energetic in the shot gathers.

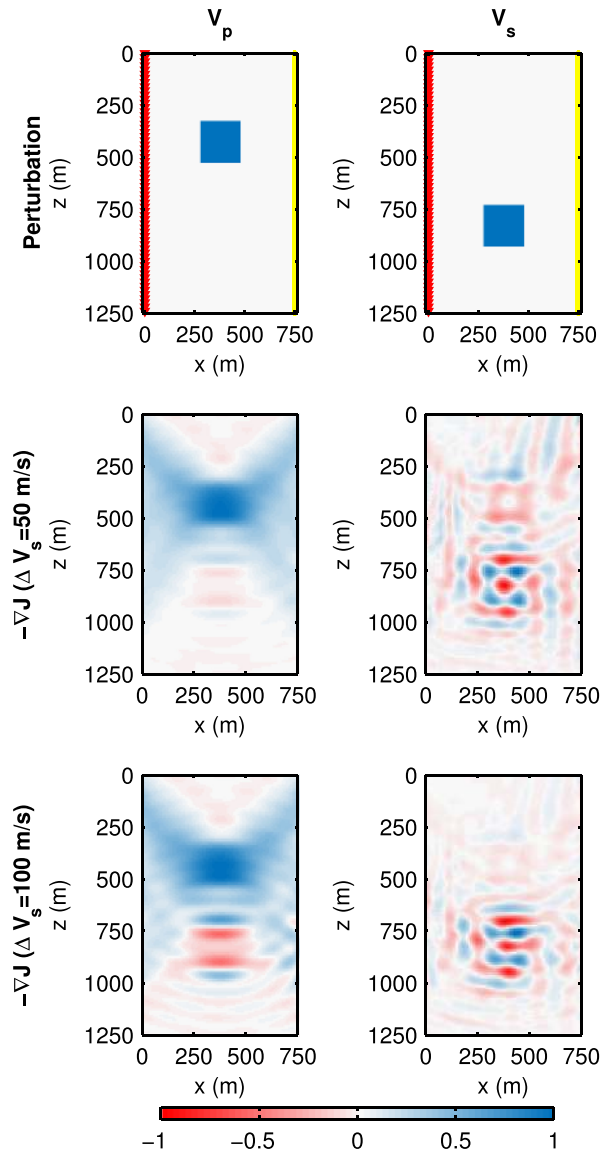
Other couplings between parameters are not shown herein but have the same kind of behaviour. For the explosive source, the parameters are, in order of importance,  $(V_p, V_s, \tau_p, \tau_s)$ . For the  $z$ -oriented source, the order of importance of the parameters is  $(V_s, V_p, \tau_s, \tau_p)$ . In both cases, the effects of  $\rho$  were not studied. Those results suggest two different methodologies to mitigate the effect of parameter crosstalk during inversion. The first is data driven: for  $V_p$  and  $\tau_p$ , explosive data should be used whereas the  $z$ -oriented force should be used for  $V_s$  and  $\tau_s$ . The second strategy is to first invert for the dominant parameters  $V_p$  and  $V_s$ , and then invert for the attenuation levels.



**Figure 4.** Effect of crosstalk between  $V_p$  (left column) and  $V_s$  (right column) for a  $z$ -force source. For a constant  $V_s$  perturbation, the  $V_p$  gradient is more affected by crosstalk for a small  $\Delta V_p$  perturbation of  $50 \text{ m s}^{-1}$  (1.4 per cent) (middle row) than for a large perturbation of  $200 \text{ m s}^{-1}$  (8 per cent). Sources are represented by the red line and receivers by the yellow line.

#### 4 INVERSION EXPERIMENT

Viscoelastic full waveform inversion is suitable in numerous situations where viscous attenuation and elasticity effects are of interest. In this section, we present one possible application: the time-lapse monitoring of  $\text{CO}_2$  sequestration by crosswell seismic tomography. Numerous studies show that crosswell seismic monitoring can detect changes in seismic velocities caused by  $\text{CO}_2$  injection (Spetzler *et al.* 2008; Onishi *et al.* 2009; Daley *et al.* 2011; Zhang *et al.* 2012; Ajo-Franklin *et al.* 2013). Furthermore,  $\text{CO}_2$  concentrations can affect substantially seismic attenuation (Carcione *et al.* 2006; Lei & Xue 2009; Müller *et al.* 2010). Hence, viscoelastic FWI is an attractive solution in this situation. Although we present a synthetic experiment, the geological setting is inspired by a study taking place in Quebec, Canada to evaluate the potential for  $\text{CO}_2$  geological storage in the province. The synthetic experiment aims at conforming

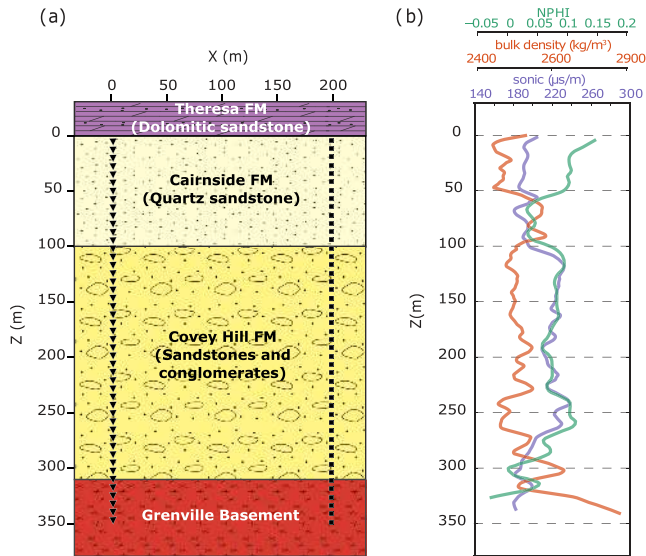


**Figure 5.** Effect of crosstalk between  $V_p$  (left column) and  $V_s$  (right column) for an explosive source. The gradient of  $V_p$  is mildly affected by crosstalk for the large  $V_s$  perturbation. The  $V_s$  gradient is not coherent here due to the very low  $S$ -wave energy produced by an explosive source. Sources are represented by the red line and receivers by the yellow line.

as much as possible to a realistic scenario. However, the goal of this study remains to show the applicability of the developed methodology, and bears to a lesser extent on  $\text{CO}_2$  sequestration.

##### 4.1 Geological context

The study site is located in the Cambrian-Ordovician sedimentary basin of the St. Lawrence Platform in southern Quebec, Canada, which has been identified as the most prospective basin for  $\text{CO}_2$  storage in the province (Malo & Bédard 2012). The targeted reservoir is the Potsdam Group, which lies unconformably upon the metamorphic Precambrian Grenville basement. It is comprised of the Covey Hill (Cambrian sandstones and conglomerates) and the Cairnside (lower Ordovician quartz sandstone) formations. The Potsdam Group is overlain by a succession of formations beginning with the Beekmantown Group (dolomitic sandstones and



**Figure 6.** The geological model showing the target reservoir for CO<sub>2</sub> injection in (a) and a representative log in the region in (b).

dolostones) to the Utica and Lorraine shales. The thickness (>800 m) and permeability ( $1 \times 10^{-19} \text{ m}^2$ ) of these units in the Becancour region are apparently capable of preventing buoyancy-driven migration of injected CO<sub>2</sub> to the surface, as they have maintained over-pressured conditions in the saline aquifers (Ngoc *et al.* 2014). Fig. 6 shows the units targeted during CO<sub>2</sub> injection monitoring, along a typical well log in the region.

## 4.2 Synthetic model and data

To build a suitable synthetic model, we consider a 2-D idealized geometry and physical model to describe the sedimentary sequence of interest. The geological model consists of a tabular succession of four horizontal layers corresponding to the Theresa formation, Cairnside formation, Covey Hill formation and the Grenville basement (Fig. 6). For each layer, the mean and the standard deviation of the physical properties ( $V_p$ ,  $V_s$ , porosity( $\phi$ ) and density ( $\rho$ )) are derived from well log data available in the studied region.

Based on this tabular model, a sequential Gaussian cosimulation (SGS) framework is used to create realistic heterogeneous models of  $V_p$ ,  $V_s$ ,  $\tau_p$ ,  $\tau_s$  and  $\rho$ . The resulting viscoelastic parameters are shown in the first column of Fig. 7. For all parameters, three zones can be identified: a zone of high velocity and medium attenuation between 0 and 100 m corresponding to the Cairnside formation, a zone of low velocity and high attenuation between 100 and 300 m corresponding to the Covey Hill formation and a zone of high velocity and low attenuation below 350 m corresponding to the basement. Substantial heterogeneity is introduced by SGS in all three units, with the highest level of heterogeneity inside the Covey Hill formation and the lowest level inside the basement. This model is referred hereafter as the baseline true model.

