

# Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation

Wei Zhou,<sup>1</sup> Romain Brossier,<sup>1</sup> Stéphane Operto<sup>2</sup> and Jean Virieux<sup>1</sup>

<sup>1</sup>*ISTerre, Université Grenoble Alpes - BP 53, F-38041 Grenoble Cedex 09, France. E-mail: wei.zhou@ujf-grenoble.fr*

<sup>2</sup>*Géoazur, Université de Nice–Sophia Antipolis - CNRS, 250 rue Albert Einstein, F-06560 Valbonne, France*

Accepted 2015 May 28. Received 2015 May 27; in original form 2014 December 19

## SUMMARY

Full waveform inversion (FWI) aims to reconstruct high-resolution subsurface models from the full wavefield, which includes diving waves, post-critical reflections and short-spread reflections. Most successful applications of FWI are driven by the information carried by diving waves and post-critical reflections to build the long-to-intermediate wavelengths of the velocity structure. Alternative approaches, referred to as reflection waveform inversion (RWI), have been recently revisited to retrieve these long-to-intermediate wavelengths from short-spread reflections by using some prior knowledge of the reflectivity and a scale separation between the velocity macromodel and the reflectivity. This study presents a unified formalism of FWI, named as Joint FWI, whose aim is to efficiently combine the diving and reflected waves for velocity model building. The two key ingredients of Joint FWI are, on the data side, the explicit separation between the short-spread reflections and the wide-angle arrivals and, on the model side, the scale separation between the velocity macromodel and the short-scale impedance model. The velocity model and the impedance model are updated in an alternate way by Joint FWI and waveform inversion of the reflection data (least-squares migration), respectively. Starting from a crude velocity model, Joint FWI is applied to the streamer seismic data computed in the synthetic Valhall model. While the conventional FWI is stuck into a local minimum due to cycle skipping, Joint FWI succeeds in building a reliable velocity macromodel. Compared with RWI, the use of diving waves in Joint FWI improves the reconstruction of shallow velocities, which translates into an improved imaging at deeper depths. The smooth velocity model built by Joint FWI can be subsequently used as a reliable initial model for conventional FWI to increase the high-wavenumber content of the velocity model.

**Key words:** Inverse theory; Controlled source seismology; Body waves; Seismic tomography; Wave scattering and diffraction; Acoustic properties.

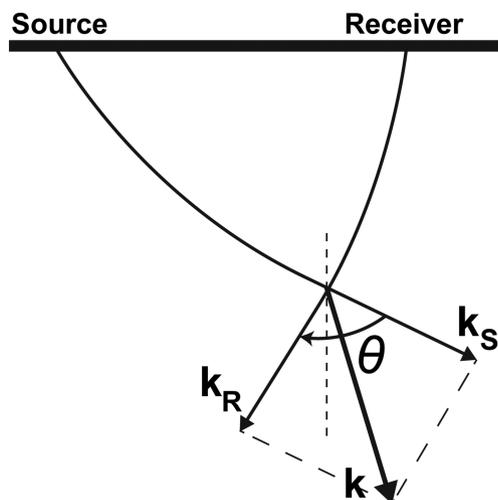
## 1 INTRODUCTION

With the emergence of long-offset wide-azimuth acquisitions and broad-band sources, full waveform inversion (FWI) has been recognized as an efficient tool for velocity model building (Virieux & Operto 2009, for a review). In these long-offset experiments, FWI is mainly driven by the information carried by diving waves and post-critical reflections to build the long-to-intermediate wavelengths of the velocity structure. The connection between the acquisition geometry and the spatial resolution of FWI has been clearly established in the theoretical framework of the generalized diffraction tomography (e.g. Devaney 1982; Miller *et al.* 1987). It has been shown that the wavenumber component  $\mathbf{k}$ , injected at a diffractor point in the subsurface is related to the local wavelength  $\lambda$  and the

scattering angle  $\theta$  by the relationship

$$\mathbf{k} = \frac{2}{\lambda} \cos\left(\frac{\theta}{2}\right) \mathbf{n}, \quad (1)$$

where  $\mathbf{n}$  is the normalization of the vector  $\mathbf{k}$  (Fig. 1). This relationship shows that the wide-scattering angles associated with diving waves and post-critical reflections contribute low-to-intermediate wavenumber updates to the subsurface. Conversely, the small scattering angles associated with short-spread reflections contribute high wavenumber updates to the subsurface. One key issue in the classical formulation of FWI (for short, referred to as FWI in the following) is that the penetration depths of diving waves are often insufficient to reach the deepest targeted structures, even from modern wide-azimuth surveys. At these depths, FWI behaves as a



**Figure 1.** Spatial resolution of diffraction tomography and its connection with acquisition geometry. The wavenumber vectors associated with the rays connecting the source and the receiver to the diffractor are denoted by  $\mathbf{k}_S$  and  $\mathbf{k}_R$ , respectively. The scattering angle is denoted by  $\theta$ . The wavenumber vector  $\mathbf{k} = \mathbf{k}_S + \mathbf{k}_R$  is the spectral component mapped at the diffractor point by the source–receiver pair during FWI.

least-squares migration of the short-spread reflections rather than as a tool for velocity model building, and would fail to update the low-to-intermediate wavenumbers of deep targets.

Alternatively, migration-based velocity analysis has been developed in the image domain to build the velocity macromodel by using the reflection data (Diaz *et al.* 2013; Liu *et al.* 2013; Allemant & Lambaré 2014). These approaches focus on the flattening of the common image gathers generated by migration (Symes & Carazzone 1991; Sava & Biondi 2004). Extended-domain approaches have also been proposed, which attempt to minimize the energy left in the non-physical dimensions added to the model space (Sava & Fomel 2006; Yang & Sava 2011; Almomin & Biondi 2012; Biondi & Almomin 2012; Sun & Symes 2012; Lameloise *et al.* 2015). The main issue of these approaches is their high computational cost that is made, on the one hand, by the migration step performed during each velocity update and, on the other hand, by the extended-domain imaging condition. Although 2-D applications have shown promising results, the cost of these approaches seems to prevent a direct extension to 3-D cases, in particular when the forward modelling is performed with the two-way wave equation.

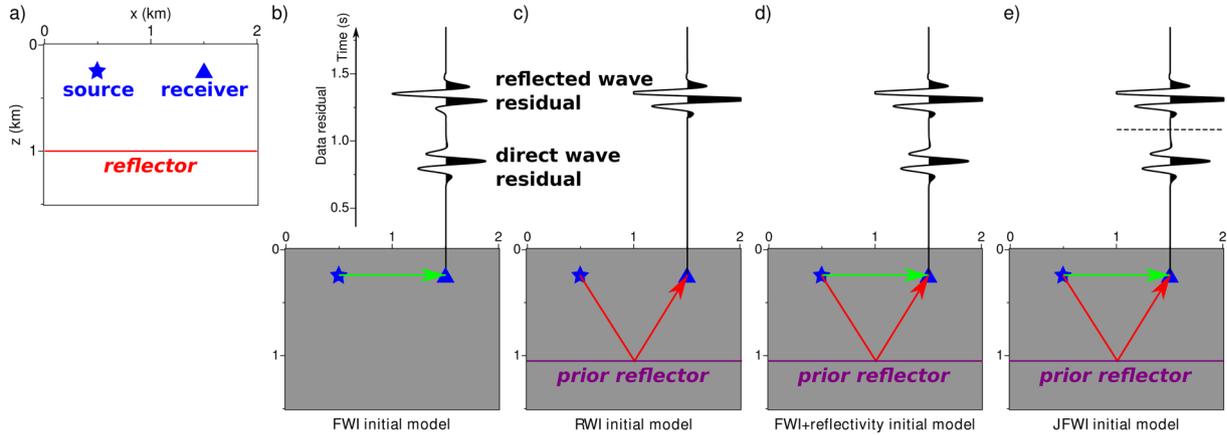
Inspired by the pioneering work of Chavent *et al.* (1994), Chavent (1996) and Clément *et al.* (2001) on migration-based traveltime inversion (MBTT), recent data-domain FWI strategies, referred to as reflection waveform inversion (RWI) in this paper, have been proposed as a new alternative to build the velocity macromodel from the reflection data (e.g. Xu *et al.* 2012; Ma & Hale 2013; Brossier *et al.* 2015). As most of the seismic reflection processing workflows, RWI relies on the explicit scale separation between a smooth velocity macromodel and a rough reflectivity. This scale separation results from the gap between wavenumber contents of the velocity macromodel built by reflection tomography or migration velocity analysis and the reflectivity built by migration (Claerbout 1985; Wu & Toksöz 1987; Jannane *et al.* 1989; Mora 1989). Such a separation leads to a two-step imaging workflow in which one repeatedly alternates the velocity model building assuming a known reflectivity and the reflectivity update by migration using the previous velocity update as the background model. More sophisticated approaches can be viewed to mitigate the computational

burden of this workflow, for example, by building the reflectivity in the pseudo-time domain (Plessix 2013) to avoid performing migration at each iteration of the velocity update (Brossier *et al.* 2015; Wang *et al.* 2015). In RWI, the governing idea behind the velocity model building task is to assume the reflectivity is known in prior, by which the reflected waves are predicted and the residuals are minimized for the velocity macromodel update. Under this assumption, high-wavenumber contributions such as migration isochrones are not present in the sensitivity kernels of RWI. Moreover, this prior reflectivity is used by RWI as the secondary sources to highlight in the sensitivity kernel the contribution of the transmission paths followed by the reflected waves. Indeed, the wide-scattering angles associated with these transmission regimes are amenable to update the low-to-intermediate wavenumbers of the subsurface located between the reflectors and the surface (eq. 1).

Brute-force approaches might be performed without explicit scale separation in the FWI formalism. They consist of applying the conventional FWI to build reflectivity during early iterations by a migration-like processing before updating the low-to-intermediate wavenumbers of the subsurface from the transmission paths of the reflected waves (AlTheyab *et al.* 2013). These approaches could be further developed for the velocity macromodel building or for the reflectivity imaging by a wavenumber-driven filtering of the gradient of the FWI misfit function (Alkhalifah 2014, 2015; Alkhalifah & Wu 2014; Wu & Alkhalifah 2014). Note that RWI can be implemented either in the time domain or in the frequency domain as shown by Wang *et al.* (2013b). Other data-domain approaches for velocity model building rely on a wavefield decomposition into upgoing and downgoing waves to separate the contribution of forward-scattering and backward-scattering in the sensitivity kernel of waveform inversion (Tang *et al.* 2013; Wang *et al.* 2013a).

One key limitation of RWI, that will be overcome in this study, is the exclusive reliance on the use of reflected waves, discarding the low-wavenumber information on the shallow targets that are carried by the diving waves.

In this study, we propose a new FWI method, referred to as joint full waveform inversion (JFWI), which integrates the ingredients of the conventional FWI and RWI into a unified formalism. Like RWI, JFWI still rests on the scale separation between the velocity macromodel and the reflectivity. The added value of JFWI compared with RWI, is the combination of the diving waves and the reflected waves such that the low-wavenumber information carried by these two wave modes, as described above, are simultaneously used in the velocity model building task. In JFWI, we regenerate the reflectivity by conventional FWI using only short-offset reflected waves (i.e. non-linear least-squares migration) at each iteration of the scale-separation workflow. A key feature of JFWI is to require the explicit separation of the early arriving phases (diving waves and post-critical reflections) and the pre-critical reflections in the data. Our approach shares some similarities with Wang *et al.* (2015), who also use refracted and reflected waves to update the low-to-intermediate wavenumbers of the subsurface. The main differences is that our workflow relies on the waveform-difference misfit function to update the velocity macromodel and the reflectivity, whereas the approach of Wang *et al.* (2015) relies on the cross-correlation based misfit function, the so-called wave-equation tomography (Luo & Schuster 1991), to perform the velocity macromodel update. Although our approach relies on the explicit separation between reflected waves and refracted waves, it does not require assigning a time window to isolate phases as it does in the workflow of Wang *et al.* (2015). This phase identification might be quite cumbersome for the reflection recordings in the complex environments. In



**Figure 2.** Illustration of the initial models and the data residuals used in different FWI approaches. (a) True velocity model including a reflector at 1 km depth. One source and receiver couple is indicated. (b) Homogeneous initial model ( $m$ ) without reflector and erroneous background velocity for FWI. Only the direct wave (green arrow) is generated. Residuals include the direct wave and the reflected wave residuals. (c) Homogeneous background ( $m_0$ ) and prior reflectivity ( $\delta m$ ) for RWI. Only the reflected wave (red arrow) is used. (d) Initial model with prior reflector for FWI. Both direct and reflected waves are modelled. (e) Homogeneous background ( $m_0$ ) and prior reflectivity ( $\delta m$ ) for JFWI. Compared with RWI (c), diving waves are used; compared with FWI (d), direct and reflected wave residuals are explicitly separated (see text for details).

addition, we use a velocity-impedance parametrization to alternate the update of the velocity macromodel and the reflectivity, unlike Wang *et al.* (2015) who choose the velocity-density parametrization.

This paper is organized as follows. First, we shall review the intrinsic limitation of FWI and RWI as well as the benefit that can be expected from JFWI with a simple synthetic model. Second, we shall discuss about the choice of a suitable subsurface parametrization to perform the velocity model and reflectivity updates, such that the scale separation between these two models is satisfied. This naturally directs us towards a subsurface parametrization in terms of wavespeed ( $V_P$ ) and impedance ( $I_P$ ). Third, we shall review the ingredients of the JFWI workflow that alternates the updates of  $V_P$  and  $I_P$  before showing its application to the synthetic Valhall case. The experimental setup of this case study is designed in a way such that the diving waves only sample the shallow part of the subsurface. We shall first show how the use of diving waves in JFWI improves the reconstruction of the shallow velocities compared with RWI, and how this shallow improvement translates into an improved imaging at greater depths. Then, we shall show that the smooth velocity model built by JFWI can be subsequently used as a reliable initial model for conventional FWI, resulting a broad-band velocity model that can be taken for the purpose of structural interpretation.

## 2 REVIEW OF FWI AND RWI

### 2.1 Full waveform inversion (FWI)

FWI is a data-fitting procedure during which the subsurface model  $m$  is iteratively updated in order to match the synthetic data,  $d = d(m)$  with the recorded data,  $d_{obs}$ . The misfit function is conventionally defined as the least-squares norm of the data residuals weighted by a linear operator  $\mathcal{W}$ , that is

$$C_{FWI}(m) = \frac{1}{2} \|\mathcal{W}(d_{obs} - d(m))\|_2^2, \quad (2)$$

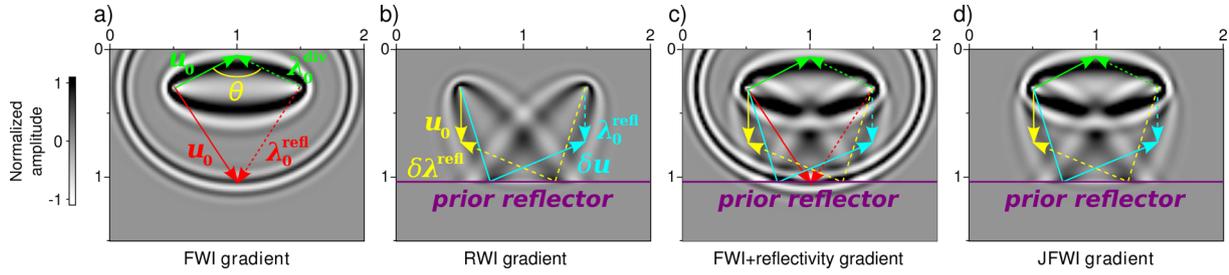
with an implicit summation over sources, receivers and time. As the wavefield does not linearly depend on the subsurface parameters, this optimization problem is highly non-linear. Many local optimization approaches have been proposed to mitigate this non-linearity and make the misfit function as convex as possible: they

might differ in the misfit definition and/or in the domain within which the minimization is performed, but all of them aim to update the model by a linear-search method (Shin & Cha 2008; van Leeuwen & Mulder 2010; Luo & Sava 2011; van Leeuwen & Herrmann 2013). The local descent direction relies on the gradient of the misfit function with respect to the model parameters, which can be efficiently computed by the adjoint-state method (for a review, see Plessix 2006). For one source–receiver pair, the gradient can be written in a compact form as

$$\nabla C_{FWI} = u_0 \star \lambda_0, \quad (3)$$

where the symbols  $u_0 = u_0(m)$  and  $\lambda_0 = \lambda_0(m)$  denote the incident wavefield and the back-propagated adjoint wavefield, respectively. For multiple sources and receivers, the gradient is a summation of  $u_0 \star \lambda_0$  over all sources and receivers. The adjoint wavefield is computed with a source term that gathers the data residuals associated with all kinds of waves (diving waves, reflected waves, scattered waves etc.). The gradient of the misfit function is computed by a zero-lag cross-correlation between the incident and adjoint fields. This correlation operation, denoted by a single star ( $\star$ ) in eq. (3), embeds, for the sake of compactness, the partial derivative of the forward modelling operator with respect to model parameters (the so-called diffraction pattern) that is cumbersome to be expressed in the time domain. Hiding this kind of complexity in the star symbol will not obscure the governing idea underlying the following derivation of the misfit function gradients. All details about the incident and adjoint fields as well as the gradient expressions are provided in Appendix A in the framework of frequency-domain FWI allowing for more compact notations. Note that all our implementations are performed in the time domain.

