

# Full waveform inversion

H. Chauris

---

## 5.1 Overview

Full Waveform Inversion (FWI) is a processing technique to derive quantitative images of the subsurface from seismic measurements. By quantitative, we mean for example P-wave velocity models expressed in m/s, and not only a structural image of the Earth as a classical stack section would provide. The principle is simple: the optimal model is the one for which the computed shot gathers are reproducing the observed shot data. More details on the formulation are provided in section 5.3. Beyond the apparent simplicity, the practical applicability of FWI is a difficult task. This is a non-linear process; the user should provide an initial model; the quality of the final inverted model depends on the reliability of the low frequency content of the observed data. A proper strategy should be established to iteratively determine the model (typically by successively introducing higher frequencies). One also needs to use the adequate wave equation to generate synthetic wave fields and associated

---

This chapter of *Seismic Imaging: a practical approach* is published under Open Source Creative Commons License CC-BY-NC-ND allowing non-commercial use, distribution, reproduction of the text, via any medium, provided the source is cited.

© EDP Sciences, 2019

DOI: 10.1051/978-2-7598-2351-2.c007

shot gathers in order to mimic the physics of wave propagation. Finally, due to the limited data frequency band and limited data acquisition from the surface only, FWI does not necessarily lead to a unique solution. For example, if the user is interested in determining P-velocity and density models, there is an intrinsic trade-off between the two quantities, especially for short offset data. This is not specific to FWI: other imaging techniques suffer from the same effect, but this is visible in the FWI context as FWI is expected to provide quantitative results. A large number of FWI results have been published on real data in seismology, as well as at the exploration scales, at least in the marine case. The use of the technique on land with onshore data, however, has only been proven for a limited number of applications due to the presence of strongly energetic surface waves. While at the geotechnical scale, it has only been tried on a few occasions (among others, Gao *et al.*, 2007; Gélis, 2005; Pérez Solano *et al.*, 2014).

The chapter is organised as follows. We first give a brief overview of the history of Full Waveform Inversion (section 5.2). We then introduce the formalism, limiting the number of equations to the most important ones. We discuss the potential impact as well as the limitations of FWI. This is an important section for anyone who would like to evaluate the potential of FWI on a particular dataset (section 5.4). Finally, we present a few illustrations at the geotechnical scale. For an overview of FWI at the seismological and exploration scales, we refer to Virieux and Operto (2010), and to Fischtner (2010) as well as to three recent didactic papers (Louboutin *et al.*, 2017; Louboutin *et al.*, 2018; Witte *et al.*, 2018).

There are two main elements to consider before applying FWI: the first important aspect is that Full Waveform Inversion considers the *full* wave field (e.g. pressure field or vertical displacement at the receiver position), and does not decompose the data in terms of travel times and amplitude (more details in 5.3). The other important aspect is that FWI can provide high-resolution quantitative results if there is a correct understanding of the main phenomena influencing the wave propagation as well as a proper strategy to iteratively converge towards a meaningful solution. For these reasons, FWI is not an automatic process and is applied after more standard processing such as travel time tomography (Bishop *et al.*, 1985).

## 5.2 History

This section mainly refers to the exploration scale (imaging of the first few kilometres of the subsurface). It is perhaps surprising that the formalism was not derived until the 1980s, through the work of Tarantola and his group in particular (Tarantola, 1984; Mora, 1987). Imaging is an inverse problem: one seeks a model such that the differences between the modelled (computed) data and the observed data are minimum in the least-squares sense. This is a very standard technique in many physical fields. On the geophysical side, the new aspect discovered in the 1980s was a method to update the model, i.e. how to compute the gradient of the misfit

objective function. Similar expressions existed before, for example proposed by Claerbout (1968) but not necessarily leading to quantitative results. The definitions of the objective function and the associated gradient are detailed in section 5.3. In practice, the applications were limited to small size 2D datasets due to limited computer capabilities (more details on this aspect in section 5.4, “memory and CPU requirements”).

New perspectives appeared around 2000, not only because computers were more powerful, but also because Sirgue and Pratt (2004), among others, proposed a practical strategy for the applicability of FWI. This strategy is further discussed in section 5.4 (“how to avoid local minima”). The main idea is to start from low frequency data. Following a few FWI iterations, the inverted model will contain the large-scale structure. Higher frequencies are then progressively introduced and the model is refined accordingly. The strategy of directly considering the full bandwidth may lead to an incorrect solution: it means that the objective function is multi-modal (i.e. contains local minima) due to the non-linearity between the data and the model (more details in section 5.4). Spectacular results were obtained by Pratt and his group, especially on synthetic data for which the observed data contain low enough frequencies to enable the use of the increasing frequency strategy. The 2004 EAGE workshop demonstrated the fundamental role played by the low frequencies during the first iterations, and consequently triggered renewed interest in FWI (Billette and Brandsberg-Dahl, 2005).

Since the early 2000s, FWI has been developed from 2D to 3D, from acoustic to elastic and visco-elastic, as well as in anisotropic contexts, from offshore to onshore datasets. It is clear that the development of computer facilities has supported this trend. FWI remains under development, particularly for multi-parameter estimation (i.e. not only pressure velocity models  $v_p$  from body waves, but also shear velocity models  $v_s$  from surface waves, as well as anisotropy or attenuation parameters). The difficulty is to extract more than one parameter (Operto *et al.*, 2013). Since the 1980s, the challenge has been to incorporate higher frequencies from the data, initially in 2D and now in 3D, as well as more physics (more details in section 5.4).