The seismic monitoring is simulated by two cross-hole surveys at time zero and at the end of injection after 15 yr. Both surveys share the same acquisition configuration. Shots are fired at a 1 m interval along the injection well located at  $x = 0$  in Fig. 6, from the top of the Cairnside formation at depth 0 to the Grenville basement at 350 m. Two types of sources are used: a force in the  $z$ -direction and an explosive source, both with the shape of a 500 Hz Ricker wavelet. This is representative of real borehole sources (Daley & Cox 2001;

Daley *et al.* 2007; Ajo-Franklin *et al.* 2013). The receiver well is 200 m away from the injection well. Two component geophones measure the  $x$  and  $z$  components of the seismic velocity ( $v_x$  and  $v_z$ ) at the same depth interval as the seismic sources.

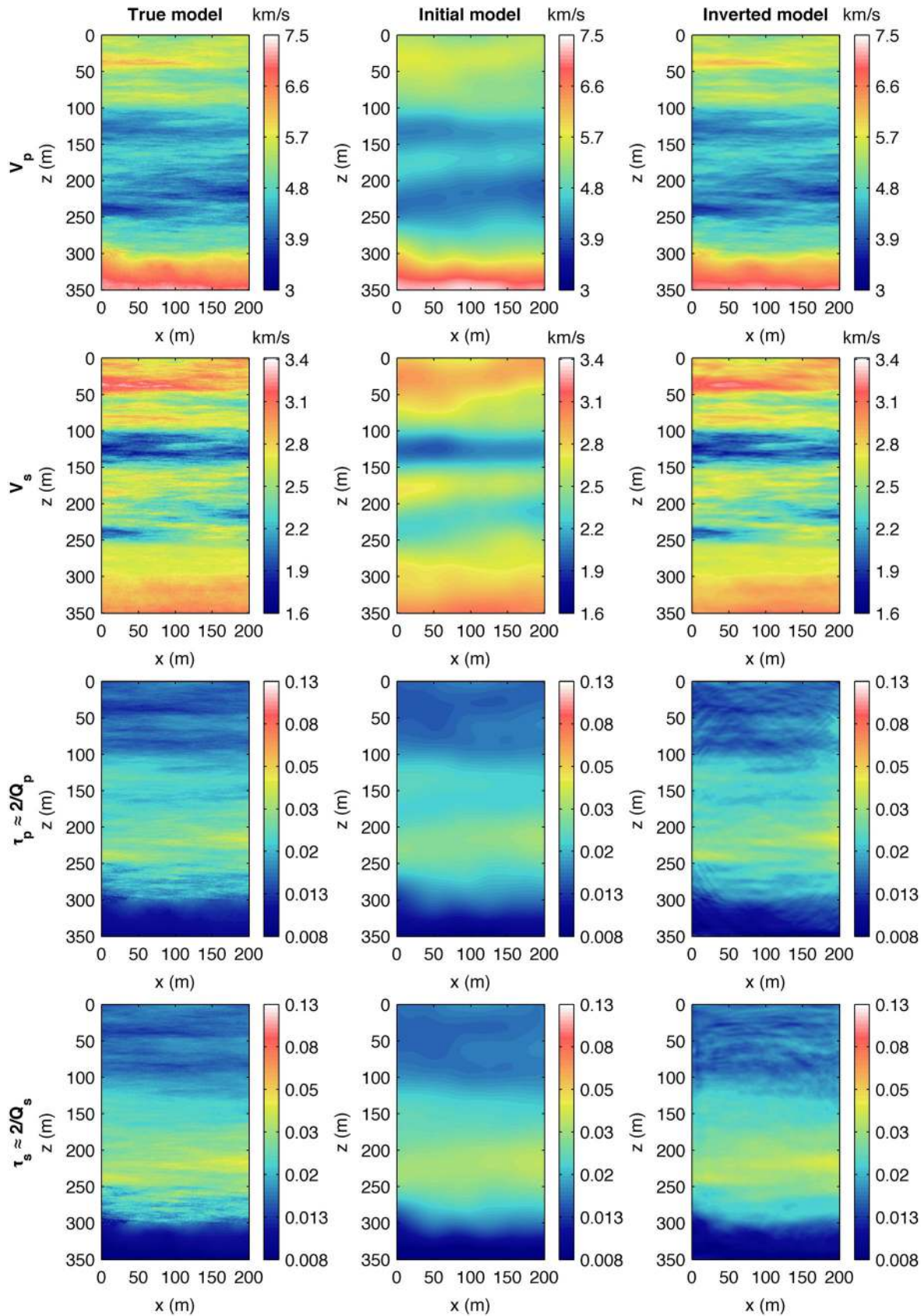
To simulate seismic propagation and to calculate the misfit gradient, we use the open source GPU accelerated seismic software SeisCL (Fabien-Ouellet *et al.* 2017), which implement the strategy presented previously (eqs (1), (24) and (26)). With a spatial finite difference operator of order 8 and the lowest velocity being  $1690 \text{ m s}^{-1}$ , we use a spatial grid with a 0.5 m step size to avoid dispersion up to 1000 Hz. To respect the stability criteria for velocities up to  $7300 \text{ m s}^{-1}$ , a time step of  $35 \mu\text{s}$  is chosen. Note that a single relaxation mechanism with a relaxation time  $\tau_{\sigma 1}$  of 500 Hz is used throughout this experiment. In that case, the attenuation levels and the quality factors at 500 Hz are related by  $\tau \approx 2/Q$  (from eq. 3). With those parameters, the synthetic data are calculated for the baseline and the injection models. Later inversions are performed with the same parameters.

The zero offset gather is shown in Fig. 8 for the baseline model. All gathers show very complex arrivals including direct waves, refracted waves, reflected waves and diffracted waves. This complexity is a direct consequence of the high spatial variability present in the synthetic models—velocities can vary by as much as 200 percent in a short distance range, creating low velocity zones. For this reason, this data set is a challenging case for FWI. The first arrivals of  $P$  and  $S$  waves were calculated by ray tracing and are marked with blue and red lines respectively in Fig. 8. Both  $P$  waves and  $S$  waves are visible for the  $z$ -force source gathers, although  $S$  waves are much more energetic. On the explosive source gathers, only first arrivals of  $P$  waves are clearly identifiable. Some low energy converted waves are also visible. Because  $P$  waves dominate the energy of the explosive source gathers and  $S$  waves dominate the  $z$ -force source gathers, using each data set separately for  $P$ -wave and  $S$ -wave inversion should minimize crosstalk.