We first illustrate the FWI gradient (eq. 3) in the case of a homogeneous subsurface model in which a flat reflector is embedded (Fig. 2a). The background model is homogeneous with a wrong velocity and does not contain the reflector, so that the source of the adjoint equation contains the residuals of the direct wave and reflected wave (Fig. 2b). These residuals give rise to two components in the adjoint field denoted by  $\lambda_0^d$  and  $\lambda_0^r$ , respectively. Correlations of the incident wavefield  $u$  with these two components of the adjoint wavefield respectively build a wide first Fresnel zone and a secondary Fresnel zone (the so-called migration isochrone, Fig. 3a),



**Figure 3.** Illustration of gradients generated by different FWI approaches. Solid and dashed arrows denote the ray paths, respectively, followed by the incident and adjoint fields that interfere constructively. (a) FWI gradient combines a low-wavenumber first Fresnel zone (represented by  $u_0 \star \lambda_0^d$ ) between source and receiver, and a high-wavenumber migration isochrone (represented by  $u_0 \star \delta\lambda^r$ ). Note that the first Fresnel zone has a limited penetration in depth. (b) RWI gradient shows two wide first Fresnel zones centred on the two-way paths followed by the reflected wave between the reflector (behaving as a secondary source) and the source/receiver positions. (c) FWI gradient, with a prior reflectivity in the initial model, combines FWI (a) and RWI (b) gradients. Low-wavenumber and high-wavenumber information enter into the gradient, hence breaking down the scale-separation prerequisite. (d) JFWI gradient combines the first Fresnel zone generated by direct-wave residuals and RWI gradient. Compared with (c) the migration isochrone is not generated such that the scale separation is honoured.

over which the direct-wave residual and the reflection residual are, respectively, back-projected (Woodward 1992). The width of these iso-phase surfaces gives the spatial resolution with which a point diffractor located in these surfaces is imaged by the current source–receiver pair, according to eq. (1). A seismic acquisition generally provides the samplings of the shallow subsurface with both diving and reflected waves, allowing for a broad-band imaging of the shallow targets. In contrast, only short-spread reflections sample the subsurface at greater depths, that is, beyond the penetration depths of the diving waves, leading to a high-wavenumber imaging of deep targets.

## 2.2 Reflection waveform inversion (RWI)

RWI focuses on the reflection data to build the subsurface model. The method relies on the prior knowledge of the reflectivity to predict the reflected waves. The misfit function is given by

$$C_{RWI}(m_0) = \frac{1}{2} \left\| W^r \left( d_{obs}^{refl} - d_{pred}^{refl}(m_0, \delta m) \right) \right\|_2^2, \quad (4)$$

where the symbols  $d_{obs}^{refl}$  and  $d_{pred}^{refl}$  denote the observed and predicted reflected waves, respectively, weighted by the linear operator  $W^r$ . The low-wavenumber background  $m_0$  and the high-wavenumber reflectivity  $\delta m$  are separated in scales, and RWI seeks to reconstruct the background  $m_0$  only. Following Brossier *et al.* (2015), the gradient with respect to the background model is (for the counterpart in the frequency domain, see Appendix A),

$$\nabla C_{RWI} = u_0 \star \delta\lambda^r + \delta u \star \lambda_0^r + \delta u \star \delta\lambda^r, \quad (5)$$

where the symbols  $u_0 = u_0(m_0)$  and  $\delta u = \delta u(m_0, \delta m)$  denote the incident wavefield computed in  $m_0$  and the wavefield scattered by  $\delta m$ , respectively. Similarly, the adjoint wavefield can be decomposed as a component  $\lambda_0^r$  propagating in the background model and a scattered component  $\delta\lambda^r$ . The scattered wavefield includes forward and backward scatterings (transmissions versus reflections) of any order. Note that the three correlation operations in eq. (5) embed two different diffraction pattern operators computed in  $m_0$  and  $m_0 + \delta m$ , respectively. The detailed expressions are provided in Appendix A.

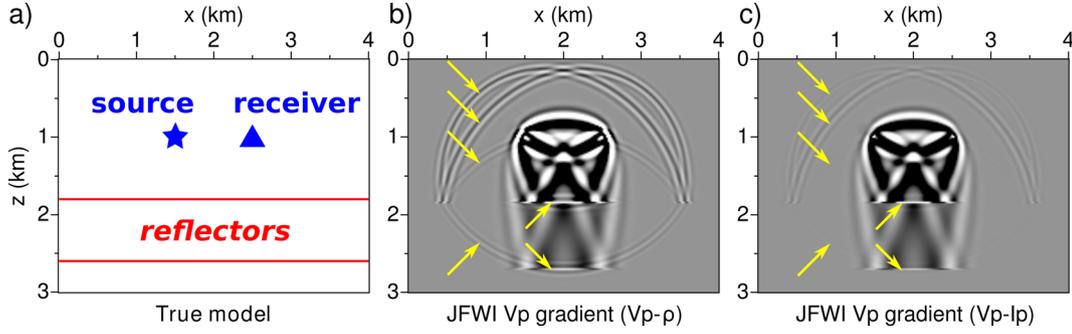
The RWI gradient corresponding to the one-reflector synthetic model is shown in Fig. 3(b). The source of the adjoint-state equation contains only the residual of the reflected wave (Fig. 2c). The first term  $u_0 \star \delta\lambda^r$  represents the correlation between the down-going field  $u_0$  and the up-going scattered field  $\delta\lambda^r$  (ray path indicated by

yellow arrows). This correlation builds a wide first Fresnel zone connecting the reflector and the source position. Similarly, the second term  $\delta u \star \lambda_0^r$  gives rise to a mirror Fresnel zone generated by the upgoing scattered field  $\delta u$  and the downgoing field  $\lambda_0^r$  (ray path indicated by cyan arrows). Known as ‘rabbit ears’, this pair of Fresnel zones allow for a long-wavelength reconstruction of the deep targets where FWI encounters difficulties. The third term represents higher-order migration isochrones. Due to the weak amplitude of the scattered fields  $\delta u$  and  $\delta\lambda^r$ , these higher-order isochrones are generally of small amplitudes and thus negligible. In Fig. 3(b), two mirror high-order migration isochrones are shown near the reflector position. They are built, on the one hand by the zero-lag correlation between the incident and adjoint fields transmitted across the reflector and, on the other hand by the zero-lag correlation between the incident and adjoint fields reflected from the reflector. The reader is referred to Appendix B for a more detailed description of these higher-order contributions.

The limitation of RWI is the reliance of the exclusive use of the reflected waves, discarding the low-wavenumber information carried by the diving waves (i.e.  $u_0 \star \lambda_0^d$  term, see analysis of conventional FWI gradient). One may simply insert the diving waves into the RWI misfit function, augmenting the reflection data residuals in eq. (4) with the diving wave residuals (i.e.  $d_{obs} - d_{pred}^{div}(m_0) - d_{pred}^{refl}(m_0, \delta m)$ ). This is equivalent to performing FWI with a prior reflectivity in the initial model (Fig. 2d). Although all low-wavenumber contributions are gathered in the gradient (Fig. 3c), the dominant imprint of the high-wavenumber information carried by the migration of the reflection residuals (i.e.  $u_0 \star \lambda_0^r$ , Fig. 3c, red arrows) makes challenging the extraction of the low-wavenumber information carried by the first Fresnel zones associated with RWI (Fig. 3b). This highlights the necessity to force a scale separation between the low-wavenumber and high-wavenumber components in the FWI formalism, and this is the aim of the following section where we propose an alternative FWI formulation that combines the diving waves and the reflected waves for the velocity macromodel building without generating the high-wavenumber isochrones.

## 3 JOINT FULL WAVEFORM INVERSION (JFWI)

We shall first review the main idea behind JFWI that will allow us to mitigate the high-wavenumber contributions during the velocity



**Figure 4.** Two-reflector model and the corresponding JFWI gradients. (a) Two-reflector model to generate multiscattered fields. (b) JFWI gradient for the  $V_P$ - $\rho$  parametrization. Higher-order isochrones are produced due to constructive interference of the multiscattered fields (Appendix B for a detailed review of all of these isochrones). (c) Same as (b) except that the  $V_P$  gradient was built for the  $V_P$ - $I_P$  parametrization. The low-wavenumber components of the gradient are kept unchanged, while the undesired high-wavenumber components were filtered out.

model building before discussing the implementation of the JFWI gradient.

### 3.1 Formulation

The governing idea of JFWI is to explicitly separate the contributions of the diving and reflected waves (Fig. 2e), which means that the misfit function is decomposed as the sum of two terms,

$$C_{JFWI}(m_0) = \frac{1}{2} \left\| W^d (d_{obs}^{div} - d_{pred}^{div}(m_0)) \right\|_2^2 + \frac{1}{2} \left\| W^r (d_{obs}^{refl} - d_{pred}^{refl}(m_0, \delta m)) \right\|_2^2, \quad (6)$$

where the symbols  $d_{obs}^{div}$  and  $d_{pred}^{div}(m_0)$  denote the observed and predicted diving waves, and  $W^d$  and  $W^r$  denote the weighting operator that are applied to the diving and reflected waves, respectively. The role of the two weighting operators is to balance the respective contributions of diving and reflected residuals in the misfit function, considering that reflected wavefields have generally weaker amplitudes than the diving waves. The operator  $W^r$  can also embed a time-dependent weighting of reflection residuals to enhance the contribution of late reflected arrivals at the expense of early ones in the misfit function. This inversion preconditioning can contribute to balance the amplitudes of the shallow and deep perturbations in the gradient of the misfit function, and hence improve the convergence rate. Other weightings can also be added in the  $W^d$  and  $W^r$  operators such as offset-dependent weighting, which can be useful to design layer-stripping strategies. These issues will be illustrated in the synthetic example presented in the sequel of this study.

Both the observed data and predicted data should be decomposed into the diving part and the reflection part during the data pre-processing stage. Many pre-processing tools, commonly used to pre-process the reflection data before migration, can be applied here, such as the dynamic time windowing or F-K filtering. While the observed data are pre-processed once and for all before the inversion starts, the decomposition of the predicted data, however, has to be performed at each iteration of JFWI. This might be cumbersome if the decomposition should be refined according to the velocity model update. For the synthetic example shown in the following of this study, we did not need to perform this refinement: we applied to the modelled data, at each JFWI iteration, the same decomposition rule based on the offset-dependent time window as is applied to the recorded data. We could afford to keep the same windowing law at each JFWI iteration because the initial model for JFWI is already

accurate enough to predict the first-arrival traveltimes within half the dominant period. If the separation between the diving waves and the reflected waves is awkward, a systematic separation procedure would consist in performing two forward modellings, one in  $m_0$  and one in  $m_0 + \delta m$ . The first simulation provides the diving wavefield, while the subtraction between the two simulated wavefields provides the reflection wavefield.

The gradient of the misfit function with respect to the background model  $m_0$  is given by (for the counterpart in the frequency domain see Appendix A)

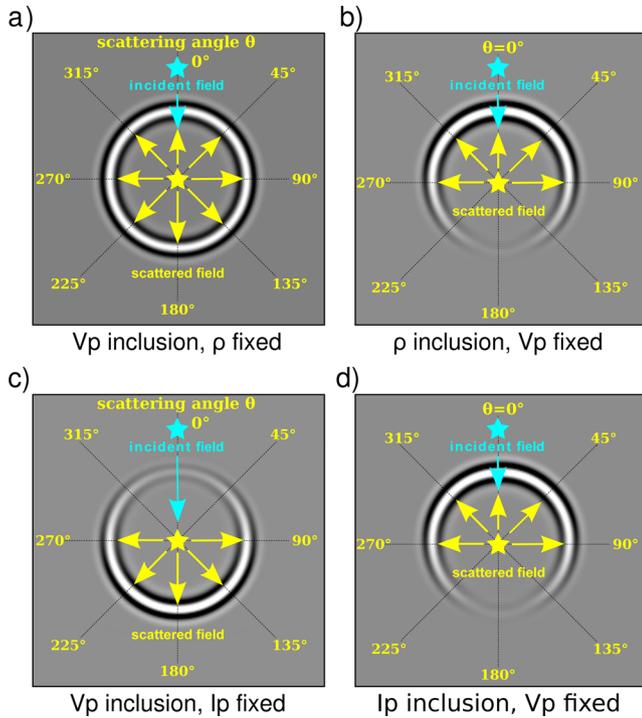
$$\nabla C_{JFWI} = u_0 \star \lambda_0^d + u_0 \star \delta \lambda^r + \delta u \star \lambda_0^r + \delta u \star \delta \lambda^r, \quad (7)$$

where the symbols  $\lambda_0^d$  and  $\lambda_0^r$  denote the background components of the adjoint field generated by the diving-wave residuals and the reflection residuals, and  $\delta \lambda^r$  denotes the scattered component of the adjoint wavefield generated by the reflection residuals, respectively. The first term in eq. (7) builds the first Fresnel zone associated with the diving waves, while the second and third terms are those generated during RWI. The key point is that the gradient in eq. (7) does not include the  $u_0 \star \lambda_0^r$  term associated with the high-wavenumber migration isochrone (Fig. 3d).

The last three terms in eq. (7) encapsulate all of the high-order scattering propagation. Among them, higher-order migration isochrones can be generated and hence, inject undesired high-wavenumber components into the gradient. A two-reflector model, in which internal multiples are generated, is used to illustrate the imprint of these high-order migration isochrones on the gradient (Figs 4a and b). A detailed review of the different high-order contributions is provided in Appendix B. Due to their inconsistent spatial locations, most of these high-order migration isochrones interfere with each other in a destructive way when the contributions from multiple shot gathers are stacked, and thus the gradient tends to be free from these high-wavenumber components. Moreover, in the next section, we shall show how to further reduce their footprint by choosing a proper subsurface parametrization for JFWI.