### 5.3 Formalism

For reasons of clarity, we limit the number of equations presented in this section, and aim to give a physical interpretation of the different quantities introduced. The objective of FWI is to minimize the least-squares misfit function

$$J(m) = \frac{1}{2} \|d_{\text{calc}}(m) - d_{\text{obs}}\|^2, \quad (5.1)$$

where  $m(\mathbf{x})$  is the model to be determined, which is a function of the spatial coordinates  $\mathbf{x} = (x, y, z)$ ,  $d_{\text{obs}}(\mathbf{s}, \mathbf{r}, t)$  the observed data at source position  $\mathbf{s}$ , receiver

position  $\mathbf{r}$  and for time  $t$ . The calculated data  $d_{\text{calc}}(s, \mathbf{r}, t)$  are a function of the  $m$  model and are the solution of the wave equation

$$L(m)d(s, \mathbf{x}, t) = \delta(s - \mathbf{x})\Omega(t), \quad (5.2)$$

$$d_{\text{calc}}(s, \mathbf{r}, t) = d(s, \mathbf{x} = \mathbf{r}, t), \quad (5.3)$$

where  $L$  is the wave equation operator,  $d$  the wave field,  $\Omega(t)$  the seismic source wavelet, and  $\delta$  the dirac distribution. It means that the wave field  $d$  is the solution of the wave equation for a point source located at  $\mathbf{x} = s$  and for a seismic source wavelet. The calculated data are obtained by sampling the wave field  $d$  at the receiver position (Eq. (5.3)). The simplest case corresponds to the constant density acoustic wave equation, with  $m$  being the pressure velocity model  $v_p$ , yielding

$$L(m) = \frac{1}{v_p^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} - \Delta. \quad (5.4)$$

The Laplacian operator  $\Delta$  is the sum of the second-order derivatives in space. There are thus three elements to evaluate the quality of a given model  $m(\mathbf{x})$ :

- Determination of the source wavelet  $\Omega(t)$  (Pratt, 1999);
- Choice of the wave equation operator  $L$ ;
- Resolution of the forward modelling (equations (5.3) and (5.4)) (Louboutin *et al.*, 2017; Louboutin *et al.*, 2018). In section 5.4, we provide more information on this aspect.

The evaluation of  $J(m)$  is a first step, but one usually needs to determine a more suitable model. There are a few recent examples of global search methods (e.g. Sajava *et al.*, 2016). For this strategy, the model space is explored with a Monte Carlo approach or with genetic algorithms. Such an approach would be feasible if the number of model unknowns was small, but this is not typically the case: the model space normally contains millions of unknowns as it is finely discretised along the  $x$ ,  $y$  and  $z$ -axis. As discussed in section 5.4, the CPU cost for solving a single equation (5.2) is the main limiting factor for a global approach (Raknes *et al.*, 2017). The only practical approach is to use a gradient-based inversion, where the model is iteratively determined. The requirements for this are:

- an initial model, usually determined by a standard tomographic approach (Bishop *et al.*, 1985);
- computation of the gradient of the objective function (Plessix, 2006).

Formally, the gradient of the objective function is written as the derivative of the objective function with respect to the model parameters. It has the same (large) size as the model space. This is not a trivial task:  $J$  depends on  $m$  through the dependency on  $d_{\text{calc}}$ , and the relation between  $d_{\text{calc}}$  and  $m$  is given by the wave equation and is clearly non-linear (Eqs. (5.2–5.4)). As mentioned in the introduction, Lailly (1983) and Tarantola (1984) established an efficient way to compute the gradient, with the “adjoint state method” (Plessix, 2006). This approach is related to the

minimization under constraints, with the introduction of Lagrangian multipliers  $\lambda(s, \mathbf{x}, t)$ . In practice, the derivation of the gradient requires three elements:

- computation of the forward wave field  $d$  (Eq. (5.2)) for each source;
- computation of the backward residual wave field  $\lambda$ , also for each source. It is obtained by solving the adjoint wave equation (for the constant density acoustic wave equation, the same wave equation applies) for a source term being the residual wave field at the receiver position  $d_{\text{calc}} - d_{\text{obs}}$ .
- cross-correlation between  $d$  and  $\lambda$ , with a summation over all times, but for fixed spatial  $\mathbf{x}$  positions.

The final gradient is obtained by adding the contribution of all sources. In practice, the derivation of  $d$  and  $\lambda$  are very similar: the gradient thus requires two modelling steps. The CPU cost associated to the cross-correlation is much less than that for solving the forward or backward problem.

Once the gradient is computed, the new model is updated with the typical strategy

$$m_{n+1} = m_n - \alpha \frac{\partial J}{\partial m}, \quad (5.5)$$

where the gradient is  $\partial J / \partial m$  and  $\alpha > 0$  a scalar step length. More advanced methods such as quasi-Newton or Newton approaches take into account the curvature of the objective function (Hessian) for a faster convergence (Nocedal, 1980).

As mentioned before, the user should provide an estimation of the source wavelet. The most popular strategy is to consider the direct arrival between a source and a receiver (Pratt, 1999). In practice, such determination depends on the unknown velocity model in the shallow part: the source wavelet as well as the model itself are together iteratively determined.