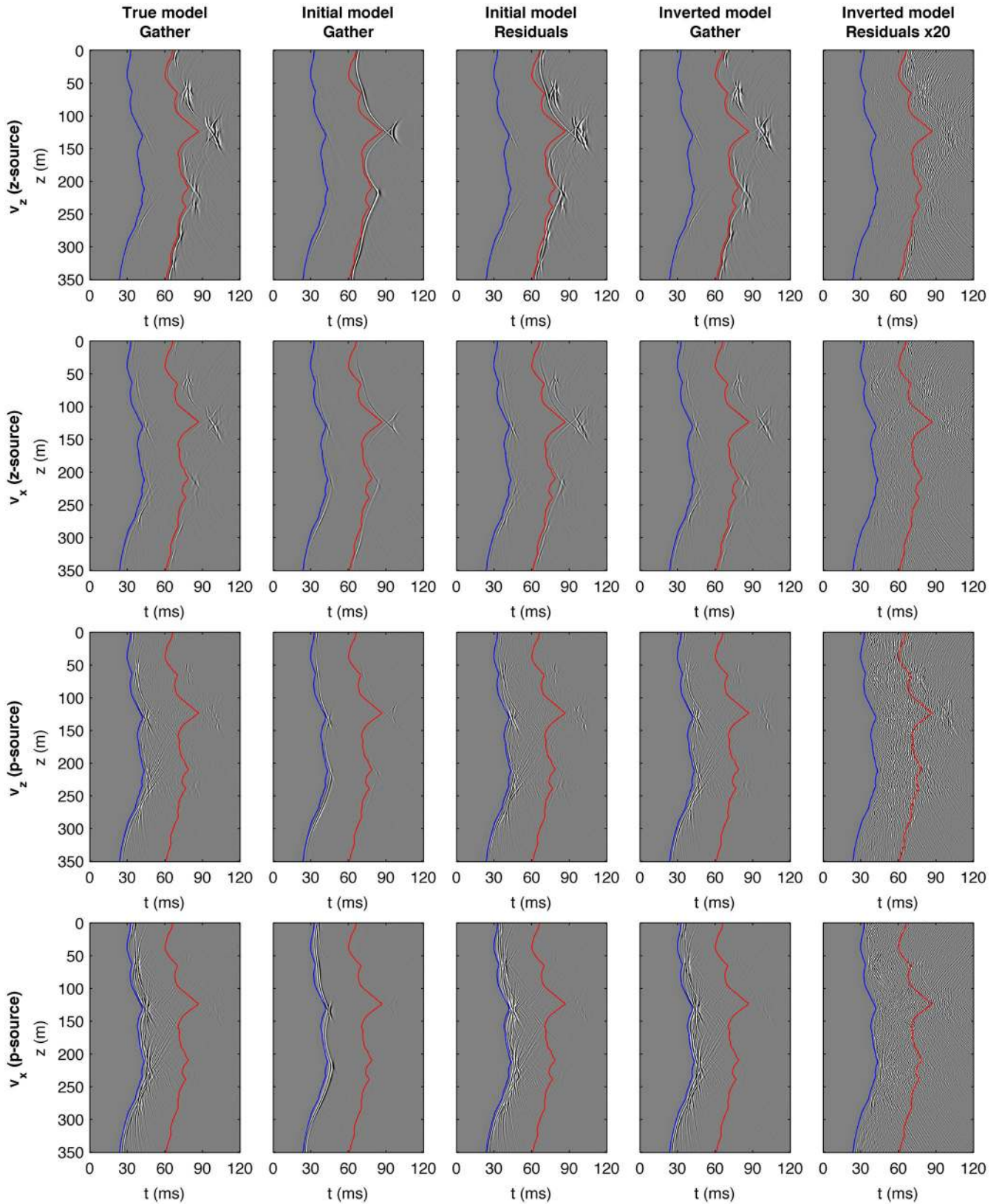
To model CO<sub>2</sub> injection, we used the vertical equilibrium solvers included in the Matlab Reservoir Simulation Toolbox (Lie *et al.* 2012). The heterogeneous geological model obtained by SGS is used as input. Injection of CO<sub>2</sub> is simulated in the Potsdam formations during 15 yr at an injection rate of  $45 \text{ t d}^{-1}$ , a rate comparable to the average rate for injection at Ketzin (Martens *et al.* 2012). The resulting CO<sub>2</sub> concentrations are presented in Fig. 9. The CO<sub>2</sub> plume after 15 yr has reached the monitoring well, with saturation ranging from 0 per cent to more than 40 per cent.

To translate the presence of CO<sub>2</sub> into a change in  $V_p$  and  $Q_p$ , the method of Carcione *et al.* (2006) based on White's model of patchy saturation (White 1975) is used, with an arbitrary patch size of 5 mm and a central frequency of 800 Hz. As pore fluid should not impact  $S$ -wave velocities, we do not introduce any changes to  $V_s$  and  $Q_s$ , even though other factors could introduce changes in real-world scenarios (pore pressure changes for example). The resulting model is presented in the first column of Fig. 10. Note that  $V_s$  and  $Q_s$  models are identical to the baseline model. The shape of the plume is clearly visible in  $Q_p$ , and its boundary is evident between depths of 150 to 250 m. The presence of CO<sub>2</sub> results in both lower  $V_p$  and  $Q_p$  values, up to 24 per cent and 73 per cent respectively. These are quite large for changes induced by the presence of CO<sub>2</sub>. It is, however, intentional, and those variations come from the arbitrary choice of White's model parameters. Such large changes allow us to better study the impact of different magnitudes of attenuation on FWI, which is the main goal of this experiment.





**Figure 7.** The baseline true model (first column), initial model (second column) and inverted model (third column). The  $P$ -wave velocity is shown on the first row, the  $S$ -wave velocity on the second row and the  $P$ - and  $S$ -wave attenuation levels on the third and fourth rows.



**Figure 8.** The baseline zero offset shot gathers for the true model (first column), the zero offset gather and residuals for the initial model (second and third columns) and the zero offset gather and residuals for the inverted model (fourth and fifth columns). The first and second rows show the  $v_z$  and  $v_x$  components for the  $z$ -force source and the third and fourth columns show the  $v_z$  and  $v_x$  components for the explosive source. The blue lines show the  $P$ -wave first arrival times and the red lines show the  $S$ -wave arrival times.

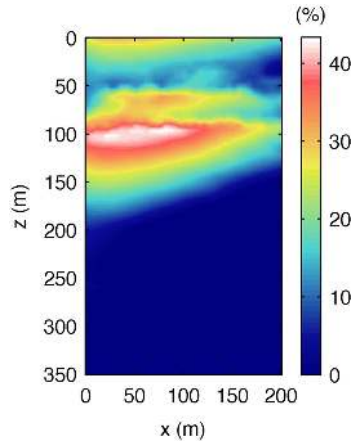


Figure 9. The CO<sub>2</sub> plume after 15 yr of injection.

### 4.3 Inversion methodology

For inversion, we use stochastic gradient descent (SGD), an inversion strategy now commonly used in the big data community for large-scale problems (Bottou 2010). Like gradient descent, SGD uses the gradient of the misfit function as the update direction. However, SGD uses a random subset of the data to compute the misfit gradient. This random subset changes at every iteration. Formally, SGD uses the following update formula:

$$m_{k+1} = m_k - \eta \nabla J_{\Omega}(m), \quad (28)$$

where  $m_k$  is the parameter to be optimized at iteration  $k$ ,  $\eta$  is the step size, and  $\nabla J_{\Omega}(m)$  is the misfit gradient in respect to  $m$  on the data subset  $\Omega$ . In this work, the step size is determined by a backtracking line search with Armijo condition (Armijo 1966). The data subset is taken as a uniform random subset of the sources. We use a variant of SGD with a growing batch size as described by Friedlander & Schmidt (2012) and applied to FWI by van Leeuwen & Herrmann (2013). The effect of the growing batch size is to speed-up convergence at latter iterations. The interest of SGD over standard gradient descent is to speed up the inversion, as the misfit gradient becomes less expensive to compute when the number of sources decreases. Because seismic data is usually highly redundant, subsampling does not affect too much convergence.

To avoid cycle skipping, we adopt a multiscale inversion strategy (Bunks *et al.* 1995), starting the inversion from low frequencies to high frequencies. The misfit gradient is calculated in the frequency domain using eq. (27). Because the resolution in the time-domain of the wave equation solves all frequencies simultaneously, it is relatively cheap to extract several frequencies during time-stepping with the discrete Fourier transform (Furse 2000). We use different frequencies for  $V_p$  and  $V_s$  due to their different wavelengths. For  $V_s$ , we use frequencies from 80 to 400 Hz every 20 Hz. For  $V_p$ , we use frequencies from 150 Hz to 950 Hz every 50 Hz. Due to the high attenuation levels, the inversion process will likely benefit from a large frequency bandwidth, and we divided the previous frequencies into two frequency bands: 80 to 200 Hz and 200 to 400 Hz for  $\tau_s$  and 150 to 500 Hz and 500 to 950 Hz for  $\tau_p$ . For each frequency band, all frequencies are inverted simultaneously. Every inversion step