### 3.2 Mitigation of high-order isochrones by choosing suitable subsurface parametrization

In most acoustic FWI approaches, the subsurface is parametrized by the velocity ( $V_P$ ) and the density ( $\rho$ ) (e.g. Wang *et al.* 2015). The diffraction patterns associated with these parameters are shown in Figs 5(a) and (b), respectively. They show two wavefields scattered by a single  $V_P$  and  $\rho$  point perturbation in a homogeneous



**Figure 5.** Diffraction patterns of  $V_{P-\rho}$  (a) and  $V_{P-I_P}$  (b) parametrizations. Single diffractors are located in the centre of the homogeneous background. A source at the vertex of the diffractor generates the scattered field, whose amplitude variation with the scattering angle is representative of the diffraction pattern of the parameter. Note how the diffraction pattern of  $V_P$  changes depending on the other parameter involved in the parametrization ( $\rho$  versus  $I_P$ ). See text for details.

background. These wavefields represent the partial derivatives of the incident wavefield with respect to the model parameters that undergo a perturbation. Zero-lag correlations of this kind of partial derivative wavefields at the receiver positions with the data residuals form the FWI gradients. The amplitude variation of the partial derivative wavefield with the scattering angle  $\theta$  (i.e. the diffraction pattern) controls the effective range of scattering angles who confers the spatial resolution to the FWI gradient for one parameter class [see Operto *et al.* (2013) for a recent tutorial on multiparameter FWI]. The wavefield scattered by the  $V_P$  diffractor shows an isotropic radiation pattern, conferring a broad-band wavenumber content to the  $V_P$  gradient. On the other hand, the wavefield scattered by the  $\rho$  diffractor has significant amplitudes at small-to-intermediate  $\theta$ , conferring a narrower (high) wavenumber content to the  $\rho$  gradient. This  $V_{P-\rho}$  parametrization is a natural choice for FWI to reconstruct a broad-band  $V_P$  model, keeping in mind that cross-talks or leakage between  $V_P$  and  $\rho$  are necessarily present in the high-wavenumber part of the  $V_P$  and  $\rho$  gradients. A careful accounting for the Hessian should help remove this leakage in the FWI subsurface models.

In contrast, the  $V_{P-I_P}$  parametrization leads to a natural scale separation between the two parameter classes. The corresponding diffraction patterns are shown in Figs 5(c) and (d). The wavefield scattered by the  $V_P$  diffractor has significant amplitudes for large  $\theta$ , leading to a  $V_P$  gradient with a low-wavenumber content. Conversely, the wavefield scattered by the  $I_P$  diffractor has significant amplitudes for small  $\theta$ , leading to an  $I_P$  gradient with a high-wavenumber content. Compared with the  $V_{P-\rho}$  parametrization, the  $V_{P-I_P}$  parametrization is more suitable to satisfy the scale-

separation condition underlying JFWI, although it will prevent the reconstruction of the high wavenumbers in the  $V_P$  model.

According to this diffraction-pattern analysis, we choose the  $V_{P-I_P}$  parametrization to perform JFWI, where the low-wavenumber part of the subsurface  $m_0$  is parametrized by  $V_P$  and the high-wavenumber part  $\delta m$  is parametrized by  $I_P$  (eq. 6). The effect of the subsurface parametrization on the JFWI gradient is illustrated in Fig. 4(c) by the effective attenuation of the high-order isochrones when the  $V_{P-I_P}$  parametrization is used. [Another advantage of the  $V_{P-I_P}$  parametrization over the  $V_{P-\rho}$  parametrization is described in Snieder *et al.* 1989, their figs 2 and 3.]

### 3.3 Implementation of $\nabla C_{JFWI}$

In order to perform a computationally efficient implementation of JFWI, we rewrite eq. (7) in a more compact form by regrouping scattered and background wavefields, which is

$$\nabla C_{JFWI} = u_0 \star \lambda_0^r + \delta u \star \lambda_0^r + u_0 \star \delta \lambda^r + \delta u \star \delta \lambda^r + u_0 \star \lambda_0^d - u_0 \star \lambda_0^r \quad (8)$$

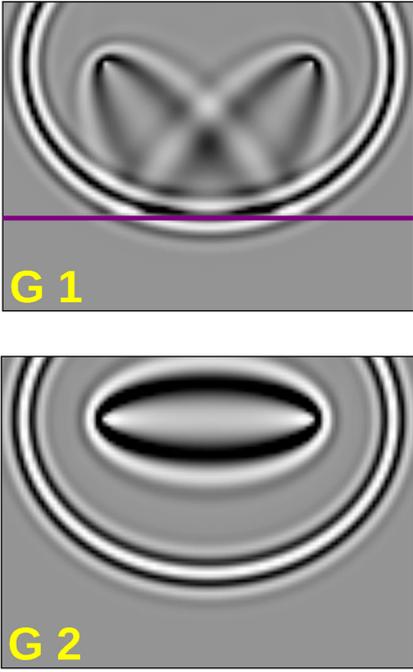
$$= \underbrace{u^r \star \lambda^r}_{G_1} + \underbrace{u_0 \star (\lambda_0^d - \lambda_0^r)}_{G_2}, \quad (9)$$

where we have (re-)introduced the total reflection field  $u^r = u_0^r + \delta u^r$  and the total reflection adjoint wavefield  $\lambda^r = \lambda_0^r + \delta \lambda^r$ . Note that eq. (8) is a true identity even though we have concealed the partial derivatives of the modelling operator inside the symbol  $\star$ . Strict demonstration is provided by the developments starting from eq. (A10) to eq. (A14) in Appendix A.

This alternative expression of the gradient leads to a workflow that sequentially computes two quantities  $G_1$  and  $G_2$  (Algorithm 1). The quantity represented by  $G_1$  is computed in  $m_0 + \delta m$  (Fig. 3c) by conventional FWI of the reflection data (hence, the first Fresnel zone associated with the diving waves represented by  $u_0 \star \lambda_0^r$  are not embedded in this expression), while the quantity represented by  $G_2$  is computed in  $m_0$  by conventional FWI of the diving waves and reflection data (Fig. 3a), in which the sign of the reflection data residuals is reversed. This reversal, indicated by the minus sign in front of  $\lambda_0^r$  in eq. (9), cancels by subtraction the first-order migration isochrone involved in  $G_1$ , making the total gradient  $G_1 + G_2$  dominated by low-wavenumber components. Moreover, computing  $G_1$  or  $G_2$  requires the same tasks as those performed during FWI to evaluate the FWI gradient in the time domain [the boundary-saving strategy is used to alleviate the memory load (Clapp 2008)]. These tasks can be summarized as follows: simulate the modelled wavefield and save the values on the model boundaries, compute the data residuals at receiver positions, compute the adjoint field by taking the residuals as the source function, and re-simulate the modelled wavefield in reverse time by using the boundary values as the Dirichlet condition, cross-correlate on the fly the two fields. Therefore, for one evaluation of the FWI gradient, we need to do three modellings and save two fields at maximum. As a combination of  $G_1$  and  $G_2$ , Algorithm 1 requires to perform six modellings and to store two fields for the cross-correlation. Consequently, the workflow to evaluate the JFWI gradient has a time complexity two times higher than the one of conventional FWI but the memory requirement is the same (see Table 1). In addition, the cost of JFWI is the same as the cost of RWI.

If more subtle data separations are used, that is, involving the decomposition of the modelled wavefield on the fly by subtraction of the full field and the background field, the implementation of  $G_1$

**Algorithm 1.** Efficient evaluation of  $\nabla C_{JFWI}$ .

<p><b>Step 1</b></p> <p>USE prior reflectivity</p> <p>DO forward modelling: <math>u = u_0 + \delta u</math></p> <p>USE reflected wave residuals</p> <p>DO backpropagation and correlation</p> <p>GET <math>G_1 =</math> rabbit ears + 1<sup>st</sup>-order isochrone</p> <p><b>Step 2</b></p> <p>USE smooth models (NO reflectivity)</p> <p>DO forward modelling: <math>u_0</math></p> <p>USE diving wave residuals</p> <p>USE same reflected wave residuals as in Step 1 but with a “-” sign</p> <p>DO backpropagation and correlation</p> <p>GET <math>G_2 =</math> diving-wave first Fresnel zone - 1<sup>st</sup>-order isochrone</p> <p><b>Step 3</b></p> <p>DO <math>G_1 + G_2 = \nabla C_{JFWI}</math></p>	
---	--

**Table 1.** Cost comparison.

Cost	Conv. FWI	RWI	JFWI
Modellings	3	6	6
Storage	2	2	2

in eq. (9) is impossible. This is because  $G_1$  is only related to the predicted reflection data, which would not be readily available at the time of the adjoint simulation when the data separation is performed on the fly with the aforementioned subtraction procedure. In this case, we would recommend computing an approximation of  $\nabla C_{JFWI}$  with the following workflow: (1) Compute the conventional FWI gradient in the model  $m_0 + \delta m$  using both the diving and reflected wave residuals (Fig. 3c). (2) Build the first-order migration isochrones by conventional FWI performed in model  $m_0$  using the reflection residuals as the source of the adjoint equation. At this stage, the predicted reflection data become available since the subtraction of the modelled full field and the background field can be readily performed. (3) Subtraction of the two quantities gives a low-wavenumber quantity, which is a good approximation of  $\nabla C_{JFWI}$ . A relative error of 2 per cent is found in the synthetic Valhall case study presented later. This implementation requires the same resources as the one discussed above (Table 1, third column).

#### 4 MULTIPARAMETER FWI FOR VELOCITY MODEL BUILDING AND IMPEDANCE IMAGING

So far we have derived the principles of JFWI for low-wavenumber velocity building, and choose the  $V_p$ - $I_p$  parametrization for high-wavenumber isochrones mitigation. In the following, we shall design a complementary imaging tool that provides a high-

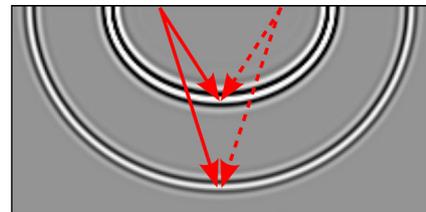
wavenumber model of the subsurface, and combine it with JFWI to have an integrated inversion workflow.

##### 4.1 $I_p$ imaging by waveform inversion of short-offset reflection data (IpWI)

We perform the high-wavenumber imaging by waveform inversion using only the short-offset reflection data (referred to as IpWI in the following). According to the previous diffraction-pattern analysis, the subsurface is parametrized by  $V_p$ - $I_p$  and only the  $I_p$  parameter is updated using the  $V_p$  model as the background model. The corresponding misfit function is given by

$$C(I_p) = \frac{1}{2} \left\| W^r \left( d_{obs}^{refl} - d_{pred}^{refl}(V_p, I_p) \right) \right\|_2^2. \quad (10)$$

The corresponding gradient (Fig. 6) contains the first-order isochrones similar to the ones which would have been imaged by a



**Figure 6.** Impedance gradient resulting from the migration of one residual seismogram in a homogeneous background. The true model contains two reflectors. The gradient only contains first-order isochrones. Ray paths of the incident and adjoint fields for the source-receiver pair are illustrated.

migration processing

$$\nabla C = u_0 \star \lambda_0^r = u_0 \star B^{\dagger^{-1}}(d_{obs}^{refl} - d_{pred}^{refl}), \quad (11)$$

where the forward modelling operator is denoted by the symbol  $B$ .

The only notable difference from a least-squares migration is that the observed reflection data is used as the source function to compute the backpropagated field in the migration process, whereas the source function of the adjoint-state equation in this IpWI is the reflection data residual.

#### 4.2 Cycle workflow of $V_P$ - $I_P$ imaging

Recall that the  $V_P$ - $I_P$  parametrization naturally leads to the scale separation between the two parameter classes. The low wavenumbers of  $V_P$  are reconstructed by JFWI, which requires a prior high-wavenumber  $I_P$  model built by IpWI. Conversely, the impedance imaging requires a background  $V_P$  that can be provided by JFWI. Therefore, it is natural to combine JFWI and IpWI to reconstruct these two parameters (a similar strategy can be found in Ma & Hale 2013). A critical issue is that, once the  $V_P$  model has been modified by JFWI to some extent, a new  $I_P$  reflectivity model should be generated accordingly in order to match the reflection data before continuing the  $V_P$  update, and the initial smooth impedance model should be used as the starting model for this new IpWI implementation (i.e. removing the imaged reflectivity; Brossier *et al.* 2015). This directs us towards a cycle workflow in which the  $V_P$  model and the  $I_P$  model are repeatedly updated in an alternate way by several non-linear iterations of JFWI and IpWI (Algorithm 2), and during each JFWI and IpWI step the two parameters are considered independently from each other. As the velocity model is not expected to be accurate during the cycle workflow, we build the impedance model from very short-offset reflections to enhance the focusing of the reflectivity image, even if mispositioned in depth, in order to avoid any biases associated to the residual moveout in the depth-migrated domain.

## 5 SYNTHETIC EXAMPLE: VALHALL CASE STUDY

### 5.1 Experimental setup

We apply our cycle workflow on a synthetic case representative of the Valhall oil field. The true  $I_P$  and  $V_P$  models are shown in Figs 7(a) and (c), respectively. The seafloor is at 62.5 m depth, above the sediment layers that overlay several low-velocity gas zones depicted in red. From 2.5 km to 3.3 km depths, the high-velocity oil reservoir is separated from the low-velocity gas zones by a cap rock of anticline structure. The sand is laid below the reservoir with smooth  $V_P$ ,  $I_P$  variations, supported by the bedrock at 5 km depth. The model, which is 8.8 km in width and 5.2 km in depth, is discretized by  $418 \times 704$  gridpoints with a grid interval of 12.5 m.

We use a Gaussian filter to smooth the true velocity model (excluding the water layer), and then extract one vertical profile to build the 1-D initial model for JFWI and IpWI (Figs 7b and d). This initial velocity model captures the large-scale variation of the true model and discards all features of the gas zones. Therefore, the main task of JFWI is to reconstruct the gas zones. Based on a former analysis (Prioux *et al.* 2011, their fig. 2b), more than 14 km of offset would be needed to record diving waves that propagate at reservoir depths. In this study, the maximum offset is set to 6 km and the diving waves reach a maximum penetration-depth of 1.5 km, which is shallower than most of the gas layers. Therefore, only the reflected waves can contribute to the reconstruction of the deep targets. The initial impedance model is shown in Fig. 7(b), which was built by taking a two-layer density model (setting  $1000 \text{ kg m}^{-3}$  for the water and  $2000 \text{ kg cm}^{-3}$  below).