By definition, Full Waveform Inversion considers the *full* wave field (e.g. pressure field or vertical displacement at the receiver position) and does not decompose the data in terms of travel times and amplitude. Travel time is a notion associated to high frequency approximation, also known as geometrical optics. FWI is thus considered as a “wave equation approach”, in the sense that it takes into account effects related to finite frequencies, such as the diffraction on a scatterer. In practice, FWI can partly select data around certain windows, for example to only include a zone around the direct arrivals or to remove ground roll. This means that equation (5.1) is modified according to

$$J(m) = \frac{1}{2} \|M(d_{\text{calc}}(m) - d_{\text{obs}})\|^2, \quad (5.6)$$

where  $M$  is a mask in the data domain. With the adjoint state approach to derive the gradient, only a minor change is needed in the so-called adjoint source.

In the next section, we illustrate the behaviour of FWI on a 2D synthetic dataset and discuss the impact of several factors, such as the initial velocity model and the role of the frequency content of the data.

## 5.4 Applicability and practical aspects

Below we provide some practical solutions for a number of key issues that users must address when applying FWI to specific datasets (Virieux and Operto, 2009; Operto *et al.*, 2013; Basker *et al.*, 2016; Raknes *et al.*, 2017).

### Model parameterization

Defining the model parameterization, i.e. the quantity FWI should determine, is the first issue to be considered. This may seem curious but is in fact essential in the case of multi-parameter estimation. We considered a practical example here: suppose that the model unknowns are the pressure velocity  $v_p$  model and the density model  $\rho$ . The P-impedance is the product of the density and velocity ( $I_p = \rho v_p$ ). We compared two strategies:

- In the first case,  $v_p$  and density  $\rho$  were determined by FWI. Then the impedance was deduced from the product of the final  $v_p$  and  $\rho$  models;
- In the second case,  $v_p$  and  $I_p$  were determined by FWI. The density model was determined a posteriori by dividing  $I_p$  by  $v_p$ .

The two approaches do not lead to the same result, as illustrated by Operto *et al.*, 2013. This should not be surprising because for zero-offset data there is a trade-off between a velocity and density contrast associated to diffraction. With larger offsets, the diffractions have different responses. This means that the algorithm converges to different solutions, all in the null space, depending on the initial model.

### Choice of wave equation

Selection of the appropriate wave equation is of course critical. It must be sensitive to the model to be determined. For example, if the attenuation factor is of interest, then the wave equation should contain visco-acoustic or visco-elastic terms. If the analysis of surface waves is important, then acoustic modelling is not sufficient and elasticity should be considered. The determination of a  $v_p$  model from marine acoustic data is a subtler example. An acoustic framework would usually be sufficient in this situation, unless the data contain converted waves, even if only the pressure is recorded at the receiver position. For pure acoustic data, is there a need to consider density? The amplitudes of the *transmitted* (diving) arrivals are not sensitive to density, whereas this is not the case for *reflected* waves. Such questions are still not fully solved.

### Time versus frequency

Two main approaches are possible to solve the direct problem (Eq. (5.4)) and the inverse problem: either in the time or the frequency domain (Virieux and Operto, 2009; Raknes *et al.*, 2017). Many scientific papers were published on this topic in the 2010s. From a theoretical point of view, the two approaches lead to the same

results if dense frequencies cover the whole spectrum. The Parseval relationship indeed indicates that

$$J(m) = \frac{1}{2} \|d_{\text{calc}}(m)[t] - d_{\text{obs}}[t]\|^2 = \frac{1}{2} \|d_{\text{calc}}(m)[\omega] - d_{\text{obs}}[\omega]\|^2, \quad (5.7)$$

where the dependency on  $\omega$  indicates the Fourier transform from time to angular frequency. We omit the dependency on the source and receiver positions. The result is different if one considers only a small number of discrete frequencies. There is currently some consensus that:

- for many frequencies, the most efficient approach is time implementation.
- for a limited number of frequencies in 3D, the time approach is also preferred, followed by a discrete Fourier transform.
- for a limited number of frequencies in 2D, the frequency approach is more suited, but one has to solve a large (sparse) linear system, for example with a LU decomposition. Once this LU decomposition is performed, the imaging is very efficient.

Beyond the time versus frequency approaches, the most important aspect is to avoid local minima.

### FWI = migration + tomography?

This title is a reference to a publication by Mora in 1989. For an adequate application of FWI, it is essential to understand how FWI behaves. In particular, FWI has two main “modes”, i.e. different methods of updating the long and short wavelength components of the velocity model:

- The *tomographic* mode means that the long wavelengths of the velocity model are updated, with an influence on the kinematics of wave propagation. In practice, this is the difficult part of FWI as this process is non-linear: if one multiplies the velocity model by 10%, then the data, for example reflected waves, are recorded at a different (shorter) time;
- The *migration* mode indicates that if the data, up to a first-order approximation, linearly depend on the short wavelength components of the velocity model. After linearization, the data only contain reflected waves. If the velocity perturbations are increased by 10%, then the recorded pressure or displacement at the receiver is also multiplied by the same factor.