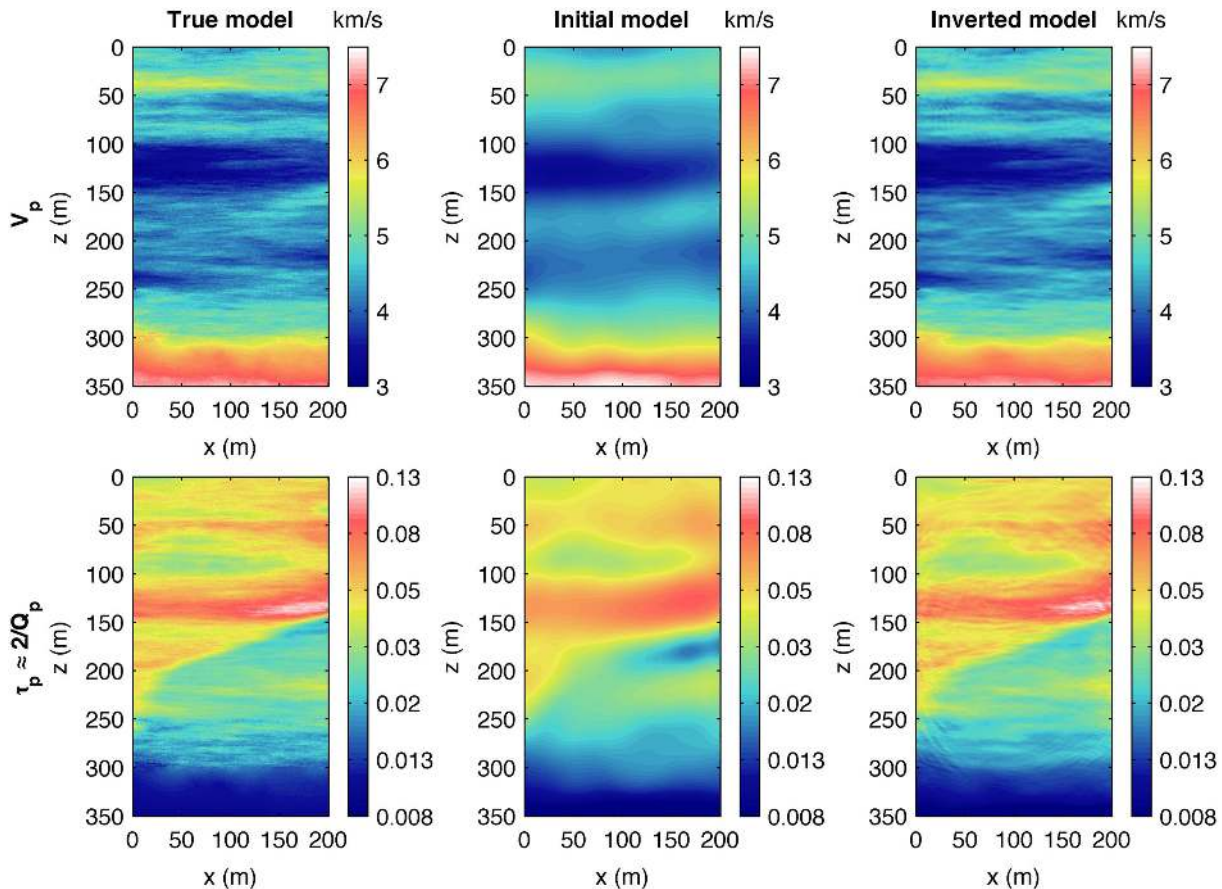


Figure 10. The model after the CO<sub>2</sub> injection: true model (first column), initial model (second column) and inverted model (third column). The  $P$ -wave velocity is shown on the first row and the  $P$ -wave attenuation level on the second row.

contains 40 iterations, each using a random subset of the sources. The number of sources grows with the frequency, from 24 at the lowest frequency to 76 at the highest frequency.

To reduce crosstalk between parameter types, we use the two strategies mentioned in the previous section. To invert for  $V_p$  and  $\tau_p$ , we use the explosive source data set and for  $V_s$  and  $\tau_s$ , we use the  $z$  oriented force. For each frequency, we invert first for  $V_s$  and then  $V_p$ . The attenuation levels  $\tau_s$  and  $\tau_p$  are inverted after the inversions for  $V_s$  and  $V_p$  are finished for the highest frequency. This sequential inversion allows to circumvent the problem of the scaling of the gradient between parameter classes (Kamei & Pratt 2013), as we focus on one type of parameter at a time. Because FWI is not very sensitive to density, we decided not to include it in this experiment. Furthermore, we do not invert for the source signature, but it can be performed easily as in Song *et al.* (1995). The inversion methodology is summarized in Algorithm 1.

---

#### Algorithm 1 Inversion methodology

---

```

1: Define  $freqvs = [80 : 20 : 400]$ 
2: Define  $freqvp = [150 : 50 : 950]$ 
3: Define  $freqall = freqvs + freqvp$ 
4: for all  $freq \in freqall$  do
5:   if  $freq$  in  $freqvs$  then
6:     SGD( $V_s, freq, 8 + 3.3ns, zdata, 40$ )
7:     Increment  $ns$ 
8:   if  $freq$  in  $freqvp$  then
9:     SGD( $V_p, freq, 8 + 3.3np, pdata, 40$ )
10:    Increment  $np$ 
11: SGD( $\tau_s, freqvs < 200, 120, zdata, 40$ )
12: SGD( $\tau_s, freqvs \geq 200, 120, zdata, 40$ )
13: SGD( $\tau_p, freqvp < 400, 120, pdata, 40$ )
14: SGD( $\tau_p, freqvp \geq 400, 120, pdata, 40$ )

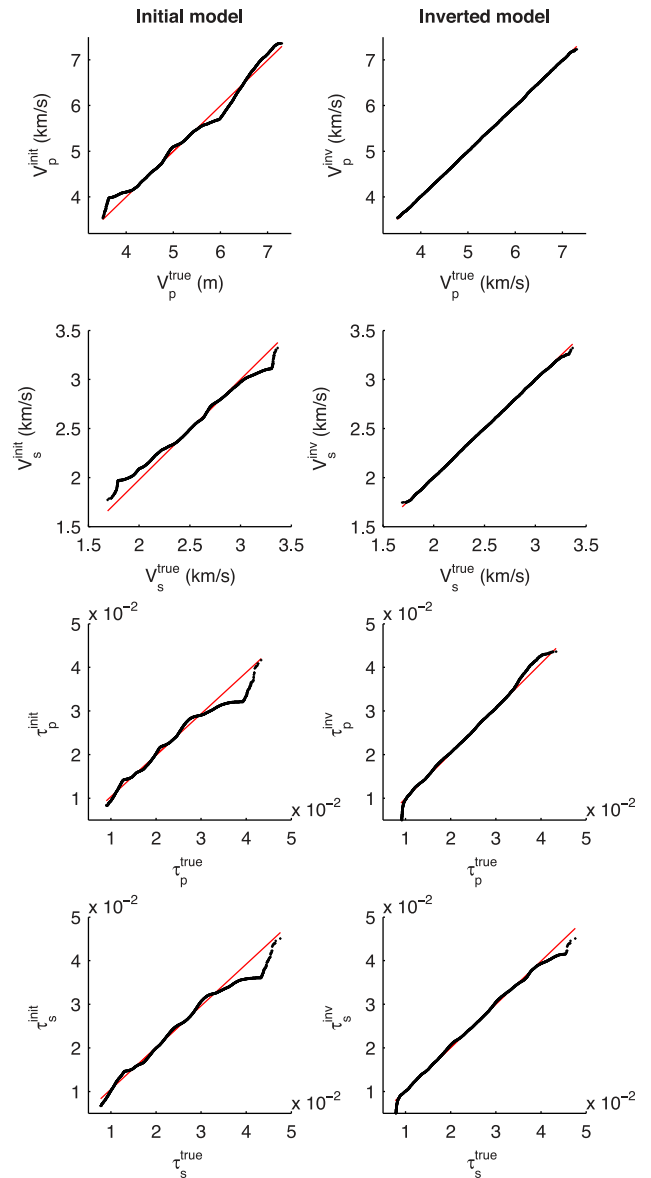
15: procedure SGD( $param, freqs, nsource, data, niter$ )
16:   while  $iter < niter$  do
17:     Draw  $nsource$  from  $data$ 
18:     Calculate  $gradient$  for  $freqs$  and  $selected data$ 
19:     Line search with  $gradient$ 
20:     Update  $model$ 
21:     Increment  $iter$ 