The forward problem is solved by a classical  $\mathcal{O}(\Delta t^2, \Delta x^4)$  staggered-grid finite-difference method. The absorbing boundary condition implemented with perfectly matched layers (PMLs, e.g. Bérenger 1994; Komatitsch & Martin 2007) is applied along each edge of the model. Therefore, no surface-related multiples are generated in the data. We exclude this kind of multiples for a purpose

**Algorithm 2.** Cycle workflow of  $V_P$ - $I_P$  imaging.

---

Smooth initial  $V_P^{(0)}$  and  $I_P^{(0)}$ ,  $k = 1$

**repeat**

**USE**  $V_P^{(k-1)}$ ,  $I_P^{(0)}$  (NO reflectivity)

**USE** short offsets

**DO**  $I_P$  inversion by IpWI of reflection data

**GET** high-wavenumber  $I_P^{(k)}$  (built reflectivity)

**USE**  $V_P^{(k-1)}$ ,  $I_P^{(k)}$

**USE** full offsets

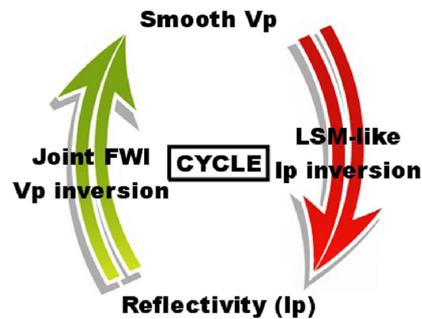
**DO**  $V_P$  inversion by JFWI

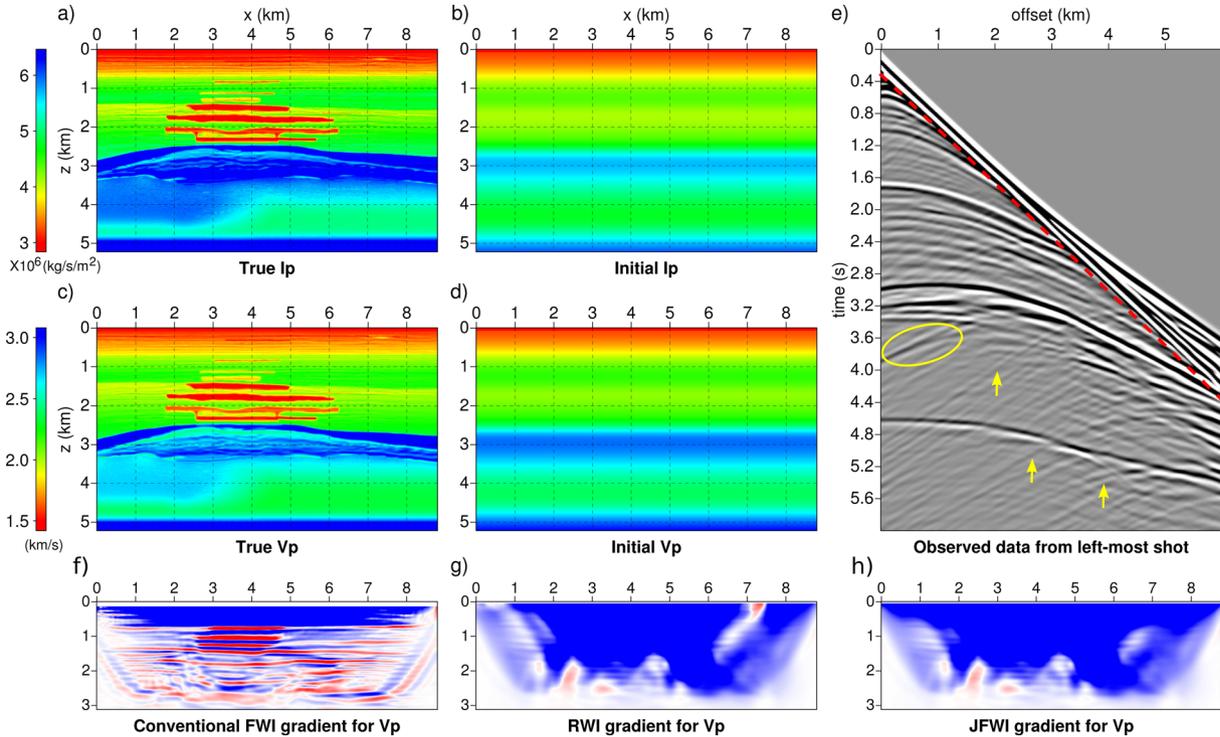
**GET** low-wavenumber  $V_P^{(k)}$  (updated background)

    Cycle number  $k = k + 1$

**until** convergence

---





**Figure 7.** Synthetic Valhall case study. (a) True  $I_P$  model. (b) Initial  $I_P$  model. (c) True  $V_P$  model. (d) Initial  $V_P$  model. (e) First shot gather. Diving and reflected waves are recorded (separated by the time-offset boundary delineated by the red dashed line), as well as several multiscattered waves indicated by the yellow ellipse and arrows. (f) Conventional FWI gradient, (g) RWI gradient and (h) JFWI gradient for  $V_P$ .

to avoid potential cycle skipping among primary reflections and secondary reflections. More realistic case involving the surface-related multiples will be discussed in Section 5.3. The source function is a Ricker wavelet with a peak frequency of 6.25 Hz. We consider a streamer acquisition of 80 shots, with the first shot gather being shown in Fig. 7(e). The reflection phases with zero-offset two-way traveltimes ranging from 0.4 s to 2.8 s are reflections from the gas layers. The reflections from the cap rock and the reservoir have zero-offset two-way traveltimes ranging from 2.8 s to 3.4 s, followed by the multiscattered phases (indicated by the yellow arrows) especially the one arriving at 3.8 s (by the yellow ellipse). These multiscattered waves are generated from the highly reflective cap rock and from the edges of the gas layers. As the direct/diving waves do not significantly overlap the reflected waves in time, we simply apply a linear time-offset window to the data set for the purpose of data separation, defined by the following formula:

$$t_{\text{sep}} \text{ (s)} = \text{offset (km)} / 1.5 \text{ (km s}^{-1}\text{)} + 0.33 \text{ (s)}, \quad (12)$$

where  $t_{\text{sep}}$  is the time instant of the window boundary as a function of offset (indicated by the red dashed line in Fig. 7e), the slope of the linear boundary is given by the water wave speed ( $1.5 \text{ km s}^{-1}$ ) and the intercept is equal to 0.33 s according to the dominant period of the Ricker wavelet. Furthermore, we estimate that the direct/diving waves have amplitudes ten times higher than those of the reflected waves. This prompts us to leave the amplitudes of the direct/diving waves unchanged in the misfit function (i.e.  $W^d$  equals to the identity matrix), while the reflection residuals are multiplied by a factor of ten in the misfit function.

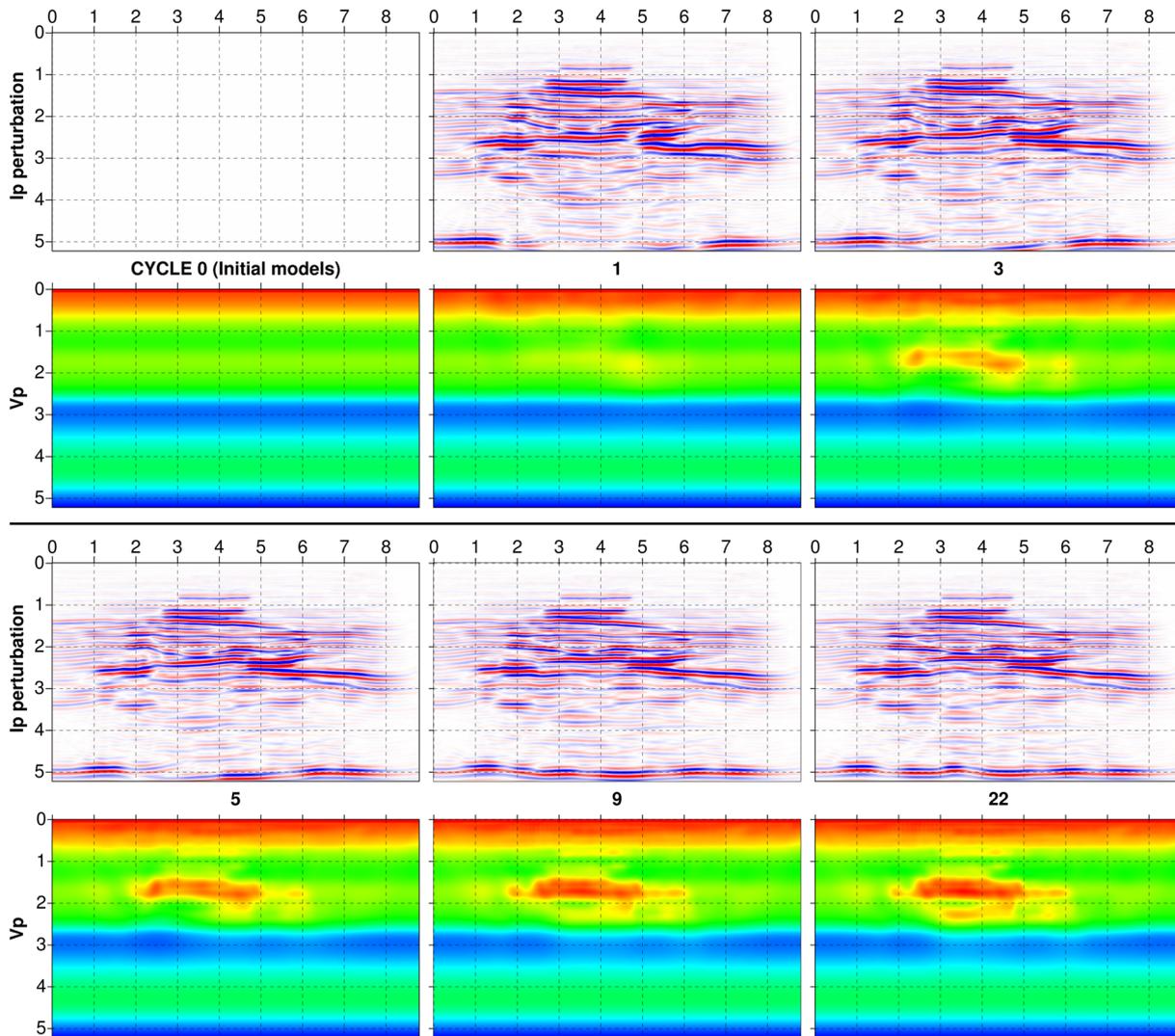
Having been observed in several tests, JFWI tends to match the recorded multiscattered phases (indicated by the yellow arrows and ellipse in Fig. 7e) with the modelled primary reflections if the full reflection wavefield is processed in one go, which implies that JFWI

also suffers from non-linearity as FWI does. Therefore, we apply a progressively increasing time window procedure (Kolb *et al.* 1986) to the data to enhance the robustness of JFWI. This is implemented by muting the full-offset reflection data after 3.5 s during first few cycles to remove the contribution of the multiscattered waves in the misfit function, then gradually restore their amplitude to their original level during later cycles.

The velocity gradients that are computed in the initial smooth model by FWI, RWI and JFWI are shown in Figs 7(f)–(h), respectively. As expected, the FWI gradient shows the limited penetration depth ( $\sim 0.8 \text{ km}$ ) of the first Fresnel zones associated with the diving waves and a high-wavenumber content at greater depths generated by the stack of the migration isochrones (Fig. 7f). The RWI gradient shows how the migration isochrones were avoided by assuming a known reflectivity in the waveform inversion formalism, while low-wavenumber components were injected at all depths along the transmission wave paths of the reflected waves that are predicted by using the prior reflectivity (Fig. 7g). The JFWI gradient yet highlights how a deficit of low-wavenumber coverage in the shallow part of the RWI gradient (Fig. 7g, see at 0.5 km and 7 km in distance) can be filled through the diving wave contribution (Fig. 7h). The optimization method relies on the L-BFGS quasi-Newton approach (Nocedal 1980). We perform 10 non-linear iterations of IpWI and 20 non-linear iterations of JFWI during each cycle of the workflow. We use only offsets smaller than 200 m to perform IpWI in order to increase the linearity of the problem (see Section 4).

## 5.2 Results and discussions

Fig. 8 displays a selection of  $I_P$  perturbations (namely  $I_P$  updates) and  $V_P$  models obtained at different cycles of the workflow (Algorithm 2). The initial  $I_P$  perturbation is zero and the initial  $V_P$  is



**Figure 8.** Selection of  $I_P$  perturbations and  $V_P$  models from the cycle IpWI+JFWI workflow. Long wavelengths are built in  $V_P$  while short wavelengths are imaged in the  $I_P$  perturbation model.

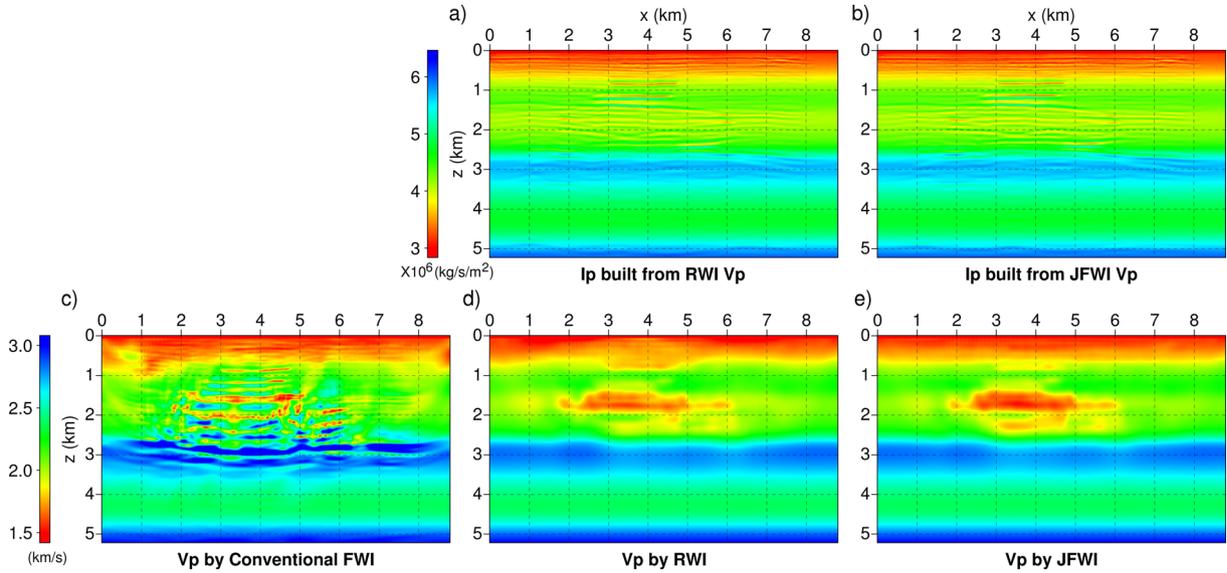
smooth (Cycle 0). After the first cycle, the  $I_P$  perturbation, computed using the initial  $V_P$  model as the background model, is not well focused especially for the sand-bedrock interface. The image of the gas-zone reflectors in the first  $I_P$  perturbation are mispositioned at excessive depths due to the overestimated velocities in the initial model. Large-scale  $V_P$  variations of the gas zones start to show up at about 1.8 km depth in the first model built by JFWI. This velocity update contributes to move the gas-zone reflectors at their correct position during IpWI in the following cycles. We perform 22 cycles of the IpWI+JFWI workflow. The final  $V_P$  and  $I_P$  models are shown in Figs 9(b) and (e), respectively. The reflectors have been moved at their correct positions, as JFWI has injected long-to-intermediate wavelengths into the velocity model. In particular, the zone of influence of the low-velocity gas-zone between 2.5 km and 5 km in distance has been well delineated. The long-wavelengths of  $V_P$  between the reservoir and bedrock are rarely imaged due to two facts: most of the incident energy is reflected backwards by the hard cap rock and the smooth variations in the sand zone provides few reflection information.