There is no strict limit between the tomographic and migration modes in FWI, but if multiple frequencies are used for the inversion, the migration will dominate and it will be difficult to update the long wavelengths. The linearized version of FWI is called “iterative least-squares migration” (LSM). Standard migration corresponds to the first iteration only.

## Avoiding local minima

A standard FWI application, without careful pre-processing or meticulous parameter selection, would not typically lead to the expected solution. The algorithm may easily converge into a *local* minimum, which is possibly far from the *global* minimum (Bunks *et al.*, 1995). To analyse this, it is useful to refer to the tomographic and migration modes explained in the previous section.

Seismic data are oscillatory signals by nature, as the zero-frequency part is not recorded. The FWI misfit function (Eq. (5.1)) evaluates the least-squares distance between calculated and observed data, both being oscillatory signals. If the initial velocity model provides the correct kinematics, then FWI mainly operates in the migration mode and the events are correctly positioned. If the initial kinematics are not correct, then the algorithm will match incorrect phases between the synthetic and observed data. This is known as the cycle-skipping effect: it means that the algorithm converges in a local minimum (Bunks *et al.*, 1995).

We illustrated this effect using a very simple example, representative of what happens in practice. We considered a single trace. The blue trace corresponds to the observed data (Figure 5.1). An incorrect velocity model has an effect on the kinematic (position) and on the dynamic (amplitude) of the events. We studied here the kinematic effect because of the non-linearity. The red trace is the one in the inverted model (Figure 5.1, top). It does not perfectly match the observed trace because of the presence of noise. We then shifted the red trace to mimic the kinematic effect, for half a period and for a period (Figures 5.1, middle and bottom).

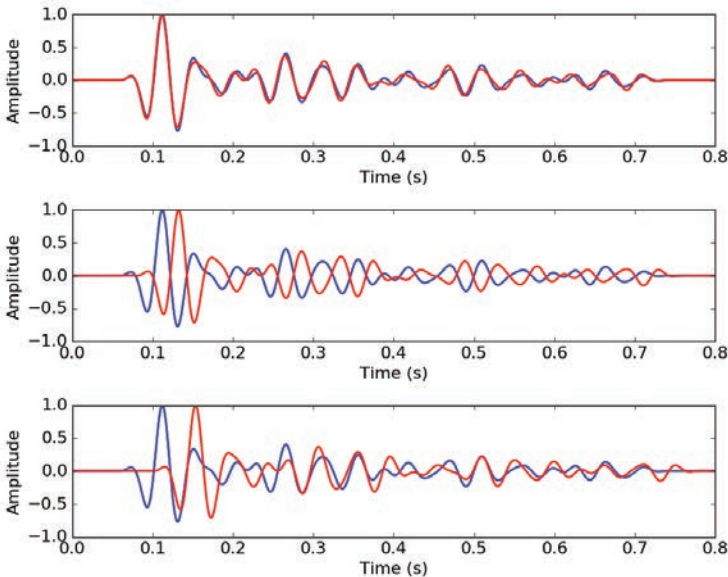


Figure 5.1 Observed trace (blue) and synthetic trace (red), for different shift values applied to the red trace (from top to bottom: 0, -21 and -42 ms, respectively).

For each possible shift, we computed the standard least-squares misfit (Eq. (5.1)) between the observed (blue) and synthetic trace (red) (Figure 5.2, blue solid line). The objective function oscillates, with a minimum for zero-shift as expected (blue dot). The shift of half a period corresponds to a local maximum (red dot) as the two signals are out of phase (Figure 5.1, middle), whereas the shift of a period is associated to a local minimum (cyan dot) for which the blue and red traces are mainly in phase. The maximum possible shift to ensure a convergence towards the global minimum (blue dot) is indicated by the red dot. This phenomenon is known as the *cycle-skipping effect*. For the same signals, if the central frequency had been halved, then the shape of the objective function would be within the blue dashed line (Figure 5.2). It means that the basin of attraction is twice as large (Bunks *et al.*, 1995).

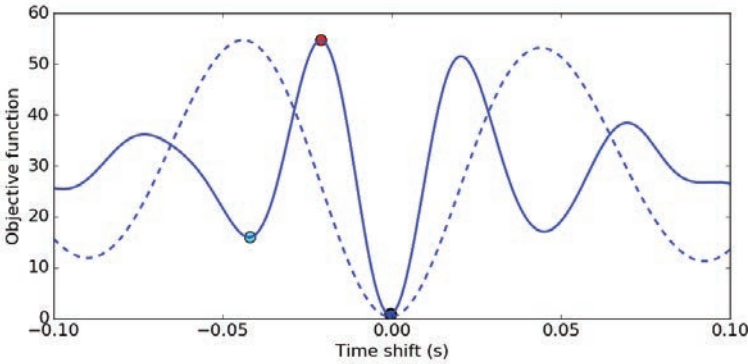


Figure 5.2 Shape of the objective function for different shift values. The blue, red and cyan dots are associated to the traces in Figure 5.1 (from top to bottom). When the central frequency is divided by two, the objective function has a basin of attraction (dashed line) that is twice as large.