```

---

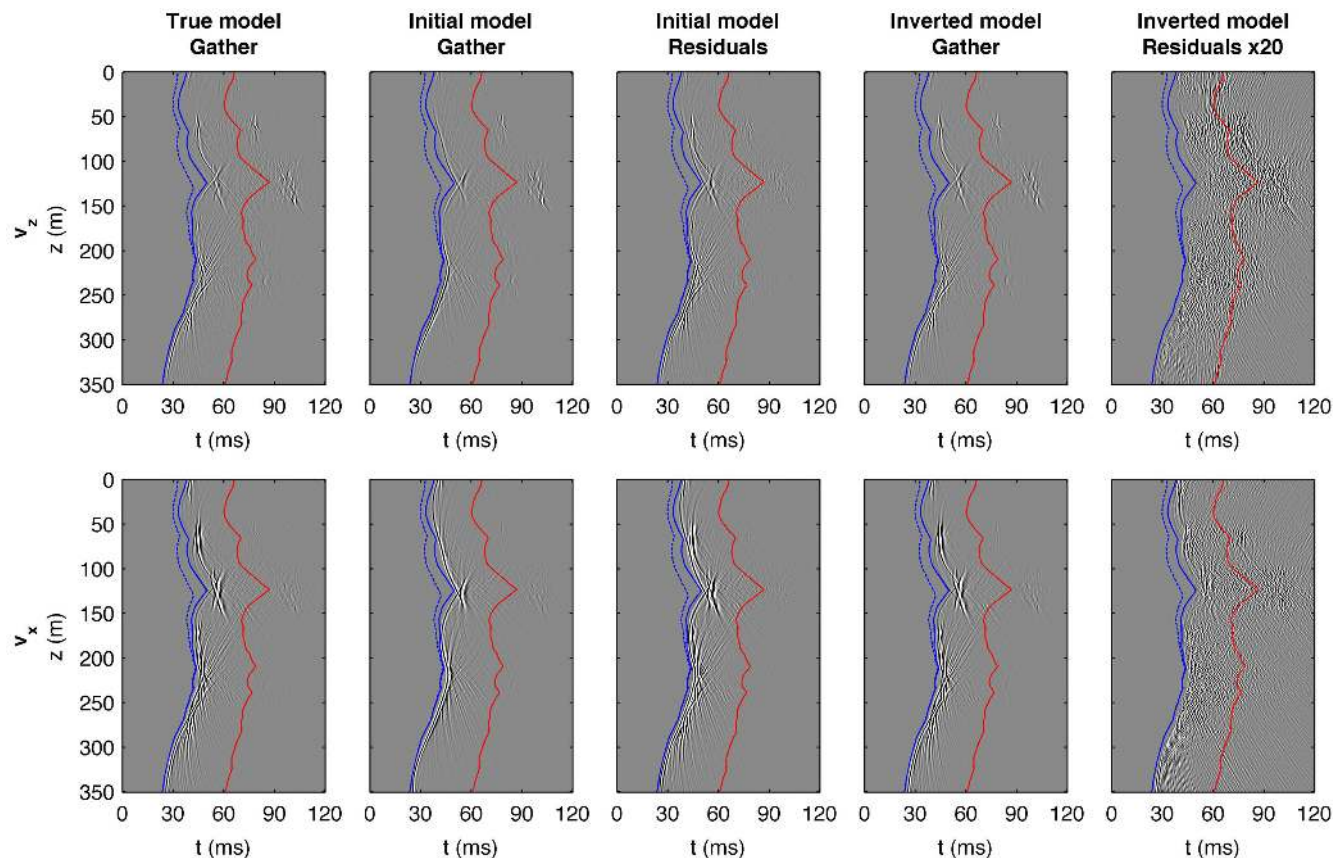
#### 4.4 Inversion of baseline data

The initial model used to invert the baseline data is a smoothed version of the true model (second column of Fig. 7). This kind of model is realistic in the sense that it could be obtained by traveltimes and amplitude tomography based on ray tracing. It still contains large scale variations and layers are identifiable. The goal of FWI in this circumstance is to attain a higher resolution throughout the model and to obtain a distribution of the optimized parameters closer to the true one, especially for the extremums. Comparing the seismic gathers for the initial and the true model (first and second column of Fig. 8, respectively), we see that the waveform contains only the first arrival for the initial model whereas it contains complex wave fronts consisting in direct, refracted, reflected and converted waves for the true model. The residuals (third column) are strongly cycle-skipped, and we can clearly see the superposition of the initial and true model gathers. In short, the true model does a very poor job of predicting the recorded wavefield.



**Figure 11.** Quantile–quantile plots between the true baseline model and the models before and after inversion. The red line represents a perfect match between distributions, the black dots represent the measured quantiles.

The inverted model is shown in the third column of Fig. 7. For both  $V_p$  and  $V_s$ , the high wavenumber details are very well reconstructed. Only the very high wavenumbers are not recovered, which is something to be expected due to the bandwidth-limited inversion. It is fair to say that the inversion converged to the right solution in both cases. This is confirmed by the quantile–quantile plot of Fig. 11. The quantile–quantile plot was constructed by computing percentiles of the true and inverted models and plotting them against each other. If both models share the same distribution of values, their quantiles should be equal. On the contrary, if both distributions are different, for example if one model is smoother or has some kind of bias, the quantile–quantile plot will diverge from a straight line. Before inversion, extreme values diverge significantly from the true model in this quantile–quantile plot. This is expected as we start from a smooth model, that is, a model lacking extreme values. After inversion, the match between the distribution of the



**Figure 12.** The zero offset shot gathers after injection for the true model (first column), the zero offset gather and residuals for the initial model (second and third columns) and the zero offset gather and residuals for the inverted model (fourth and fifth columns). The first and second rows show the  $v_z$  and  $v_x$  components for the explosive source. Dashed and solid blue lines show respectively the  $P$ -wave first arrival times before and after injection and solid red lines show the  $S$ -wave arrival times.

inverted and true model is nearly exact, except for very small or very high values, showing that much of the variability of the velocities parameters was well recovered.

On the other hand,  $\tau_p$  and  $\tau_s$  are not resolved with the same level of accuracy. Some of the high-resolution details are recovered, especially in regions of higher attenuation. However, both parameters suffer from the crosstalk with the small anomalies remaining in the  $V_p$  and  $V_s$  models. As shown in the previous section, this crosstalk does happen for quite small velocity errors. Hence, to be able to recover the attenuation levels, very accurate velocity models are needed. The effect of crosstalk can be seen on the quantile–quantile plot of Fig. 11 for small values, where the match seems to deteriorate after inversion. This plot also shows that the high values are not very well recovered, but that inversion does improve the match for intermediate values. Overall, FWI increased the frequency content of the initial model.

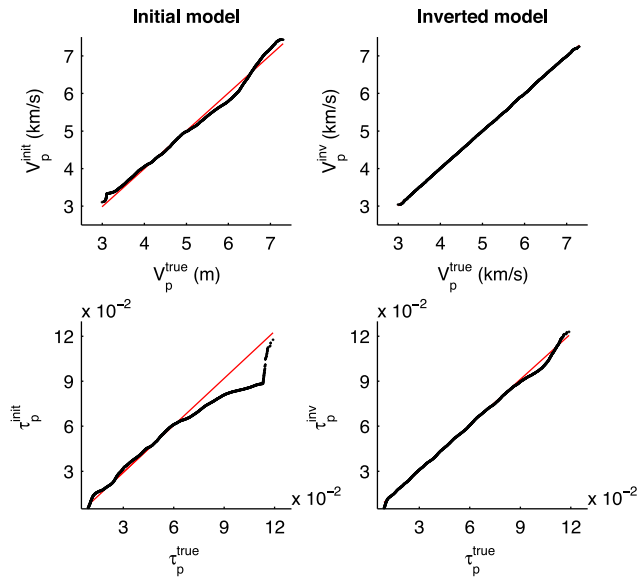
Comparing the seismic gathers in the first and third columns of Fig. 8, we see that the wavefield is very well reproduced and contains all the complex arrivals mentioned previously. The residuals have a much smaller amplitude than either the recorded data or the initial model residuals. It can also be observed that most of the remaining energy corresponds to converted waves and multiply scattered arrivals. These wave types are generally very difficult to reproduce with FWI. Furthermore, most of the remaining energy is of high frequency. Due to more severe cycle skipping, high frequencies are much more difficult to invert and frequencies higher

than 400 Hz were not even inverted for the  $z$ -force source. Hence, most of the energy should be of higher frequency, as observed. The improved fit between observed and predicted data leads to the conclusion that the inversion has successfully converged to the right model.

#### 4.5 Inversion after CO<sub>2</sub> injection

The initial model used to invert the seismic data after 15 yr of injection is also a smoothed version of the true model for  $V_p$  and  $\tau_p$  (second column of Fig. 10). The inverted model from the baseline for  $V_s$  and  $\tau_s$  are used for this case, as we made the assumption that the shear wave velocity is not affected by the presence of CO<sub>2</sub>. Hence, this inversion only focus on  $V_p$  and  $\tau_p$ . Contrary to the baseline model, the explosive source gathers for the initial model (second column of Fig. 12) contains not only  $P$ -wave direct arrivals but also quite a lot of converted waves. This is a consequence of using the inverted  $V_s$  model that reproduce much of the spatial heterogeneity of the true model. This shows the importance of elastic effects even when only  $V_p$  is of interest.

The inversion results are shown in the third column of Fig. 10 and the quantile–quantile plot is shown in Fig. 13. Once again,  $V_p$  is very well reconstructed and the quantile–quantile distribution matches almost exactly. Moreover, reconstruction of the attenuation level shows much more details than for the baseline data. Fig. 13 shows that extreme values are also better reproduced. Crosstalk anomalies



**Figure 13.** Quantile–quantile plots between the true model after injection and the models before and after inversion. The red line represents a perfect match between distributions, the black dots represent the measured quantiles.

are much weaker, although still present. This is explained by the overall higher attenuation in this model: as seen previously, higher attenuation levels reduce the crosstalk caused by errors in the velocity model, without affecting too much the velocity reconstruction. Finally, after inversion, the wavefield is very similar to the synthetic data, as shown by the small amplitude of the data residuals (fifth column of Fig. 12). As for the baseline inversion, the residuals contain primarily high frequencies and converted waves. This shows that the inversion converged.