We also apply FWI and RWI on this data set to highlight the add-value provided by JFWI. Results are shown in Figs 9(a), (c) and

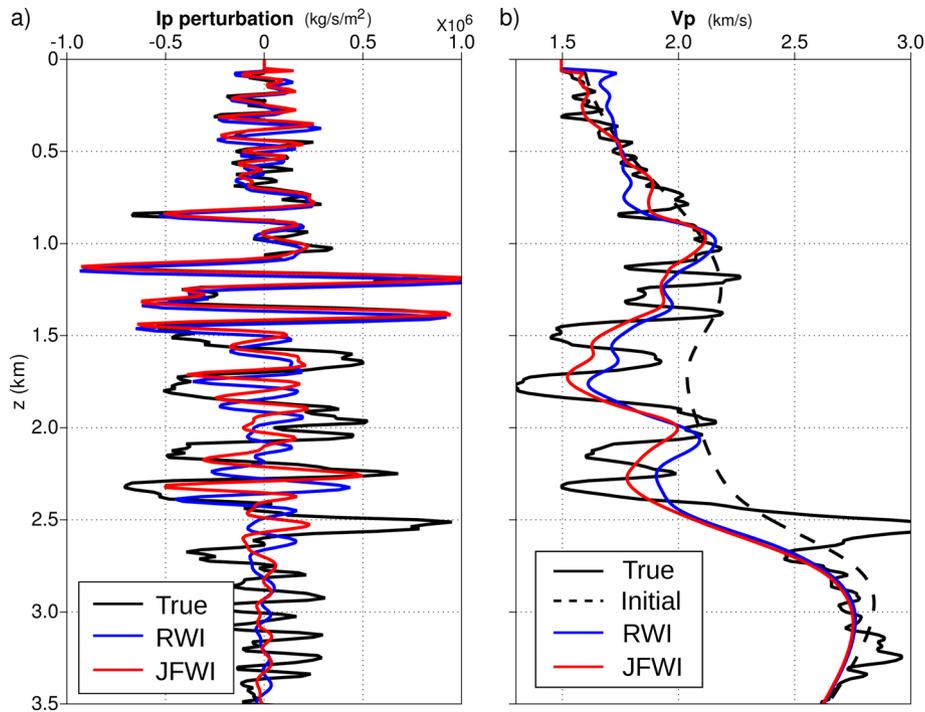
(d). FWI is implemented with the  $V_P$ - $\rho$  parametrization and only the  $V_P$  model is reconstructed. RWI is performed with the  $V_P$ - $I_P$  parametrization. The same cycle workflow as was used for JFWI is used again to perform RWI, except that the contribution of diving waves is discarded by RWI.

The final  $V_P$  model obtained by FWI is roughly the superimposition of short-wavelength velocity perturbations on the smooth initial velocity model. The inversion clearly fails to update the long-to-intermediate wavelengths of the velocity model because of the inaccuracy of the initial model and the lack of long offsets. The short-wavelength components of the reconstructed  $V_P$  are poorly focused due to the inaccurate long-wavelength components.

In contrast, both RWI and JFWI have reconstructed the long wavelengths of the gas zones to some extent. However, the final  $V_P$  model built by JFWI is significantly more accurate than the one built by RWI, especially in the shallow part where diving waves penetrate. An inspection of the vertical profiles extracted from the true model, RWI model and JFWI model supports this statement (Fig. 10, to be discussed later). The final  $I_P$  models obtained by IpWI, from either JFWI or RWI  $V_P$  models, are purely superimpositions of the short-wavelength components imaged by IpWI on the



**Figure 9.** Final subsurface models of different waveform inversion approaches. (a and b) Impedance models obtained by RWI (a) and JFWI (b). (c–e)  $V_p$  model obtained by FWI (c), RWI (d) and JFWI (e). See text for explanations.



**Figure 10.** (a and b) Logs of  $I_p$  perturbations (a) and  $V_p$  models (b) at  $x = 3.75$  km for IpWI+RWI and IpWI+JFWI. Accounting for diving waves in JFWI improves the velocity reconstruction in the shallow part, which translates into an improved imaging of the deep velocity and the impedance structures.

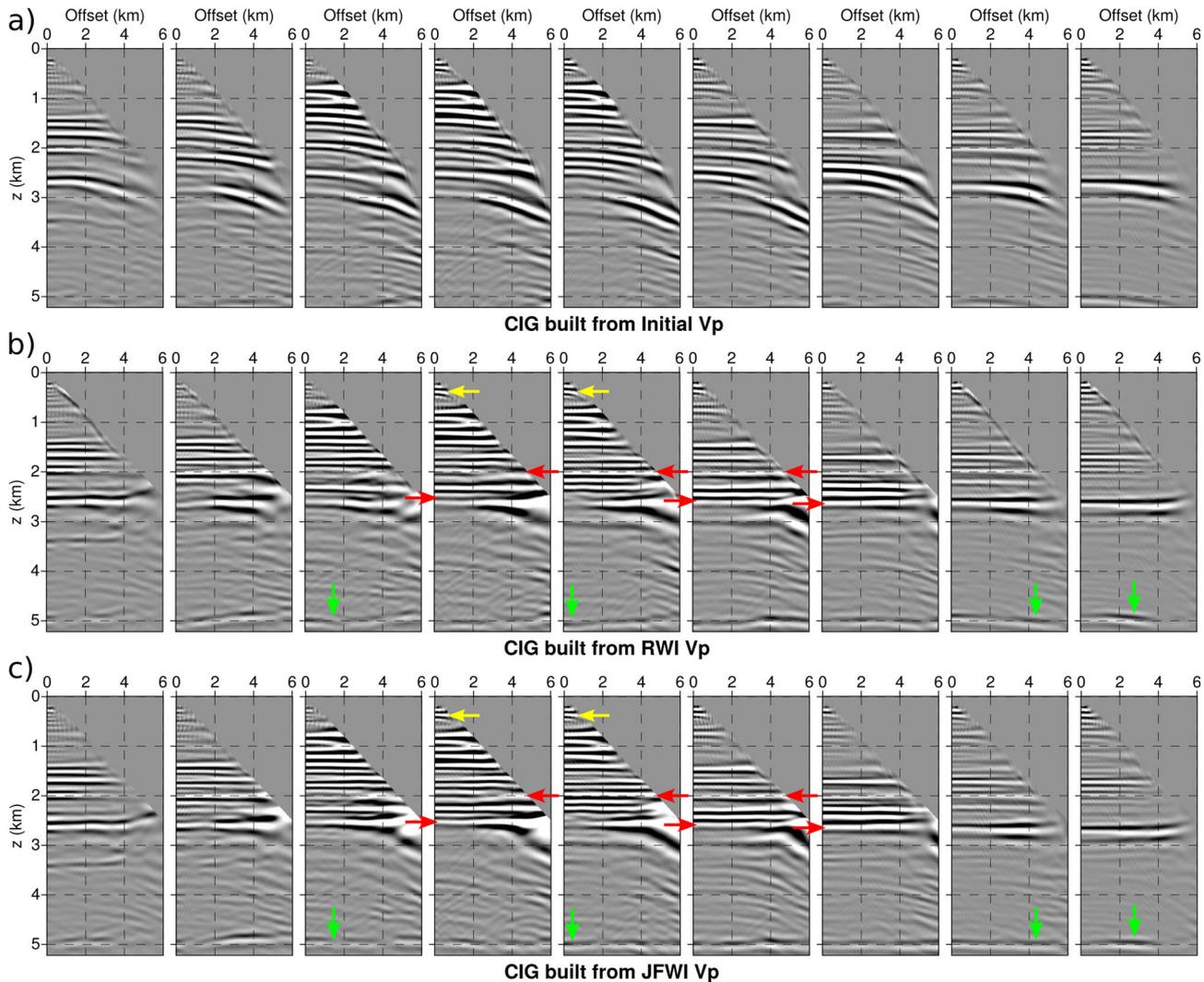
smooth initial  $I_p$  model, according to the diffraction pattern of the  $I_p$  parameter (Fig. 5d). In summary, the scale-separation condition has been fulfilled through this proposed cycle inversion.

Figs 10(a) and (b) show the logs of the  $I_p$  perturbations obtained by IpWI and the  $V_p$  models inferred from RWI and JFWI at  $x = 3.75$  km, respectively. Above  $z = 1$  km, the  $I_p$  perturbations computed from the RWI and JFWI  $V_p$  models are almost equivalent. However, the RWI  $V_p$  model is clearly less accurate than the JFWI  $V_p$  model, particularly in the shallow part where the diving waves penetrate. From 1.5 km to 2.5 km depth, due to the accumulation of inaccuracies from the near surface, the velocities of the RWI model are overestimated and prevent the correct positioning

in depth of the  $I_p$  perturbations. This highlights that, indeed, an accurate near-surface reconstruction is also critical to properly image deeper zones.

### 5.2.1 Quality control by common image gathers

We further assess the quality of the velocity models inferred from RWI and JFWI by generating common image gathers (CIGs) in the offset-depth domain (Fig. 11). These CIGs are computed in the initial  $V_p$ , RWI  $V_p$  and JFWI  $V_p$  models by reverse-time migration using the same modelling engine as was used during JFWI and IpWI.



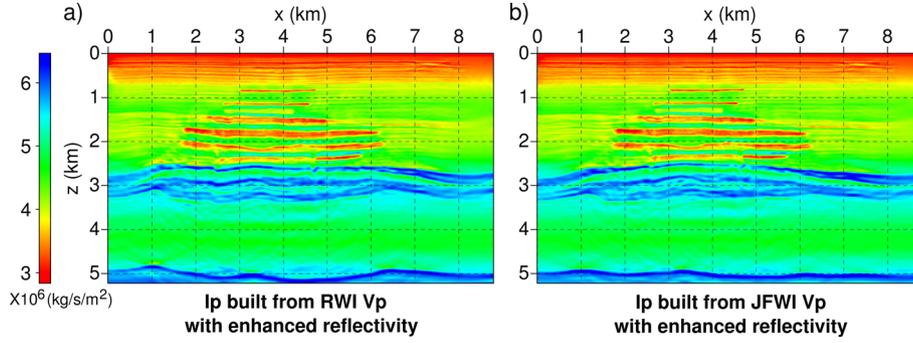
**Figure 11.** Common image gathers at positions  $x = \{1.88\ 2.50\ 3.13\ 3.75\ 4.38\ 5.00\ 5.63\ 6.25\ 6.88\}$  km from the initial  $V_p$  (a), RWI  $V_p$  (b) and JFWI  $V_p$  (c) models. Improvements to the event flatness provided by JFWI compared with RWI are pointed by horizontal yellow arrows in the shallow part, red arrows at the cap rock level and vertical green arrows at 5 km depth.

Compared with the initial  $V_p$  model, both the RWI and JFWI  $V_p$  models can significantly improve the flatness of the events in the CIGs. The improvement provided by JFWI compared with RWI is more subtle as the CIGs generated by RWI are already quite flat. Nonetheless, we still show that shallow events are flatter in the CIGs inferred from the JFWI  $V_p$  model relative to those inferred from the RWI  $V_p$  model (Fig. 11, horizontal yellow arrows). This manifests the diving wave contribution used in JFWI as well as the difficulty of the reflection-based imaging methods to update the shallow part of the subsurface. This point was illustrated with a real-data case study from Valhall by Prioux *et al.* (2011), who showed that FWI of diving waves and reflected waves improved the flatness of the CIGs in the first kilometres in depth of the subsurface compared with the CIGs inferred from a reflection traveltime tomography velocity model. Some improvements achieved by JFWI compared with RWI are also shown at the cap rock level (Fig. 11, red arrows), which highlight how the more accurate shallow velocity reconstruction impacts on the focusing of the deep reflector images. The horizontal reflector at 5 km depth is also better imaged in the JFWI model than in the RWI model (Fig. 11, vertical green arrow).

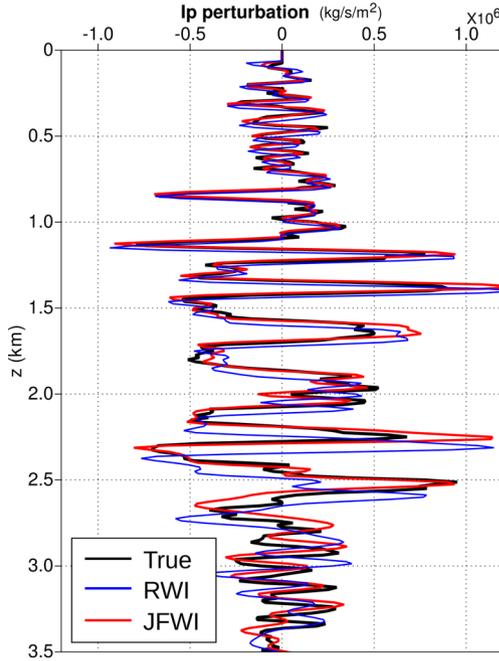
### 5.2.2 Fitting amplitudes

Due to the geometrical spreading effects that are incompletely removed during the FWI process, the amplitude of the impedance perturbations decreases with depth. Without an accurate reconstruction of the impedance contrasts, the amplitude of the modelled reflected waves can be significantly smaller than the observed amplitude. In order to improve the amplitude fit, we precondition the gradient to strengthen the deep perturbations at the expense of shallow ones and perform 40 IpWI iterations with the preconditioned L-BFGS optimization scheme, starting from the final  $V_p$  model of JFWI (Fig. 9e). We still use offsets smaller than 200 m.

The number of iterations is taken four times as before such that tiny differences in the data amplitude can influence the model update. The refined impedance perturbations computed in the RWI and JFWI velocity models are shown in Fig. 12. Compared with the previous  $I_p$  models (Figs 9a and b), the image of the deep reflectors, such as the cap rock–reservoir interface, has been significantly enhanced. Note also how the geometry of the reservoir is much more accurately delineated in the  $I_p$  model computed by using the JFWI velocity model compared with the one using the RWI velocity model.



**Figure 12.**  $I_P$  models after 40 iterations of IpWI performed with preconditioned L-BFGS optimization, using RWI (a) and JFWI (b)  $V_P$  models as background models (Figs 9d and e, respectively).



**Figure 13.** Logs of enhanced reflectivity ( $I_P$  update) at  $x = 3.75$  km. Compared with previous  $I_P$  logs (Fig. 10a), deeper reflectivity can be matched by preconditioned L-BFGS optimization starting from the RWI or JFWI  $V_P$  models. Still, JFWI performs better than RWI in terms of depth positioning.

For the sake of completeness, we plot the vertical profile of the new  $I_P$  perturbation models at  $x = 3.75$  km in Fig. 13. The amplitude of the impedance perturbations is now much better estimated in the deep part of the model.

The first shot gather computed in the true models and in the final RWI and JFWI models are compared in Figs 14 and 15. Phases and amplitudes in the JFWI-calculated data agree quite well with those of the observed data, except for the multiscattered waves (e.g. time = 4.4–5.2 s at offset = 6 km). Nevertheless, the absence of the multiscattered waves in the calculated data helps us avoid somehow the cycle-skipping issues: the phase ended in time = 4.6 s at offset = 6 km seems not to be cycle-skipped. As the sand zone is hard to be recovered (due to few reflection information from the seismogram), the moveout of the latest reflection coming from the sand–bedrock interface is less properly matched especially at offset = 4 km. More convex misfit functions (Luo & Schuster 1991; Brossier *et al.* 2009, 2010, 2015; Luo & Sava 2011; Ma & Hale 2013; Warner & Guasch 2014) would be helpful to relax the cycle-skipping issue raised by waveform-difference misfit func-

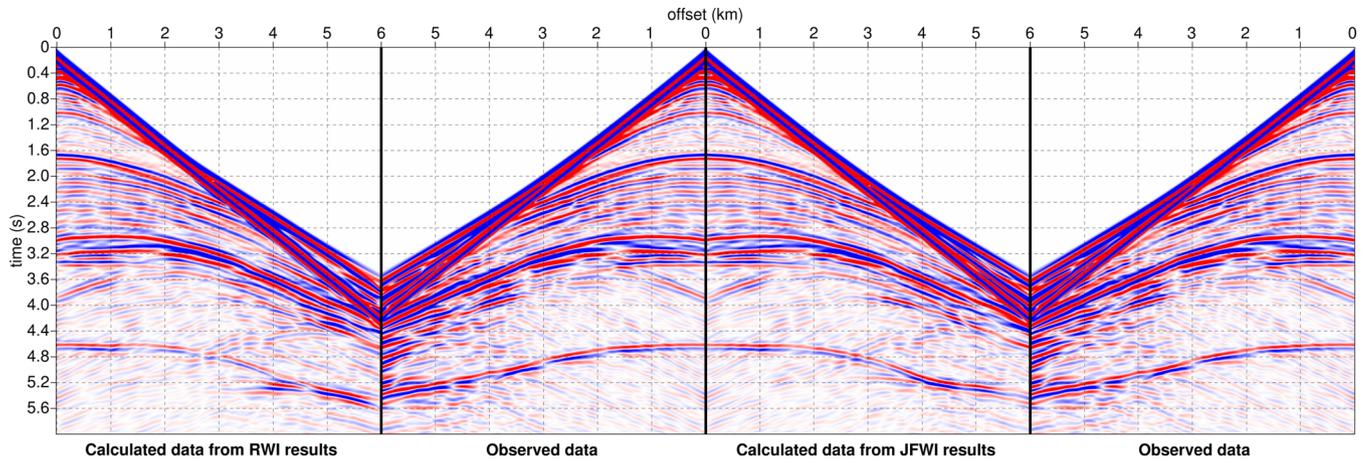
tions, and allow large traveltimes shifts for the inversion procedure to match the full-offset reflection phases. On the other hand, the data calculated in the RWI final models match the observed data at short offsets, but fail at long offsets (e.g. time = 3.6 s, 5.4 s at offset = 6 km). This means that the diving-wave information is also critical for the matching of the reflection data (phases, moveouts), and should not be discarded in the high-resolution imaging techniques.