In conclusion, there are two main strategies to ensure a proper convergence:

- Start with an initial velocity model that is not too far away from the correct solution.
- Use low frequencies at first, then progressively increase the frequency content. Low frequencies have larger basins of attraction than higher frequencies, as illustrated in Figures 5.1 and 5.2. Different authors have proposed practical rules (Sirgue and Pratt, 2004). The current practice is to consider a frequency band  $[f_0 - f]$ , where  $f_0$  is the minimum reliable frequency and where  $f$  increases progressively. The alternative is to let  $f_0$  also increase. Note that for a given frequency band, at least a few non-linear iterations are performed before modifying the frequency range. Then the same process is repeated.

Beyond these two points, the redundancy of the data is an important aspect to constrain the inversion. The *null space* refers to equivalent models associated to the

same misfit value. The reduction of the size of the null space can be achieved by introducing log data or a priori information.

### Memory and CPU requirements

FWI is an expensive process, both in terms of memory and CPU requirements. Let's consider a simple case where the forward and backward wave equations are solved with a standard Finite Difference scheme (Virieux, 1986). We applied the constant density acoustic wave equation (Eq. (5.4)). In the simplest case, one first needs to define a regular grid in time and space; the second-order time derivative and Laplacian operator are approximated in 2D by  $[+1, -2, +1] / (\Delta t)^2$ ,  $[+1, -2, +1] / (\Delta x)^2$  and  $[+1, -2, +1] / (\Delta z)^2$ , where  $\Delta t$ ,  $\Delta x$  and  $\Delta z$  are the grid increments along the time and space axis. They cannot be chosen arbitrarily and should satisfy the conditions

$$\Delta x = \Delta z \leq \frac{1}{10} \frac{v_{min}}{f_{max}} \quad (5.8)$$

$$\Delta t \leq \frac{1}{\sqrt{\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta z}\right)^2}} \frac{1}{v_{max}} \quad (5.9)$$

The first equation is the *dispersion* condition: the space discretization should be 10 times smaller than the maximum wavelength  $v_{min} / f_{max}$ , obtained for the minimum expected velocity value in the model and for the maximum frequency. This is more restrictive than the Nyquist condition (more than 2 points per wavelength) as this is not a static representation but a dynamic one. If the condition is not satisfied, then numerical dispersion is observed on the signals.

Equation (5.9) is related to the *stability* condition and is a function of the maximum velocity value. If the condition does not hold, the scheme is not stable and does not provide any useful information.

Let's consider an example for a target at 1 km depth, with desired maximum frequencies at 10 Hz and 100 Hz, for typical velocities between 300 m/s to 3000 m/s.

**Table 5.1** 2D discretization in space and time for the same extreme velocity values and for two different maximum frequency values, as well as the storage requirement for a single wave field (right column).

	$\Delta x$ (m)	$\Delta z$ (m)	$\Delta t$ (ms)	Storage (GB)
$f_{max} = 10$ Hz $v_{min} = 300$ m/s $v_{max} = 3000$ m/s	3.0	3.0	0.70	3.2
$f_{max} = 100$ Hz $v_{min} = 300$ m/s $v_{max} = 3000$ m/s	0.3	0.3	0.07	3200

Note that  $f_{min}$  does not play a role here. For  $f_{max} = 100$  Hz, the discretization is very fine. Suppose that the image is 1 km deep and 5 km long, with a record up to 1 s. The number of samples can be easily computed. For a simple example, the storage for a single wave field is evaluated in Table 5.1, right column. This is feasible for the 10 Hz case, but not for the 100 Hz. The important aspect is that the storage depends on  $f_{max}^3$  in 2D and  $f_{max}^4$  in 3D (with the introduction of  $\Delta y$ ). In practice, the maximum processed frequency is limited because of limited memory and CPU resources.

## 5.5 Examples

Below we present an example of the capabilities of FWI, as well as its applicability for a 2D synthetic dataset. The choice of synthetic versus real data is driven by the a posteriori control of the results. At the end of the section we discuss the additional difficulties related to real data (Basker *et al.* 2016).

### Geophysical context

The synthetic data is inspired from a real data acquisition in the geotechnical context (Deidda and Balia, 2001). It consists of three main parts: an unconsolidated sediment layer with silt and clay in the deeper part, a concrete layer at around 4-5 m depth and a compacted back fill material in the shallow part (Figure 5.3). For the analysis, we use the same synthetic model (with P-wave velocity and density) as presented in Pérez Solano *et al.*, 2014. Despite the apparent relatively simple structure, the difficulty in this model comes from its lateral variations.

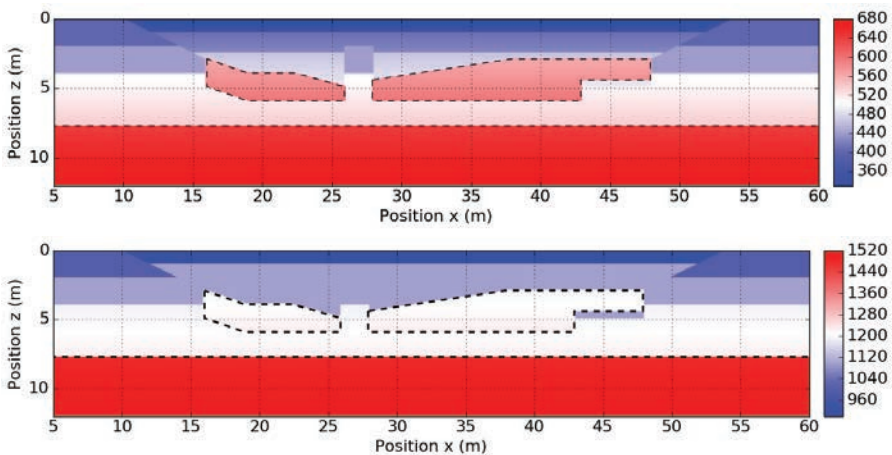


Figure 5.3 Exact  $v_p$  (top, in m/s) and density (bottom, in  $\text{kg/m}^3$ ) models, from (Pérez Solano *et al.*, 2014).