#### 4.6 CO<sub>2</sub> monitoring

Because both  $V_p$  and  $\tau_p$  are sensitive to the presence of CO<sub>2</sub>, their changes in time can be used to track the evolution of the plume. This is depicted in Fig. 14. The first column shows the CO<sub>2</sub> saturation and the change in  $V_p$  and  $Q_p$  of the true models. Because we used White’s model of patchy saturation, changes in  $V_p$  and  $Q_p$  are not linked linearly to CO<sub>2</sub> saturation. Hence, it is not possible to recover the concentration directly from the changes in the seismic parameters. Note, however, that larger changes in  $V_p$  are more or less correlated with larger concentrations. The inverted changes were obtained by the difference of the inverted model after injection with the baseline model. Comparing the second column of Fig. 14, we see that changes in  $V_p$  are very well recovered. Some artefacts in the inversion results are present which are caused by the limited aperture of the acquisition geometry and crosstalk. Even though relative changes are larger for  $Q_p$  than for  $V_p$ , reconstruction for  $Q_p$  (third column of Fig. 14) shows more artefacts. Still, FWI manages to recover most of the large-scale variations found in the true model. The larger errors for the changes in  $Q_p$  are indeed caused by the larger errors of the inverted model in  $\tau_p$ , which are caused by the smaller sensitivity of the viscoelastic wave equation to  $\tau_p$  than to  $V_p$ . This experiment highlights the very high resolution that can be obtained with FWI in time-lapse monitoring, even with this very rudimentary time-lapse inversion methodology.

## 5 CONCLUSIONS

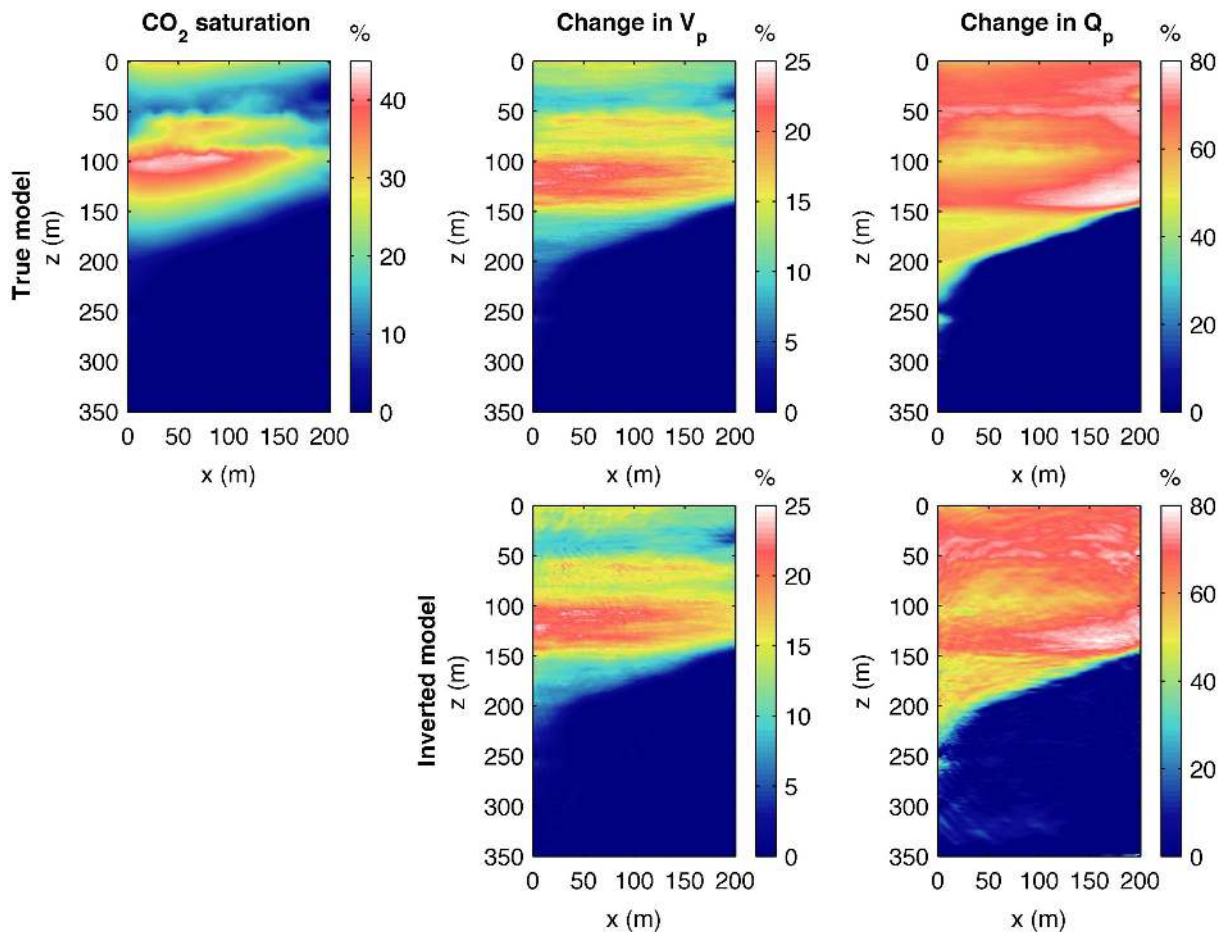
In this paper, we studied the application of FWI to the viscoelastic wave equation in the velocity–stress formulation based on the generalized standard linear solid rheology. We proposed a new strategy to compute the misfit gradient with the adjoint state method that allows the same modelling code to be used to compute both the forward and adjoint wavefields. To do so, we transformed the viscoelastic wave operator into a sum of self and anti-self-adjoint operators. By deriving the adjoint state equations for the transformed operators and applying the inverse transformations, we obtained a set of partial differential equations that is identical to the forward equations, up to a sign change. Note that the total forward ( $A + B - C$ ) and the transformed adjoint ( $A - B + C$ ) wave operators are not self-adjoint because of the sign change. However, the only modification required to the forward modelling code to calculate the adjoint wavefield is minor. In comparison, the direct application of the adjoint state method to the viscoelastic wave equation leads to two distinct sets of partial differential equations that must be solved by two different modelling codes (Yang *et al.* 2016). This fact is not negligible, considering that a highly optimized modelling code can be very time consuming to produce.

We have also investigated the crosstalk between parameter classes of the ( $\rho$ ,  $V_p$ ,  $V_s$ ,  $\tau_p$ ,  $\tau_s$ ) parametrization. We have shown that the velocity parameters are dominant over the attenuation parameters, meaning that the velocities are less affected by the attenuation levels errors than the contrary. This result, however, is not general and depends on the acquisition configuration used in this example. Indeed, the footprint of the velocities and the attenuation levels on the seismic data changes with the source–receiver offset. For a large enough propagation distance, attenuation may become dominant. Also, as shown in the time-lapse inversion experiment, sensitivity of FWI to attenuation increases for higher attenuation levels. Hence, the ability of FWI to recover the  $Q$  factor depends on the acquisition configuration and on the degree of attenuation of the Earth. This should be assessed on a case-by-case basis.

We have also shown that crosstalk can be reduced between  $S$ - and  $P$ -wave parameters by using different source types during acquisition. Explosive sources are more appropriate for  $V_p$  and  $Q_p$ , whereas oriented force sources may be used to generate more  $S$  waves than  $P$  waves and thus favour the inversion of  $V_s$  and  $\tau_s$ . Both types of sources exist for land and borehole environments, although it is not always realistic to acquire two different sets of data. Better ways to uncouple the parameter classes could be used to increase convergence when only one set of data is available, for example when trying to recover  $V_s$  from converted waves. The use of the Hessian is one possibility for that purpose (Operto *et al.* 2013).

In the last section, we showed how the viscoelastic FWI formulation presented in the paper can be used for a multi-parameter inversion in a synthetic cross-hole experiment. For a low attenuation model (the baseline model), the inversion successfully recovered both  $V_p$  and  $V_s$  and to a lesser extent  $\tau_p$  and  $\tau_s$ . In this case, the attenuation of the baseline model is too low, which led to strong crosstalk anomalies in the inverted attenuation levels, even though the velocity models were nearly exactly recovered. After CO<sub>2</sub> injection, the  $P$ -wave attenuation level is much higher and the inverted  $\tau_p$  model is consequently less affected by crosstalk. This experiment showed that for realistically complex models, FWI can indeed recover velocities and attenuation levels if attenuation is high enough.