### 5.2.3 Broad-band imaging of $V_P$

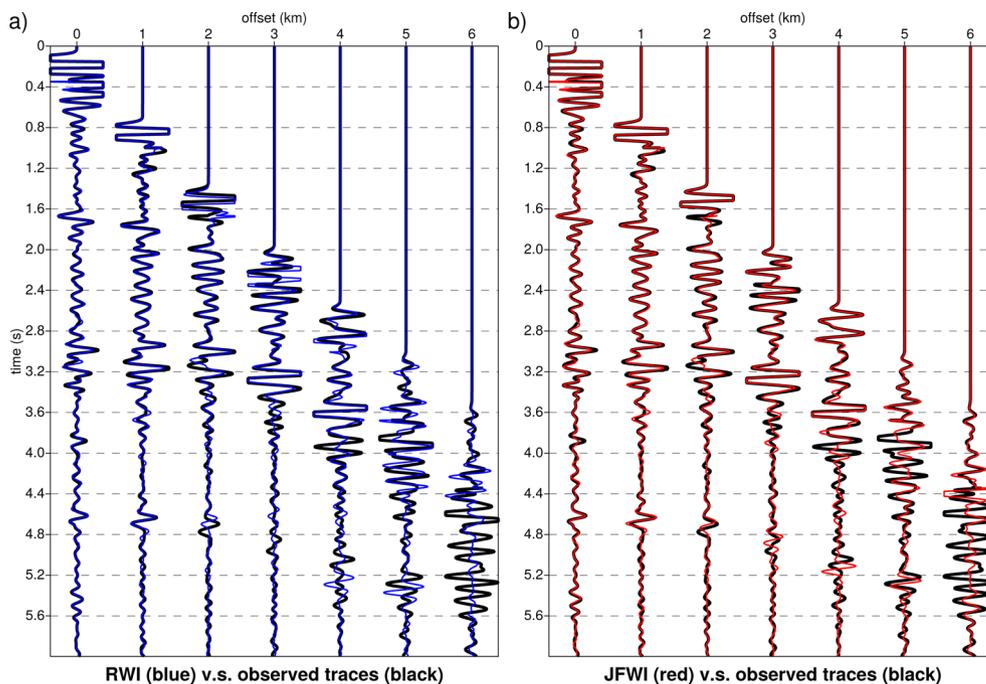
The impedance model in Fig. 12 could be used for geophysical interpretation. Alternatively, a broad-band  $V_P$  model is also very helpful to understand the structure of the subsurface and the rock properties, but conventional approaches like FWI may fail in building such a  $V_P$  model from a crude initial model due to the lack of low frequencies (Fig. 9a). Joint FWI can be used as a robust tool for initial model building since it can build the long wavelengths that are required to perform reliable FWI (Fig. 9e).

Fig. 16(a) shows the result of FWI starting from the JFWI  $V_P$  model (Fig. 9e) under the  $V_P$ – $\rho$  parametrization. The two-layer density model that was used as the initial model for JFWI is used here as the background model, and is kept fixed during this FWI implementation. The final FWI velocity model matches quite closely to the true velocity model, except in the deep part due to the lack of illumination. This result confirms the relevance of JFWI as a robust tool to build an initial velocity model for FWI, leading to a two-step velocity model building workflow of successive JFWI and FWI. However, this  $V_P$  model is prone to contain imprints from  $V_P$ – $\rho$  cross-talks: the short-scale heterogeneities that are missed in the background density model might be interpreted as the short-scale velocity perturbations by FWI, leading to an erroneous velocity estimation. This might explain some amplitude mismatches between the true and the FWI velocities in the log profile (Fig. 16b), for example at 2–2.5 km depths.

The data fit of the first shot gather computed in the true  $V_P$  model and in the final JFWI followed by FWI  $V_P$  model is shown in Fig. 17. The multiscattered waves are generated (e.g. time = 4.4–5.2 s at offset = 6 km), due to the injection of the intermediate wavenumbers into the final  $V_P$  model, which allows us to model these higher-order scattering effects from the top of the cap rock and the vertical edges of the gas layers. Except at zero offset, the match of the phase and amplitude is further improved comparing with Fig. 15, especially for the multiscattered waves and the late reflection at long offsets. In summary, this experiment manifests that JFWI has a great potentiality to be used as a robust tool for initial velocity model building for conventional FWI.



**Figure 14.** Comparison of the first shot gather computed in the true  $V_p$  model (second and fourth panels) and in the final RWI (first panel) and JFWI (third panel) models. See text for details.



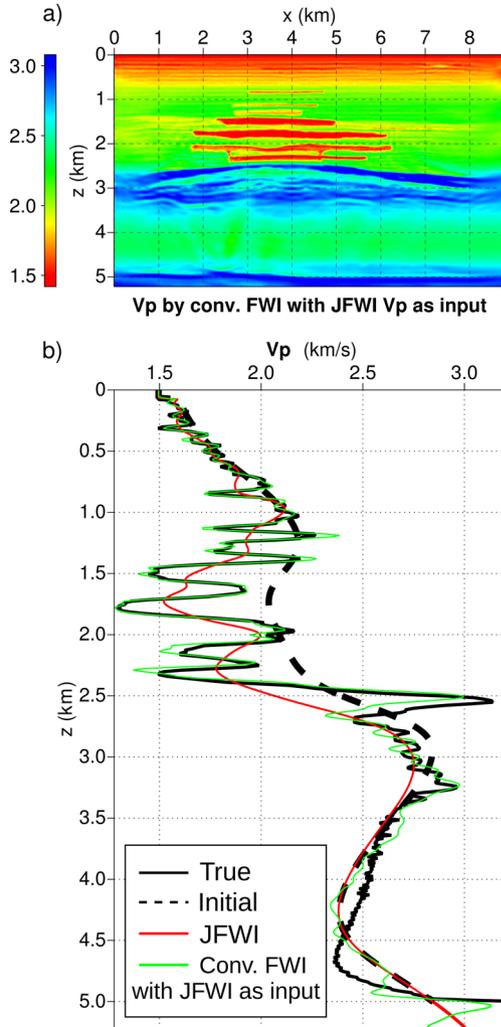
**Figure 15.** (a) Direct comparison between seismograms computed in the true  $V_p$  model (black lines) and the final RWI model (blue lines) for the first shot gather (Fig. 14). (b) Same as (a) for seismograms computed in the JFWI model (red lines). See text for details.

The match at zero offset has been degraded (within the two-way traveltimes ranging from 1.8 s to 3.4 s) due to the offset weighting that has been applied to enhance the reflection data at long offsets. Moreover, this degradation might be caused by the fact that we have kept the smooth density model to its initial values, suggesting that the model space is not large enough to account for the amplitude effects at short offsets. In other words, these amplitudes cannot be matched by a velocity-only inversion procedure. Therefore, a multiparameter FWI for  $V_p$  and  $\rho$  should be considered to improve the data fit.

### 5.3 JFWI in the presence of multiples

In this section, we discuss the robustness of JFWI in the presence of surface multiples. The interest is that if JFWI allows the multiples to be present in the data, the pre-processing workflow could be simplified.

We still consider the synthetic Valhall model (Figs 7a and c) and generate a data set with surface-related multiples (Fig. 18a). This data set is processed by JFWI and IpWI by considering a free-surface boundary condition on the surface during the seismic modelling. Compared with the data set computed without free-surface multiples (Fig. 7a), the diving waves have weaker amplitudes in the 1–2 km offset range (using the same clip), and more complex reflection wavefields are recorded. This prompts us to reduce the scaling factor applied to the reflected waves through the operator  $W^r$  from 10 to 5, such that the contributions from the diving and reflected waves are re-balanced in the JFWI misfit function. Moreover, as free-surface effects generate more multiscattering, we progressively feed the inversion with late-arriving reflections at a slower rate than in the former case performed without surface multiples. Here, we mute the reflection data after 3.3 s over the full offset range during first few cycles, then gradually restore their amplitude to their original level during later cycles (also weighted by 5).



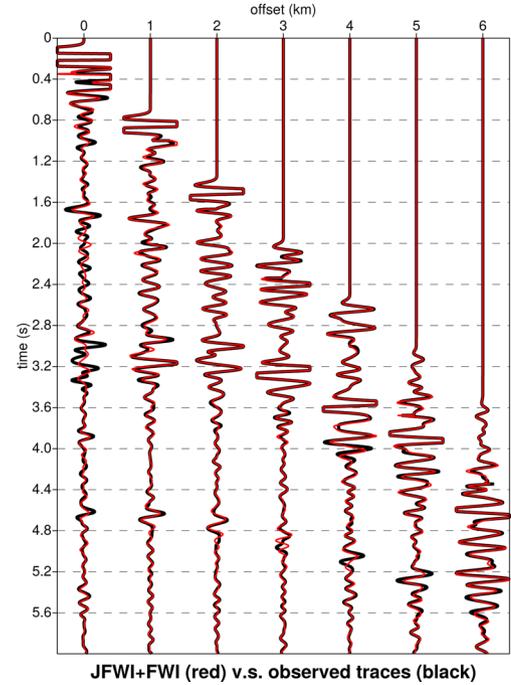
**Figure 16.** (a) Broad-band reconstruction of the velocity model by conventional FWI using JFWI  $V_P$  model as the starting model. The  $V_{P-\rho}$  parametrization is used for inversion keeping the density model fixed). (b) Comparison between velocity profiles ( $x = 3.75$  km) extracted from the true model (solid black line), the initial  $V_P$  model (dashed line), the JFWI  $V_P$  model (red line) and the broad-band  $V_P$  model shown in (a) (green line). Leakage of  $\rho$  reflectivity is expected (e.g. at 2–2.5 km depths).

We apply the same time-offset window (eq. 12) to separate the data.

Fig. 18(b) shows the JFWI result after 14 cycles. Although the final JFWI  $V_P$  model shows the long wavelengths of the true  $V_P$  model as in Fig. 9(e), we have witnessed some degradations of the quality from this velocity result:

(i) The thin reflector at the water bottom between  $x = 3$  km and 4 km is an artefact coming from the very early reflections in the diving-wave time window. Because we have applied a simple linear time-offset window, the migration isochrones generated from the reflections left in the diving-wave time window are not totally cancelled during the summation of  $G_1$  and  $G_2$ . When multiples are present, these reflections may be enhanced, making the migration isochrones apparent in the gradient. This highlights that a successful application of JFWI heavily relies on the accuracy of data separation.

(ii) Above 2 km depth, the resolution seems to be higher in Fig. 18(b) than in Fig. 9(e), even though we have applied the same



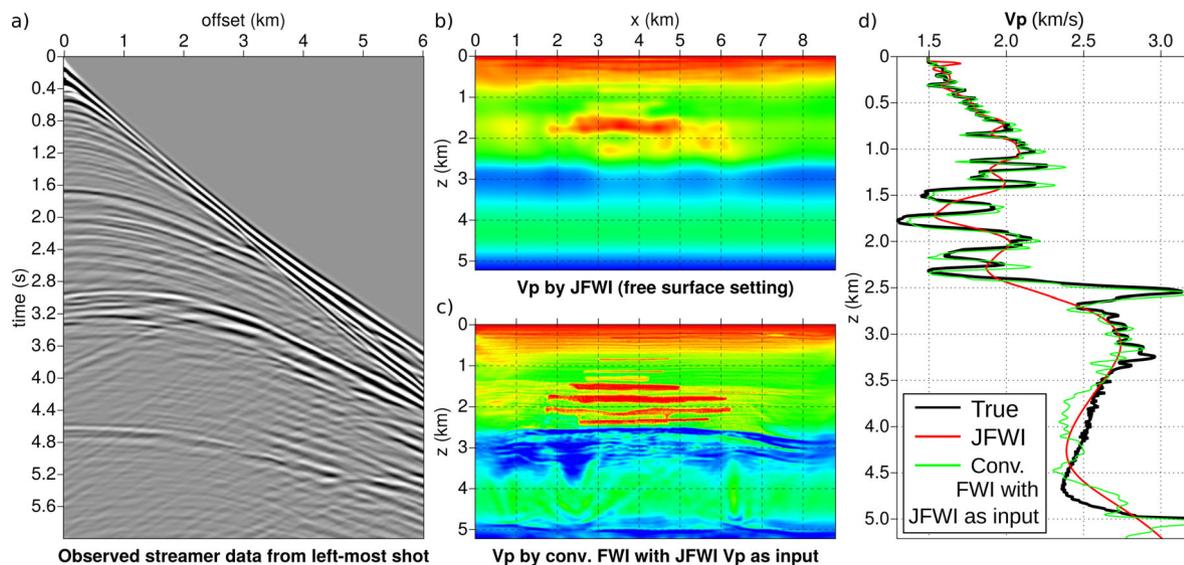
**Figure 17.** Data fit of the first shot gather computed in the true  $V_P$  model (black) and in the final JFWI+FWI  $V_P$  model (red), which is comparable with the one obtained with the JFWI  $V_P$  model (Fig. 15). Here, the fit at long offsets of multiscattered waves has been nicely improved at the expense of the fit at short offsets. See text for interpretation.

Gaussian smoothing regularization to the gradient (vertical and horizontal correlation lengths equal to twice of the dominant wavelength). This improved resolution might result from the improved subsurface coverage provided by the surface-related multiples. This statement deserves however further investigations.

(iii) In contrast, the low-velocity zone at 2.3 km depth is reconstructed less accurately in Fig. 18(b) than in Fig. 9(e). We consider that imaging at this depth, covered by only reflection kernels, is a difficult task especially when multiples are present.

As was in Section 5.2.3, we launch the conventional FWI starting from the JFWI  $V_P$  macromodel (Fig. 18b) using the  $V_{P-\rho}$  parametrization. The two-layer density model that was used as the initial model for JFWI is also used here as the background model, and is kept fixed during this FWI implementation. The result is shown in Fig. 18(c). The gas zone is properly reconstructed, with some noise due to the multiples. The image of the cap rock, however, is not as accurate as the one built from multiple-free data. We believe that this degradation results from the surface multiples and the multiscattered waves generated by the hard interfaces and the edge of the gas layers. Remembering the former application without multiples, these multiscattered waves had already made the velocity model building quite non-linear, which prompted us to design a time-windowing approach to mitigate this non-linearity. Here, as expected, the presence of multiples has significantly increased the non-linearity, which forces us to apply a more prudent time-windowing approach. With such an effort, the degradation of the resolution is limited.