For simplicity, we have used acoustic modelling. With elasticity, the presence of surface waves largely complicates the imaging strategy (see below in the discussion part) (Brossier *et al.*, 2009). The modelling is based on a staggered grid strategy with a velocity-stress formulation (Virieux, 1986). The particle velocity is converted at the surface into a pressure wave field. The acquisition geometry consists of regular shots every 1 m along the profile and receivers at  $\pm 10$  m around the source position, knowing that the maximum depth to be investigated is at around 10 m. It means that in this configuration, the interesting information is contained in the reflected waves and not in the transmitted waves for which longer offsets would be needed (see Chapters 2 and 3 on “refraction surface” and “seismic tomography”).

### Inversion results

For the definition of the initial model (here the  $v_p$  and density models), we used a simple gradient model, where the parameters linearly increase with depth (Figure 5.4). This model is consistent with a smooth version of the exact model on the left or right parts, but differs in the central part (Figure 5.3). The only modification brought to this gradient model was the introduction of the shallow layer (first 50 cm) with a dipping interface around  $x = 12$  and  $53$  m (Figure 5.4).

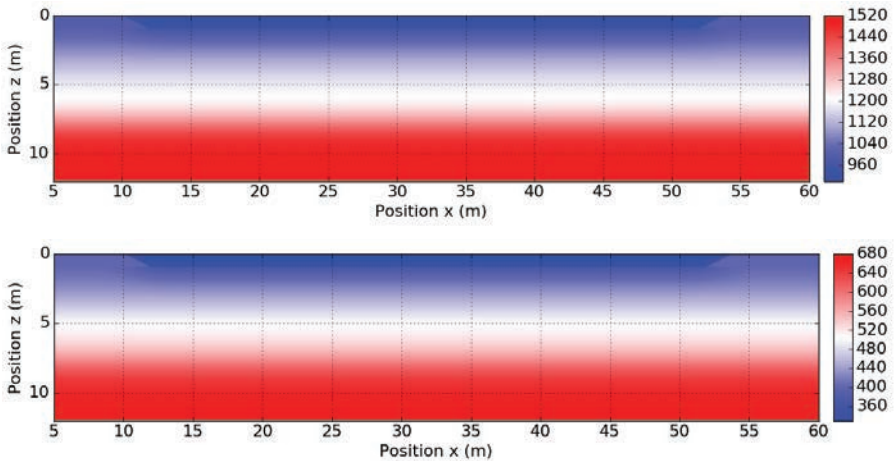


Figure 5.4 Initial  $v_p$  (top, in m/s) and density (bottom, in  $\text{kg/m}^3$ ) models.

In the first strategy (A), the minimum and maximum data frequencies were 30 and 300 Hz, respectively. All frequencies were inverted simultaneously. Note that the central frequency was about 120 Hz. The non-linear minimization was obtained with a standard quasi-Newton scheme (L-BFGS, (Nocedal, 1980)), for which the last 5 iterations are used to build the inverse of the Hessian to speed up convergence

and more importantly to equalize the contributions with depth (at the first iteration, only the very shallow part is updated). Note that only the P-wave model was updated: the density model remained unchanged, even if the exact density model differs from the initial model (Figures 5.1 and 5.2, bottom). After 30 non-linear iterations, the final model displays some interesting structures, for example around  $x = 27$  m (Figure 5.5). We superimposed the main elements in a dashed black line extracted from the exact model (Figure 5.3). But this final model does not display homogeneous velocities within the cemented structures. The bedrock is not horizontal and its depth is slightly underestimated in the central part. In fact, the FWI algorithm converges to a local minimum as the initial model (Figure 5.4) is too far from the correct one for the frequency range of the data.

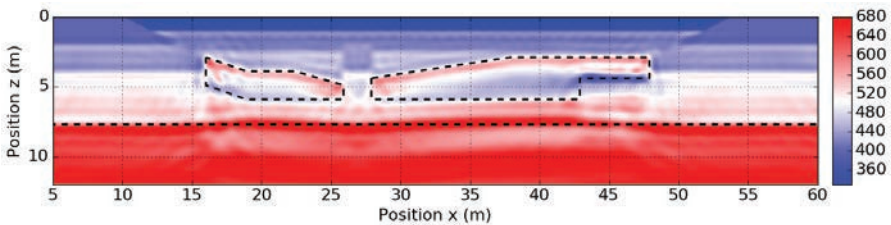


Figure 5.5 Inverted  $v_p$  model (m/s) after 30 iterations, with  $f_{min} = 30$  Hz and  $f_{max} = 300$  Hz. The density model is not updated.