Finally, our synthetic experiment shows that FWI is a promising methodology to monitor changes in  $V_p$  and  $Q_p$  caused by the



**Figure 14.** The  $\text{CO}_2$  saturation after 15 yr of injection (first column), the change in  $V_p$  (second column) and in the quality factor  $Q_p$  (third column) between the baseline and the model after injection. The first row shows the true model and the second row the change recovered after inversion.

injection of  $\text{CO}_2$ . Indeed, the spatial resolution obtained by FWI is the highest attainable by any seismic method. However, caution is advised when interpreting the quality of our results. First, the experiment is noise-free and based on a perfect physics model. In reality, coherent and incoherent noise will be present in the seismic data. Also, 3-D effects were not considered herein. At least a line source to point source conversion should be performed before inversion in a real case. The 3-D nature of the  $\text{CO}_2$  plume should also have to be taken into account. Finally, to convert modelled  $\text{CO}_2$  saturation into changes in  $V_p$  and  $\tau_p$ , we used White's model of patchy saturation, using arbitrary parameters that were not calibrated on real data. In fact, we chose conveniently those parameters to obtain a model with high attenuation to highlight the effect of light versus strong attenuation on FWI. To better assess the ability of FWI to track changes in  $V_p$  and  $Q_p$ , a calibrated relationship should be used.

## ACKNOWLEDGEMENTS

This work was funded by a Vanier Canada Graduate Scholarship and supported by the Canada Chair in geological and geophysical data assimilation for stochastic geological modelling. We thank Lorenzo Perozzi for allowing us to use his geological model.

## REFERENCES

Ajo-Franklin, J.B., Peterson, J., Doetsch, J. & Daley, T.M., 2013. High-resolution characterization of a  $\text{CO}_2$  plume using crosswell seismic to-

mography: Cranfield, MS, USA, *Int. J. Greenhouse Gas Control*, **18**, 497–509.

Armijo, L., 1966. Minimization of functions having Lipschitz continuous first partial derivatives, *Pac. J. Math.*, **16**(1), 1–3.

Askan, A., Akcelik, V., Bielak, J. & Ghattas, O., 2007. Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures, *Bull. seism. Soc. Am.*, **97**(6), 1990–2008.

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

Barton, N., 2007. *Rock Quality, Seismic Velocity, Attenuation and Anisotropy*, CRC Press.

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

Bohlen, T., 2002. Parallel 3-D viscoelastic finite difference seismic modelling, *Comput. Geosci.*, **28**(8), 887–899.

Bottou, L., 2010. *Large-scale Machine Learning with Stochastic Gradient Descent*, pp. 177–186, Springer.

Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.

Carcione, J.M., Kosloff, D. & Kosloff, R., 1988. Viscoacoustic wave propagation simulation in the earth, *Geophysics*, **53**(6), 769–777.

Carcione, J.M., Picotti, S., Gei, D. & Rossi, G., 2006. Physics and seismic modeling for monitoring  $\text{CO}_2$  storage, *Pure appl. Geophys.*, **163**(1), 175–207.

Castellanos, C., Etienne, V., Hu, G., Operto, S., Brossier, R. & Virieux, J., 2011. Algorithmic and methodological developments towards full

- waveform inversion in 3D elastic media, in *Expanded Abstracts of 81st Annual International SEG Meeting*, pp. 2793–2798.
- Causse, E., Mittet, R. & Ursin, B., 1999. Preconditioning of full-waveform inversion in viscoacoustic media, *Geophysics*, **64**(1), 130–145.
- Charara, M., Barnes, C. & Tarantola, A., 2000. *Full Waveform Inversion of Seismic Data for a Viscoelastic Medium*, pp. 68–81, Springer.
- Daley, T.M. & Cox, D., 2001. Orbital vibrator seismic source for simultaneous *P*- and *S*-wave crosswell acquisition, *Geophysics*, **66**(5), 1471–1480.
- Daley, T.M., Solbau, R.D., Ajo-Franklin, J.B. & Benson, S.M., 2007. Continuous active-source seismic monitoring of CO<sub>2</sub> injection in a brine aquifer, *Geophysics*, **72**(5), A57–A61.
- Daley, T.M., Ajo-Franklin, J.B. & Doughty, C., 2011. Constraining the reservoir model of an injected CO<sub>2</sub> plume with crosswell CASSM at the Frio-II brine pilot, *Int. J. Greenhouse Gas Control*, **5**(4), 1022–1030.
- Emmerich, H. & Korn, M., 1987. Incorporation of attenuation into time-domain computations of seismic wave fields, *Geophysics*, **52**(9), 1252–1264.
- Fabien-Ouellet, G., Gloaguen, E. & Giroux, B., 2016. The adjoint state method for the viscoelastic wave equation in the velocity-stress formulation, in *78th EAGE Conference and Exhibition 2016 EAGE*, Extended abstract, doi:10.3997/2214-4609.201600826.
- Fabien-Ouellet, G., Gloaguen, E. & Giroux, B., 2017. Time-domain seismic modeling in viscoelastic media for full waveform inversion on heterogeneous computing platforms with OpenCL, *Comput. Geosci.*, **100**, 142–155.
- Fichtner, A., 2011. *Full Seismic Waveform Modelling and Inversion*, Springer Science & Business Media.
- Fichtner, A., Bunge, H.P. & Igel, H., 2006. The adjoint method in seismology—I. Theory, *Phys. Earth planet. Inter.*, **157**(1–2), 86–104.
- Friedlander, M.P. & Schmidt, M., 2012. Hybrid deterministic-stochastic methods for data fitting, *SIAM J. Sci. Comput.*, **34**(3), A1380–A1405.
- Furse, C.M., 2000. Faster than fourier: ultra-efficient time-to-frequency-domain conversions for FDTD simulations, *IEEE Antennas Propag. Mag.*, **42**(6), 24–34.
- Kamei, R. & Pratt, R.G., 2013. Inversion strategies for visco-acoustic waveform inversion, *Geophys. J. Int.*, **194**(2), 859–884.
- Lei, X. & Xue, Z., 2009. Ultrasonic velocity and attenuation during CO<sub>2</sub> injection into water-saturated porous sandstone: measurements using difference seismic tomography, *Phys. Earth planet. Inter.*, **176**(3–4), 224–234.
- Levander, A.R., 1988. Fourth-order finite-difference *P*-*SV* seismograms, *Geophysics*, **53**(11), 1425–1436.
- Liao, O. & McMechan, G.A., 1995. 2.5D full-wavefield viscoacoustic inversion, *Geophys. Prospect.*, **43**(8), 1043–1059.
- Liao, Q. & McMechan, G.A., 1996. Multifrequency viscoacoustic modeling and inversion, *Geophysics*, **61**(5), 1371–1378.
- Lie, K.-A., Krogstad, S., Ligeard, I.S., Natvig, J.R., Nilsen, H.M. & Skaflestad, B., 2012. Open-source matlab implementation of consistent discretisations on complex grids, *Comput. Geosci.*, **16**(2), 297–322.
- Malo, M. & Bédard, K., 2012. Basin-scale assessment for CO<sub>2</sub> storage prospectivity in the Province of Québec, Canada, *Energy Procedia*, **23**, 487–494.
- Martens, S. *et al.*, 2012. Europe's longest-operating on-shore CO<sub>2</sub> storage site at Ketzin, Germany: a progress report after three years of injection, *Environ. Earth Sci.*, **67**(2), 323–334.
- Müller, T.M., Gurevich, B. & Lebedev, M., 2010. Seismic wave attenuation and dispersion resulting from wave-induced flow in porous rocks—a review, *Geophysics*, **75**(5), 75A147–75A164.
- Ngoc, T.T., Lefebvre, R., Konstantinovskaya, E. & Malo, M., 2014. Characterization of deep saline aquifers in the Bécancour area, St. Lawrence Lowlands, Québec, Canada: implications for CO<sub>2</sub> geological storage, *Environ. Earth Sci.*, **72**(1), 119–146.
- Onishi, K., Ueyama, T., Matsuoka, T., Nobuoka, D., Saito, H., Azuma, H. & Xue, Z., 2009. Application of crosswell seismic tomography using difference analysis with data normalization to monitor CO<sub>2</sub> flooding in an aquifer, *Int. J. Greenhouse Gas Control*, **3**(3), 311–321.
- Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L. & Virieux, J., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: from theory to practice, *Leading Edge*, **32**(9), 1040–1054.
- Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**(2), 495–503.
- Robertsson, J.O.A., Blanch, J.O. & Symes, W.W., 1994. Viscoelastic finite-difference modeling, *Geophysics*, **59**(9), 1444–1456.
- Song, Z., Williamson, P.R. & Pratt, R.G., 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.
- Spetzler, J., Xue, Z., Saito, H. & Nishizawa, O., 2008. Case story: time-lapse seismic crosswell monitoring of CO<sub>2</sub> injected in an onshore sandstone aquifer, *Geophys. J. Int.*, **172**(1), 214–225.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms including elasticity and attenuation, *Pure appl. Geophys.*, **128**(1–2), 365–399.
- Toksöz, M.N. & Johnston, D.H., 1981. *Seismic Wave Attenuation*, Society of Exploration Geophysicists.
- Tromp, J., Tape, C. & Liu, Q., 2004. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**(1), 195–216.
- van Leeuwen, T. & Herrmann, F.J., 2013. Fast waveform inversion without source-encoding, *Geophys. Prospect.*, **61**, 10–19.
- Virieux, J., 1986. *P*-*SV* wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, **51**(4), 889–901.
- White, J., 1975. Computed seismic speeds and attenuation in rocks with partial gas saturation, *Geophysics*, **40**(2), 224–232.
- Yang, P., Brossier, R., Métivier, L. & Virieux, J., 2016. A review on the systematic formulation of 3-D multiparameter full waveform inversion in viscoelastic medium, *Geophys. J. Int.*, **207**, 129–149.
- Zhang, F., Juhlin, C., Cosma, C., Tryggvason, A. & Pratt, R.G., 2012. Crosswell seismic waveform tomography for monitoring CO<sub>2</sub> injection: a case study from the Ketzin site, Germany, *Geophys. J. Int.*, **189**(1), 629–646.