Regardless the difficulties raised by these two wave modes, we still obtain a reasonable velocity model by FWI using the JFWI model as the starting model. The log extracted at  $x = 3.75$  km is shown in Fig. 18(d). Without considering possible cross-talk effects between parameters, the final  $V_P$  values (green curve) fit quite well



**Figure 18.** Synthetic Valhall case study with free surface multiples. (a) First shot gather. (b) Long-wavelength  $V_p$  model by JFWI. (c) Broad-band reconstruction of the  $V_p$  model (conventional FWI implementation starting from (b),  $V_p$ - $\rho$  parametrization and density fixed). (d) Logs of  $V_p$  models at  $x = 3.75$  km for (b) and (c). JFWI suffers from the high-order scattering effects from the surface-related multiples and the multiscattered waves, therefore the  $V_p$  results are worse than the previous results (Figs 9e and 16a). However, from the log we see that the gas layers are still well imaged. See text for details.

the true  $V_p$  values (black curve) above 3.2 km depth, and is much comparable with the one inferred from multiple-free data (Fig. 16).

## 6 CONCLUSIONS AND PERSPECTIVES

Most applications of the conventional FWI are driven by diving waves and post-critical reflections. However, imaging the long-to-intermediate wavelengths in deep regions remains challenging if the transmitted waves do not penetrate these regions due to insufficient offsets. In contrast, RWI succeeds to some extent in imaging the long-to-intermediate wavelengths of the deep regions focusing on the two transmission wave paths followed by the reflected waves from the surface to the reflector positions. This kind of approaches relies on a scale separation between a smooth velocity model and a known reflectivity, that allows one to suppress the contribution of the unwanted (high-wavenumber) migration isochrones during the velocity model building. The limitation of RWI is that diving waves and post-critical reflections are not used, although they carry the essential long-wavelength information of the shallow subsurface. We have presented in this paper an extension of RWI, namely Joint FWI, the aim of which is to account for the contribution of the diving waves and post-critical reflections during the velocity model building. The relevance of the method is demonstrated with the synthetic Valhall case study, which has shown that how the improvement of the near surface imaging provided by the diving waves translates into an improved imaging of the deep targets performed by RWI. Although modern wide-azimuth long-offset seismic acquisition may still not allow to record the diving waves with a sufficient penetration depth to sample the deepest targeted structures, diving waves and post-critical reflections can have an increasing contribution in the seismic wavefield. In this study, we have proposed a seismic workflow which makes an optimal use of the information carried by all kinds of waves to build a reliable velocity macromodel. Our approach requires however the explicit separation of the diving waves (or post-critical reflections) and the pre-critical reflections.

The conventional  $V_p$ - $\rho$  parametrization of the subsurface favours a broad-band reconstruction of the  $V_p$  parameter, but the mitigation

of the cross-talks between  $V_p$  and  $\rho$  is challenging from reflected waves. In contrast, the  $V_p$ - $I_p$  parametrization leads to a more natural uncoupling between the two parameter classes, which is consistent with the scale separation between the velocity model and the reflectivity underlying RWI and JFWI. It becomes therefore natural to combine within an iterative workflow the velocity model building performed by JFWI and the impedance imaging performed by conventional FWI of reflected waves (IpWI). As the impedance model needs to be updated according to the velocity updates, JFWI and IpWI are performed in an alternate way leading to the cycle workflow.

For more efficient implementations of the cycle workflow, one possibility is to reduce the iteration number of the  $I_p$  inversion. Although the amplitudes of the imaged reflectivity will be incompletely estimated, the kinematics of the data can be retrieved to reconstruct a reliable  $V_p$  background model. As soon as the kinematic attributes of the data are matched, amplitudes can be further used to build more accurate reflectivity images.

Having respectively built the low-part and the high-part of the  $V_p$  and  $I_p$  spectra, it is natural to wonder whether a broad spectrum of the subsurface is possible. Starting from the velocity macromodel built by JFWI, a broad-band velocity model can be tentatively imaged by conventional FWI using the  $V_p$ - $\rho$  parametrization. The quality of the reconstructed velocity model shows that the low-to-intermediate wavenumber components of the JFWI model are accurate enough to successfully image the subsurface from reflected waves by FWI. In this experiment, only the  $V_p$  parameter is updated keeping the density fixed to its original value. Therefore, this  $V_p$  model is prone to contain imprints from  $V_p$ - $\rho$  cross-talks. These cross-talks probably manifest by the overestimation and/or underestimation of the reconstructed velocities. To tackle this kind of cross-talks, multiparameter inversions which involves the Hessian operator should be considered. Depending on the computational facilities, L-BFGS (Nocedal 1980) or Truncated Newton (Métivier *et al.* 2013) methods could be considered to introduce the Hessian operator.

On the other hand, imaging low-wavenumber part of the  $I_p$  spectrum could be more difficult, as neither  $V_p$ - $\rho$  nor  $V_p$ - $I_p$

parametrization can provide low-wavenumber sensitivity to the  $I_p$  parameter (cross-talks arise in the transmission regimes for the  $\rho$ - $I_p$  parametrization). In this case, we have to rely on empirical parametric relations such as the Gardner relation. From this kind of relation, we could rebuild the initial  $I_p$  in each cycle to translate low wavenumbers from  $V_p$  to  $I_p$ , or set a trust region to bound the searching area of the  $I_p$  inversion.

We also look forward to the applications of JFWI to real data sets such as the real Valhall one. The cycle-skipping issue would appear and could hinder the local search method to reach the global minimum. Therefore, more robust misfit definitions would be required. Besides, how the cycle inversion scheme is affected by the elastic information of the data should be studied. For example, we foresee that the misinterpretation of elastic wavefields as acoustic ones would generate reflection residuals due to inaccurate amplitudes. The inversion of these residuals by JFWI would generate some artificial velocity perturbations even if the initial  $V_p$  model is correct. Accounting for density to absorb the elastic effects is an option, although this inevitably questions the use of the  $V_p$ - $I_p$  parametrization during JFWI (Borisov *et al.* 2014; Plessix *et al.* 2014; Plessix & Solano 2015). Another possible strategy is to correct the acoustic wavefields for the elastic effects by using artificial source terms (Chapman *et al.* 2014; Hobro *et al.* 2014). A last-but-not-least possibility could be to rely on the kinematic-associated misfit functions such as the ones based on cross-correlation (van Leeuwen & Mulder 2010), deconvolution (Luo & Sava 2011; Warner & Guasch 2014), instantaneous phase (Fichtner *et al.* 2008; Bozdag *et al.* 2011), or dynamic warping (Hale 2013). Extensions to 3-D geometry could be performed, but the repetition of the  $I_p$  inversion inside the cycle workflow would be a computational obstacle. One possible solution is to build the reflectivity in the pseudo-time domain instead of the depth domain, by which reflection phases are always matched in short offsets (Plessix 2013; Wang *et al.* 2015).

## ACKNOWLEDGEMENTS

This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07\_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS under the allocation 046091 made by GENCI. Authors appreciate fruitful discussions with Ludovic Métivier (LJK-CNRS, Université Grenoble Alpes) and Yang Li (Harbin Institute of Technology), they thank the editor R.E. Plessix, Partha Routh and an anonymous reviewer for their constructive comments which helped us improve the paper.

## REFERENCES

Alkhalifah, T., 2014. Monofrequency waveform acquisition and inversion: a new paradigm, in *Expanded Abstracts, 84th Annual SEG Meeting*, Denver.  
Alkhalifah, T., 2015. Scattering-angle based filtering of the waveform inversion gradients, *Geophys. J. Int.*, **200**(1), 363–373.

Alkhalifah, T. & Wu, Z., 2014. FWI and MVA the natural way, in *EAGE, 76th EAGE Conference and Exhibition incorporating SPE EUROPEC 2014*.  
Allemand, T. & Lambaré, G., 2014. Full waveform inversion guided migration velocity analysis, in *SEG Technical Program Expanded Abstracts 2014*, pp. 4712–4717.  
Almomin, A. & Biondi, B., 2012. Tomographic full waveform inversion: practical and computationally feasible approach, *SEG Technical Program Expanded Abstracts 2012*, pp. 1–5.  
AlTheyab, A., Wang, X. & Schuster, G.T., 2013. Time-domain incomplete Gauss-Newton full-waveform inversion of Gulf of Mexico data, in *SEG Technical Program Expanded Abstracts 2013*, pp. 5175–5179.  
Bérenger, J.-P., 1994. A perfectly matched layer for absorption of electromagnetic waves, *J. Comput. Phys.*, **114**, 185–200.  
Biondi, B. & Almomin, A., 2012. Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis, *SEG Technical Program Expanded Abstracts 2012*, pp. 1–5.  
Borisov, D., Stopin, A. & Plessix, R.-E., 2014. Acoustic pseudo-density full waveform inversion in the presence of hard thin beds, in *Expanded Abstracts, 76th Annual EAGE Meeting*, Amsterdam.  
Bozdag, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.*, **185**(2), 845–870.  
Brossier, R., Operto, S. & Virieux, J., 2009. Robust frequency-domain full-waveform inversion using the  $l_1$  norm, *Geophys. Res. Lett.*, **36**, L20310, doi:10.1029/2009GL039458.  
Brossier, R., Operto, S. & Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion?, *Geophysics*, **75**(3), R37–R46.  
Brossier, R., Operto, S. & Virieux, J., 2015. Velocity model building from seismic reflection data by full waveform inversion, *Geophys. Prospect.*, **63**, 354–367.  
Chapman, C.H., Hobro, J.W.D. & Robertsson, J.O.A., 2014. Correcting an acoustic wavefield for elastic effects, *Geophys. J. Int.*, **197**, 1196–1214.  
Chavent, G., 1996. Duality methods for waveform inversion, *Tech. rep.*, Research report, RR-2975, INRIA-00073723.  
Chavent, G., Clément, F. & Gómez, S., 1994. Automatic determination of velocities via migration-based traveltimes waveform inversion: a synthetic data example, *SEG Technical Program Expanded Abstracts 1994*, pp. 1179–1182.  
Claerbout, J., 1985. *Imaging the Earth's Interior*, Blackwell Scientific Publication.  
Clapp, R., 2008. Reverse time migration: saving the boundaries, Tech. Rep. SEP-136, Stanford Exploration Project.  
Clément, F., Chavent, G. & Gómez, S., 2001. Migration-based traveltimes waveform inversion of 2-D simple structures: a synthetic example, *Geophysics*, **66**, 845–860.  
Devaney, A.J., 1982. A filtered backprojection algorithm for diffraction tomography, *Ultrason. Imaging*, **4**, 336–350.  
Diaz, E., Sava, P. & Yang, T., 2013. Data-domain and image-domain wavefield tomography, *Leading Edge*, **32**(9), 1064–1072.  
Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.*, **175**, 665–685.  
Hale, D., 2013. Dynamic warping of seismic images, *Geophysics*, **78**(2), S105–S115.  
Hobro, J.W.D., Chapman, C.H. & Robertsson, J.O.A., 2014. A method for correcting acoustic finite-difference amplitudes for elastic effects, *Geophysics*, **79**, T243–T255.  
Jannane, M. *et al.*, 1989. Wavelengths of Earth structures that can be resolved from seismic reflection data, *Geophysics*, **54**(7), 906–910.  
Kolb, P., Collino, F. & Lailly, P., 1986. Prestack inversion of 1-D medium, in *Extended Abstracts*, Vol. 74, pp. 498–508.  
Komatitsch, D. & Martin, R., 2007. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation, *Geophysics*, **72**(5), SM155–SM167.

- Lameloise, C.-A., Chauris, H. & Noble, M., 2015. Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis, *Geophys. Prospect.*, **63**, 391–404.
- Liu, Y., Symes, W., Huanf, Y. & Li, Z., 2013. Linearized extended waveform inversion via differential semblance optimization in the depth-oriented extension, in *Expanded Abstracts*, Vol. 171, pp. 4869–4873, SEG.
- Luo, S. & Sava, P., 2011. A deconvolution-based objective function for wave-equation inversion, *SEG Technical Program Expanded Abstracts*, **30**(1), 2788–2792.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation travelttime inversion, *Geophysics*, **56**(5), 645–653.
- Ma, Y. & Hale, D., 2013. Wave-equation reflection travelttime inversion with dynamic warping and full waveform inversion, *Geophysics*, **78**(6), R223–R233.
- Métivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full waveform inversion and the truncated Newton method, *SIAM J. Sci. Comput.*, **35**(2), B401–B437.
- Miller, D., Oristaglio, M. & Beylkin, G., 1987. A new slant on seismic imaging: migration and integral geometry, *Geophysics*, **52**(7), 943–964.
- Mora, P.R., 1989. Inversion = migration + tomography, *Geophysics*, **54**(12), 1575–1586.
- Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage, *Math. Comput.*, **35**(151), 773–782.
- Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A. & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *Leading Edge*, Special Section Full Waveform Inversion (September), 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.
- Plessix, R.E., 2013. A pseudo-time formulation for acoustic full waveform inversion, *Geophys. J. Int.*, **192**, 613–630.
- Plessix, R.-E. & Solano, C.P., 2015. Modified surface boundary conditions for elastic waveform inversion of low-frequency wide-angle active land seismic data, *Geophys. J. Int.*, **201**, 1324–1334.
- Plessix, R.E., Stopin, A., Milcik, P. & Matson, K., 2014. Acoustic and anisotropic multiparameter seismic full waveform inversion case studies, in *Expanded Abstracts, 84th Annual SEG Meeting*, Denver.
- Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O. & Kommedal, J., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophys. J. Int.*, **187**, 1495–1515.
- Sava, P. & Biondi, B., 2004. Wave-equation migration velocity analysis. I. Theory, *Geophys. Prospect.*, **52**(6), 593–606.
- Sava, P. & Fomel, S., 2006. Time-shift imaging condition in seismic migration, *Geophysics*, **71**(6), S209–S217.
- Shin, C. & Cha, Y.H., 2008. Waveform inversion in the Laplace domain, *Geophys. J. Int.*, **173**(3), 922–931.
- Snieder, R., Xie, M., Pica, A. & Tarantola, A., 1989. Retrieving both the impedance contrast and background velocity: a global strategy for the seismic reflection problem, *Geophysics*, **54**, 991–1000.
- Sun, D. & Symes, W.W., 2012. Waveform inversion via nonlinear differential semblance optimization, *SEG Technical Program Expanded Abstracts 2012*, pp. 1–7.
- Symes, W.W. & Carazzone, J.J., 1991. Velocity inversion by differential semblance optimization, *Geophysics*, **56**, 654–663.
- Tang, Y., Lee, S., Baumstein, A. & Hinkley, D., 2013. Tomographically enhanced full waveform inversion, in *Expanded Abstracts*, Vol. 171, pp. 1037–1041, SEG.
- van Leeuwen, T. & Herrmann, F.J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, **195**, 661–667.
- van Leeuwen, T. & Mulder, W.A., 2010. A correlation-based misfit criterion for wave-equation travelttime tomography, *Geophys. J. Int.*, **182**(3), 1383–1394.
- Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Wang, F., Chauris, H., Donno, D. & Calandra, H., 2013a. Taking advantage of wave field decomposition in full waveform inversion, in *EAGE Technical Program Expanded Abstracts 2013*, p. Tu0708.
- Wang, S., Chen, F., Zhang, H. & Shen, Y., 2013b. Reflection-based full waveform inversion (RFWI) in the frequency domain, in *Expanded Abstracts*, Vol. 171, pp. 877–881, SEG.
- Wang, H., Singh, S., Audebert, F. & Calandra, H., 2015. Inversion of seismic refraction and reflection data for building long-wavelengths velocity models, *Geophysics*, **80**(2), R81–R93.
- Warner, M. & Guasch, L., 2014. Adaptive waveform inversion: FWI without cycle skipping—theory, in *76th EAGE Conference and Exhibition 2014*.
- Woodward, M.J., 1992. Wave-equation tomography, *Geophysics*, **57**, 15–26.
- Wu, R.S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, **52**, 11–25.
- Wu, Z. & Alkhalifah, T., 2014. Full waveform inversion based on the optimized gradient and its spectral implementation, in *Expanded Abstracts*, p. Tu P01 11, EAGE.
- Xu, S., Wang, D., Chen, F., Lambaré, G. & Zhang, Y., 2012. Inversion on reflected seismic wave, *SEG Technical Program Expanded Abstracts 2012*, pp. 1–7.
- Yang, T. & Sava, P., 2011. Wave-equation migration velocity analysis with time-shift imaging, *Geophys. Prospect.*, **59**(4), 635–650.