We thus developed the second strategy (B), consisting of two steps: (1) 15 non-linear inversions with frequencies between 30 and 60 Hz, and then (2) 30 non-linear inversions with frequencies between 30 and 300 Hz as before. The intermediate model after 15 iterations is smooth and contains higher velocities in the cemented structure, even if the limit is not clear (Figure 5.6, top). From that result, we increased the frequency range. The final model (Figure 5.6, bottom) provides a very satisfactory result. The velocity model is much more homogeneous and the main interfaces are correctly positioned. Also, the low (blue) velocity anomaly around  $x = 45$  m and  $z = 5$  m is well retrieved. Note that some oscillations were created to compensate for density contrasts as the density model remains unchanged.

This is clearly visible on vertical sections extracted from the inverted and exact models (Figure 5.7). There is a good agreement between the inverted (blue) and exact (red) velocity models (Figure 5.7, left) and a quantitative match between the impedance models  $\rho v_p$  (Figure 5.7, right), where  $v_p$  is the inverted model and  $\rho$  the initial model. For example, for  $x = 30$  m, the velocity jump at 1 m depth is overestimated in  $v_p$  and correctly determined in  $\rho v_p$ . Here, the image mainly comes from the analysis of *reflected* waves; this is why the *impedance* section is preferred compared to the *velocity* section.

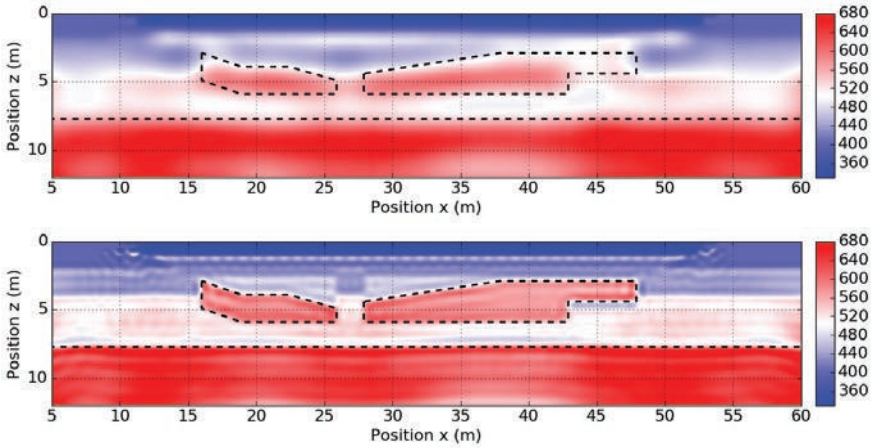


Figure 5.6 Inverted  $v_p$  models (m/s); top: after 15 iterations with  $f_{min} = 30$  Hz and  $f_{max} = 60$  Hz; bottom: after 30 iterations with  $f_{min} = 30$  Hz and  $f_{max} = 300$  Hz, starting from the model on the top. As for Figure 5.5, the density model is not updated.

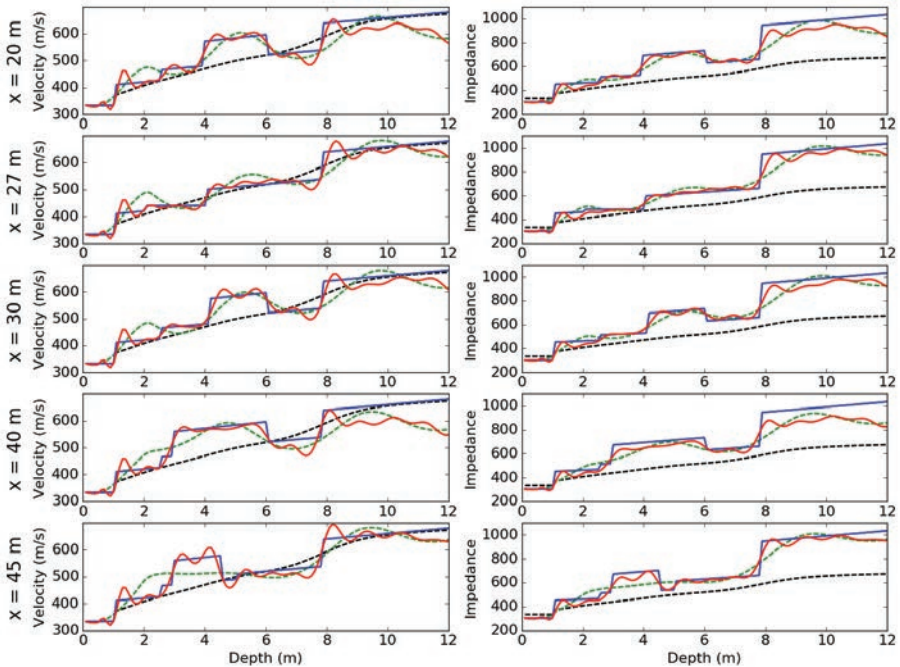


Figure 5.7 Vertical selections across  $v_p$  models (m/s), left, and impedance models  $\rho v_p$  (in  $10^3$  kg/s/m<sup>2</sup>), right, for  $x = 20, 27, 30, 40$  and  $45$  m, from top to bottom, respectively. Dashed black line: initial model; dashed green line: intermediate model (Figure 5.6, top); solid red line: inverted model (Figure 5.6, bottom) and solid blue line: exact model. As for Figure 5.5, the density model is not updated.