## APPENDIX A: GRADIENT MISFIT EXPRESSION

This appendix presents the explicit expression for the misfit gradient for  $\rho$ ,  $M$ ,  $\mu$ ,  $\tau_p$  and  $\tau_s$ . Any other parametrization can be obtained with the chain rule. We give the expression for 2-D and 3-D, with  $N$  the number of dimensions. For 2-D, variables containing a subscript of one of the three spatial dimensions should be set to zero.

$$\frac{\partial J}{\partial \rho} = \left\langle \overleftarrow{v}_x, \partial_t v_x \right\rangle + \left\langle \overleftarrow{v}_y, \partial_t v_y \right\rangle + \left\langle \overleftarrow{v}_z, \partial_t v_z \right\rangle, \quad (\text{A1a})$$

$$\frac{\partial J}{\partial M} = -c_1^M P_1 + c_2^M P_2, \quad (\text{A1b})$$

$$\frac{\partial J}{\partial \tau_p} = -c_1^{\tau_p} P_1 + c_2^{\tau_p} P_2, \quad (\text{A1c})$$

$$\frac{\partial J}{\partial \mu} = -c_1^\mu P_3 + c_2^\mu P_1 - c_3^\mu P_4 + c_4^\mu P_5 - c_5^\mu P_2 + c_6^\mu P_6, \quad (\text{A1d})$$

$$\frac{\partial J}{\partial \tau_s} = -c_1^{\tau_s} P_3 + c_2^{\tau_s} P_1 - c_3^{\tau_s} P_4 + c_4^{\tau_s} P_5 - c_5^{\tau_s} P_2 + c_6^{\tau_s} P_6. \quad (\text{A1e})$$

with the scalar products:

$$P_1 = \left\langle \overleftarrow{\sigma}'_{xx} + \overleftarrow{\sigma}'_{yy} + \overleftarrow{\sigma}'_{zz}, \partial_t (\sigma'_{xx} + \sigma'_{yy} + \sigma'_{zz}) \right\rangle \quad (\text{A2a})$$

$$P_2 = \left\langle \overleftarrow{R}_{xxl} + \overleftarrow{R}_{yyt} + \overleftarrow{R}_{zzt}, (1 + \tau_{ol} \partial_t) (r_{xxl} + r_{yyt} + r_{zzt}) \right\rangle, \quad (\text{A2b})$$

$$P_3 = \left\langle \overleftarrow{\sigma}'_{xy}, \partial_t \sigma'_{xy} \right\rangle + \left\langle \overleftarrow{\sigma}'_{xz}, \partial_t \sigma'_{xz} \right\rangle + \left\langle \overleftarrow{\sigma}'_{yz}, \partial_t \sigma'_{yz} \right\rangle \quad (\text{A2c})$$



$$P_4 = \left\langle \overleftarrow{\sigma}'_{xx}, \partial_t((N-1)\sigma'_{xx} - \sigma'_{yy} - \sigma'_{zz}) \right\rangle + \left\langle \overleftarrow{\sigma}'_{yy}, \partial_t((N-1)\sigma'_{yy} - \sigma'_{xx} - \sigma'_{zz}) \right\rangle + \left\langle \overleftarrow{\sigma}'_{zz}, \partial_t((N-1)\sigma'_{zz} - \sigma'_{xx} - \sigma'_{yy}) \right\rangle \quad (A2d)$$

$$P_5 = \left\langle \overleftarrow{R}_{xyl}, (1 + \tau_{\sigma l} \partial_t) r_{xyl} \right\rangle + \left\langle \overleftarrow{R}_{xzl}, (1 + \tau_{\sigma l} \partial_t) r_{xzl} \right\rangle + \left\langle \overleftarrow{R}_{yzl}, (1 + \tau_{\sigma l} \partial_t) r_{yzl} \right\rangle \quad (A2e)$$

$$P_6 = \left\langle \overleftarrow{R}_{xxl}, (1 + \tau_{\sigma l} \partial_t)((N-1)r_{xxl} - r_{yy l} - r_{zz l}) \right\rangle + \left\langle \overleftarrow{R}_{yy l}, (1 + \tau_{\sigma l} \partial_t)((N-1)r_{yy l} - r_{xx l} - r_{zz l}) \right\rangle + \left\langle \overleftarrow{R}_{zz l}, (1 + \tau_{\sigma l} \partial_t)((N-1)r_{zz l} - r_{xx l} - r_{yy l}) \right\rangle. \quad (A2f)$$

With  $b_1$  and  $b_2$  defined as:

$$b_1 = \left( NM \frac{(1 + L\tau_p)}{(1 + \alpha\tau_p)} - 2(N-1)\mu \frac{(1 + L\tau_s)}{(1 + \alpha\tau_s)} \right)^{-1}, \quad (A3a)$$

$$b_2 = \left( NM \frac{\tau_p}{(1 + \alpha\tau_p)} - 2(N-1)\mu \frac{\tau_s}{(1 + \alpha\tau_s)} \right)^{-1}, \quad (A3b)$$

the coefficients for the misfit gradient are given by:

$$c_1^M = \frac{(1 + L\tau_p)}{(1 + \alpha\tau_p)} b_1^2, \quad (A4a)$$

$$c_2^M = \frac{\tau_p}{(1 + \alpha\tau_p)} b_2^2, \quad (A4b)$$

$$c_1^\mu = \frac{1}{\mu^2} \frac{(1 + \alpha\tau_s)}{(1 + L\tau_s)}, \quad (A4c)$$

$$c_2^\mu = \frac{N+1}{3} \frac{(1 + L\tau_s)}{(1 + \alpha\tau_s)} b_1^2, \quad (A4d)$$

$$c_3^\mu = \frac{1}{2N} \frac{1}{\mu^2} \frac{(1 + \alpha\tau_s)}{(1 + L\tau_s)}, \quad (A4e)$$

$$c_4^\mu = \frac{1}{\mu^2} \frac{(1 + \alpha\tau_s)}{\tau_s}, \quad (A4f)$$

$$c_5^\mu = \frac{N+1}{3} \frac{\tau_s}{(1 + \alpha\tau_s)} b_2^2, \quad (A4g)$$

$$c_6^\mu = \frac{1}{2N} \frac{1}{\mu^2} \frac{(1 + \alpha\tau_s)}{\tau_s}, \quad (A4h)$$

$$c_1^{\tau p} = (L - \alpha) \frac{M}{(1 + \alpha\tau_p)^2} b_1^2, \quad (A4i)$$

$$c_2^{\tau p} = \frac{M}{(1 + \alpha\tau_p)^2} b_2^2, \quad (A4j)$$

$$c_1^{\tau s} = \frac{1}{\mu} \frac{(L - \alpha)}{(1 + L\tau_s)^2}, \quad (A4k)$$

$$c_2^{\tau s} = \frac{N+1}{3} (L - \alpha) \frac{\mu}{(1 + \alpha\tau_s)^2} b_1^2, \quad (A4l)$$

$$c_3^{\tau s} = \frac{L - \alpha}{2N\mu(1 + L\tau_s)^2}, \quad (A4m)$$

$$c_4^{\tau s} = \frac{1}{\mu\tau_s^2}, \quad (A4n)$$

$$c_5^{\tau s} = \frac{N+1}{3} \frac{\mu}{(1 + \alpha\tau_s)^2} b_2^2, \quad (A4o)$$

$$c_6^{\tau s} = \frac{1}{2N\mu\tau_s^2}. \quad (A4p)$$