## APPENDIX A: DERIVATION OF GRADIENTS THROUGH LAGRANGIAN FORMULATION

Although we perform full waveform inversion in the time domain, we shall derive the gradient formulations in the frequency domain for the sake of compactness. The Lagrangian quantities of FWI, RWI, FWI with reflectivity (denoted by FWI+reflectivity in Fig. 2(d), here FWI2 for short), and JFWI can be expressed with common notations [see Plessix 2006 for a review of adjoint-state method]:

$$\mathcal{L} = \text{Misfit function} + \Re\{a|\text{First constraint}\} + \Re\{b|\text{Second constraint}\}, \quad (\text{A1})$$

where  $a$  and  $b$  denote the adjoint-state variables and  $\Re\{\cdot|\cdot\}$  denotes the real part of the inner product. The specific expressions of other notations are listed in Table A1. The physical meaning of  $u_0$  and  $\delta u$  are the background and scattered components of the modelled wavefield, respectively. Starting from the smooth initial model, the scattered field is not generated during the first iteration of FWI, therefore  $\delta u = 0$  and only the first constraint is required. The operator  $B$  denotes the modelling operator,  $s$  the source term and  $R$  the real-valued sampling operator that extracts the calculated data from the modelled field.

The expression of the first constraint is derived from the forward problem equation [i.e.  $B(m)u = s$ ] in  $m$  or  $m_0$  to constrain  $u_0$ . Similarly, the expression of the second constraint is from the equation  $B(m + \Delta m)(u + \delta u) = s = B(m)u$  to constrain  $\delta u$ . The source term  $-\Delta B u_0$  with  $\Delta B = B(m_0 + \delta m) - B(m_0)$  emits the scattered field when the background field  $u_0$  hits local diffractors. Since the modelling operator  $B(m_0 + \delta m)$  depends on  $\delta m$ , high-order scattering effects are accounted for in  $\delta u$ .

Setting the derivatives of eq. (A1) with respect to the state variables to zero gives the adjoint-state equations, listed in Table A2, where  $\Delta d^*$  denotes the conjugate of the data residual,  $T$  the transpose operation and  $\dagger$  the adjoint operation. In the first column,  $\Delta d^*$  represents the data residuals at the receiver positions, reversed in time, augmented with zeroes in the subsurface model by the prolongation operator  $R^T$ , and used as virtual sources to produce the

**Table A1.** Specific expressions of the Lagrangian terms for each formulation.

	Misfit function	State variable(s)	First constraint	Second constraint
FWI	$0.5\ W(d_{obs} - Ru)\ ^2$	$u_0$	$B(m)u_0 - s$	–
RWI	$0.5\ W^r(d_{obs}^{refl} - R\delta u)\ ^2$	$u_0$ and $\delta u$	$B(m_0)u_0 - s$	$B(m_0 + \delta m)\delta u + \Delta Bu_0$
FWI2	$0.5\ W^d(d_{obs}^{div} - Ru_0) + W^r(d_{obs}^{refl} - R\delta u)\ ^2$	$u_0$ and $\delta u$	$B(m_0)u_0 - s$	$B(m_0 + \delta m)\delta u + \Delta Bu_0$
JFWI	$0.5\ W^d(d_{obs}^{div} - Ru_0)\ ^2 + 0.5\ W^r(d_{obs}^{refl} - R\delta u)\ ^2$	$u_0$ and $\delta u$	$B(m_0)u_0 - s$	$B(m_0 + \delta m)\delta u + \Delta Bu_0$

**Table A2.** Adjoint-state equations for each formulation.

	$\partial_{\delta u}\mathcal{L} = 0$ gives	$\partial_{u_0}\mathcal{L} = 0$ gives
FWI	–	$B(m)^\dagger a = -R^\top W^\top W\Delta d^*$
RWI	$B(m_0 + \delta m)^\dagger b = -R^\top W^{r\top} W^r \Delta d^{refl*}$	$B(m_0)^\dagger a = -\Delta B^\dagger b$
FWI2	$B(m_0 + \delta m)^\dagger b = -R^\top W^{r\top} (W^d \Delta d^{div*} + W^r \Delta d^{refl*})$	$B(m_0)^\dagger a = -\Delta B^\dagger b - R^\top W^{d\top} (W^d \Delta d^{div*} + W^r \Delta d^{refl*})$
JFWI	$B(m_0 + \delta m)^\dagger b = -R^\top W^{r\top} W^r \Delta d^{refl*}$	$B(m_0)^\dagger a = -\Delta B^\dagger b - R^\top W^{d\top} W^d \Delta d^{div*}$

**Table A3.** Expressions of  $a$  and  $b$  for each formulation.

	$b$	$a$
FWI	–	$W^2(\lambda_0^d + \lambda_0^r)$
RWI	$W^{r2}(\lambda_0^r + \delta\lambda^r)$	$W^{r2}\delta\lambda^r$
FWI2	$W^{rd}(\lambda_0^d + \delta\lambda^d) + W^{r2}(\lambda_0^r + \delta\lambda^r)$	$W^{d2}\lambda_0^d + W^{rd}\delta\lambda^d + W^{dr}\lambda_0^r + W^{r2}\delta\lambda^r$
JFWI	$W^{r2}(\lambda_0^r + \delta\lambda^r)$	$W^{d2}\lambda_0^d + W^{r2}\delta\lambda^r$

adjoint field  $b$ . In the second column, the other adjoint quantity  $a$  is emitted by the residual source at receiver positions or by the secondary sources located at diffractor positions  $\delta m$ . The adjoint propagator  $B^\dagger$  indicates that the computation of the adjoint quantities can be implemented by modifying the forward modelling code without much effort, and the cost to compute one adjoint quantity is the same as the cost of one forward modelling computation.

Analysis of these equations allows us to give a more physical interpretation of the adjoint fields  $a$  and  $b$ , listed in Table A3, where  $\lambda_0^d$  denotes the background adjoint field from the diving wave residuals,  $\delta\lambda^d$  the scattered adjoint field from the diving wave residuals,  $\lambda_0^r$  the background adjoint field from the reflected wave residuals, and  $\delta\lambda^r$  the scattered adjoint field from the reflected wave residuals.  $W^2$  is the short-hand for  $W^\top W$ , and similarly  $W^{d2}$  for  $W^{d\top} W^d$ ,  $W^{r2}$  for  $W^{r\top} W^r$ ,  $W^{dr}$  for  $W^{d\top} W^r$  and  $W^{rd}$  for  $W^{r\top} W^d$ . Instead of computing the adjoint quantities  $a$  and  $b$ , we actually compute the adjoint fields  $\lambda$ 's because their source terms are easier to be built: the evaluation of the operator  $\Delta B$  is not required and only the data residuals are computed.

In the adjoint-state method, the gradient is found by taking the derivative of eq. (A1) with respect to  $m$  or  $m_0$ . Inserting the respective expressions of  $a$ ,  $b$ , we find that

$$\nabla C_{FWI} = (\lambda_0^d + \lambda_0^r)^\dagger W^2 \frac{\partial B(m)}{\partial m} u_0, \quad (\text{A2})$$

$$\nabla C_{RWI} = \delta\lambda^{r\dagger} W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} u_0 + \lambda_0^{r\dagger} W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} \delta u \quad (\text{A3})$$

$$+ \delta\lambda^{r\dagger} W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} \delta u \quad (\text{A4})$$

$$+ \lambda_0^{r\dagger} W^{r2} \frac{\partial \Delta B}{\partial m_0} u_0, \quad (\text{A5})$$

$$\nabla C_{FWI2} = \left[ \lambda_0^{d\dagger} W^{d2} + \lambda_0^{r\dagger} W^{r2} \right] \frac{\partial B(m_0)}{\partial m_0} u_0 \quad (\text{A6})$$

$$+ \left[ \delta\lambda^{d\dagger} W^{dr} + \delta\lambda^{r\dagger} W^{r2} \right] \frac{\partial B(m_0 + \delta m)}{\partial m_0} u_0 \quad (\text{A7})$$

$$+ \left[ (\lambda_0^d + \delta\lambda^d)^\dagger W^{dr} + (\lambda_0^r + \delta\lambda^r)^\dagger W^{r2} \right] \frac{\partial B(m_0 + \delta m)}{\partial m_0} \delta u \quad (\text{A8})$$

$$+ \left[ \lambda_0^{d\dagger} W^{dr} + \lambda_0^{r\dagger} W^{r2} \right] \frac{\partial \Delta B}{\partial m_0} u_0, \quad (\text{A9})$$

$$\nabla C_{JFWI} = \lambda_0^{d\dagger} W^{d2} \frac{\partial B(m_0)}{\partial m_0} u_0 + \delta\lambda^{r\dagger} W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} u_0 \quad (\text{A10})$$

$$+ \lambda_0^{r\dagger} W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} \delta u + \delta\lambda^{r\dagger} W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} \delta u \quad (\text{A11})$$

$$+ \lambda_0^{r\dagger} W^{r2} \frac{\partial \Delta B}{\partial m_0} u_0. \quad (\text{A12})$$

Apart from the fact that they are formulated in the frequency domain, these gradients expressions are the exact forms of the compact ones provided in eqs (3), (5) and (7), respectively. For FWI2, the gradient expression is not given but illustrated in Fig. 3(d).

The terms associated with  $\partial \Delta B / \partial m_0$  have non-zero values only at reflector positions (where  $\delta m$  differs from 0). Regularization of the inverse problem helps suppress these unwanted contributions

because they have high-wavenumber contents. Therefore, the last terms of RWI, FWI2 and JFWI gradients (A5), (A9) and (A12) can be neglected. Nonetheless, the workflow proposed to evaluate the JFWI gradient (Algorithm 1) provides the exact expressions. In the frequency domain,  $G_1$  and  $G_2$  are given by:

$$G_1 = (\lambda_0^r + \delta\lambda^r)^\dagger W^{r2} \frac{\partial B(m_0 + \delta m)}{\partial m_0} (u_0 + \delta u), \quad (\text{A13})$$

$$G_2 = \lambda_0^{d\dagger} W^{d2} \frac{\partial B(m_0)}{\partial m_0} u_0 - \lambda_0^{r\dagger} W^{r2} \frac{\partial B(m_0)}{\partial m_0} u_0. \quad (\text{A14})$$

One can verify that adding  $G_2$  to  $G_1$  results in the same expression as (A10) to (A12).

Another approximation can be made on eqs (A6)–(A9). Terms representing interferences of  $\delta\lambda^d$  and  $u_0$ ,  $\lambda_0^d$  and  $\delta u$  as well as  $\delta\lambda^d$  and  $\delta u$  can be discarded simply because these interferences happen to high-order scattered fields, giving negligible contribution to the gradient (e.g. around 2 per cent of the total energy in the Valhall case).

## APPENDIX B: HIGH-ORDER SCATTERING EFFECT

Formulations of the high-order migration isochrones can be theoretically deduced from expressions (A10) to (A12). Taking the two-reflector case (Fig. 4a) as an example, the scattered component of the adjoint field can be decomposed as infinite series in an increasing order:

$$\begin{cases} \delta\lambda^{r1} = \delta\lambda_1^{r1} + \delta\lambda_2^{r1} + \delta\lambda_{1,2}^{r1} + \dots, \\ \delta\lambda^{r2} = \delta\lambda_1^{r2} + \delta\lambda_2^{r2} + \delta\lambda_{1,2}^{r2} + \dots, \end{cases}$$

where  $r1$  and  $r2$  respectively indicate the early and late reflection phases that provide the adjoint sources. Numbers in the subscript indicate the reflectors at which the scattering takes place. 1: scattering at the shallow reflector; 2: scattering at the deep reflector; 1, 2: successive scatterings at the shallow and deep reflectors. Similarly,

the scattered component of the incident field can be decomposed using the expression  $\delta u = \delta u_1 + \delta u_2 + \delta u_{1,2} + \dots$ . Inserting these decompositions into expressions (A10) to (A12), omitting small-amplitude terms and switching to the time domain, the gradient turns out to be

$$\nabla C_{JFWI} = u_0 \star \lambda_0^d(\text{A}) \quad (\text{B1})$$

$$+ u_0 \star \delta\lambda_1^{r1}(\text{B}) + \delta u_1 \star \lambda_0^{r1}(\text{C}) + \delta u_1 \star \delta\lambda_1^{r1}(\text{D}) \quad (\text{B2})$$

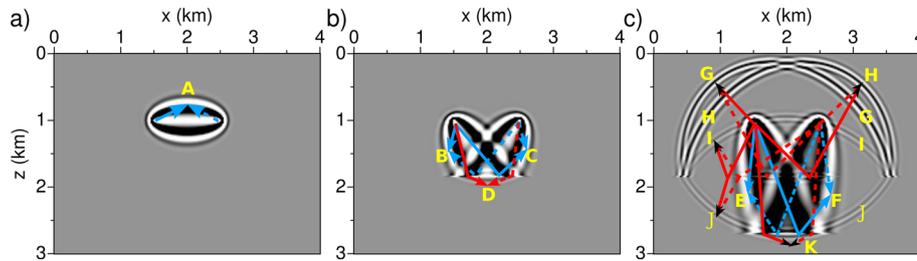
$$+ u_0 \star \delta\lambda_2^{r2}(\text{E}) + u_0 \star \delta\lambda_{1,2}^{r2}(\text{E}) + \delta u_2 \star \lambda_0^{r2}(\text{F}) + \delta u_{1,2} \star \lambda_0^{r2}(\text{F}) \quad (\text{B3})$$

$$+ u_0 \star \delta\lambda_1^{r2}(\text{G}) + \delta u_1 \star \lambda_0^{r2}(\text{H}) + \delta u_1 \star \delta\lambda_1^{r2}(\text{I, J}) \quad (\text{B4})$$

$$+ \delta u_2 \star \delta\lambda_2^{r2}(\text{K}) + \delta u_{1,2} \star \delta\lambda_2^{r2}(\text{K}) + \delta u_2 \star \delta\lambda_{1,2}^{r2}(\text{K}) \quad (\text{B5})$$

$$+ \delta u_{1,2} \star \delta\lambda_{1,2}^{r2}(\text{K}). \quad (\text{B6})$$

Each term represents one or two Fresnel zones labelled by the capital letters in parentheses. They are shown in Fig. B1. Zone A is the classical first-Fresnel zone wave path generated by direct/diving waves (Fig. B1a). Zones B and C are the RWI wave paths associated with the first reflection, whereas Zone D is a high wavenumber isochrone located near the shallow reflector (Fig. B1b). The size of Zone D decreases as higher frequencies are used. The late reflection phase gives rise to other zones E to K (Fig. B1c). Zones E and F are the RWI wave paths associated with the second reflector. Zones I, J, G and H are higher-order migration isochrones. In migration imaging, these isochrones are conventionally considered as migration artefacts, suppressed by destructive interference from different source–receiver couples. Like Zone D, Zone K is located near the deep reflector position and shrinks with higher frequencies.



**Figure B1.** Decomposed JFWI gradient in the two-reflector case, components associated to (a) direct/diving wave, (b) early reflection phase denoted by  $r1$ , and (c) late reflection phase denoted by  $r2$ . Fresnel zones A–K are represented by correlation terms in eqs (B1)–(B6). Solid and dashed arrows denote the ray paths followed by the modelled and adjoint fields, respectively. Blue paths are useful for low-wavenumber imaging unlike the red ones.