The FWI objective function decreases in the two steps (Figure 5.8). Here, we have shown all iterations, even though the L-BFGS algorithm does not consider all of them as real iterations. The reason is that each iteration has the same CPU cost. The objective function is normalized by its maximum value. After the first step (iteration 15), the objective function suddenly increases dramatically, even if the models are the same, simply because the frequency content of the data is modified. Most of the decrease is observed during the first iterations, while the final iterations are still important to retrieve the details in the model.

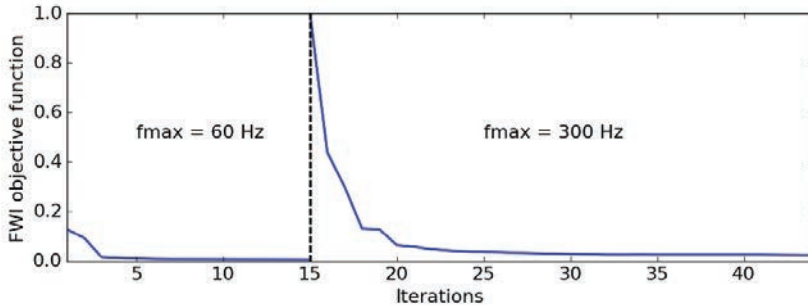


Figure 5.8 Shape of the FWI objective function (normalised to 1) for the first 15 iterations (with  $f_{min} = 30$  Hz and  $f_{max} = 60$  Hz) and for the subsequent 30 iterations (with  $f_{min} = 30$  Hz and  $f_{max} = 300$  Hz).

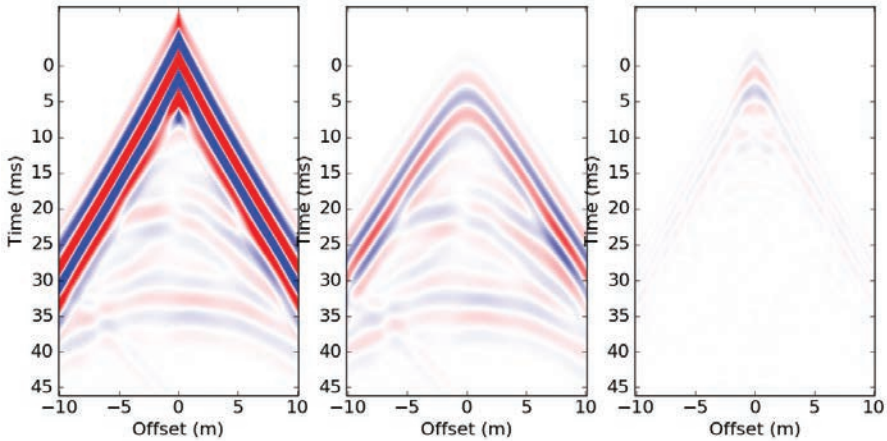


Figure 5.9 Observed data (left), initial data residuals (middle) and final data residuals (right), all represented at the same scale and for a shot at position  $x = 32$  m. Here the pressure is displayed.

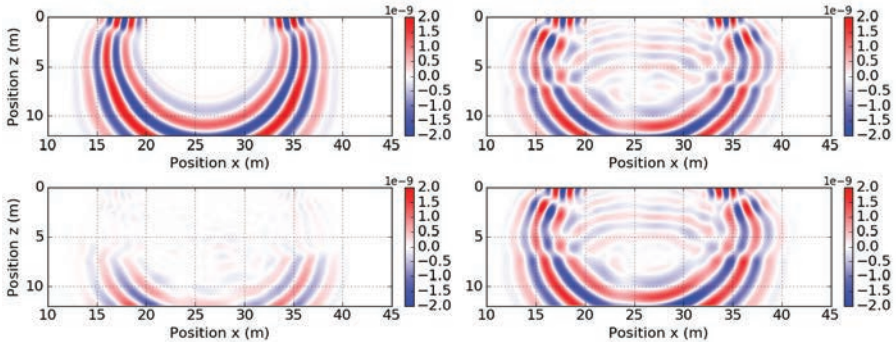


Figure 5.10 Snapshots (from the same source, the same time and at the same scale); top left: in the initial model (Figure 5.4, top); top right: in the inverted model (Figure 5.6, bottom); bottom right: in the exact model (Figure 5.3, top) and bottom left: differences between the snapshots on the right.

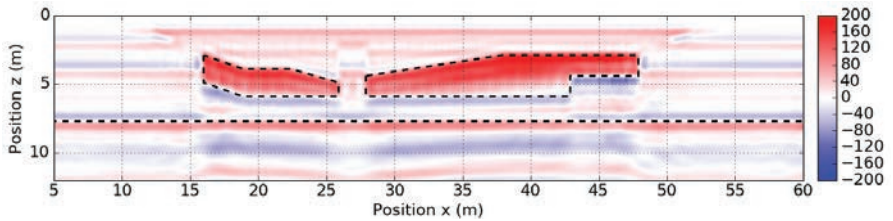


Figure 5.11 Least-squares migration results (30 iterations), expressed in  $v_p$  perturbations (in m/s) and computed in the model from Figure 5.6, top.

Most of the energy in the shot gather for the shot position at  $x = 32$  m is contained in the first arrival (Figure 5.9, left). As the initial model for the first meter is relatively correct, the initial residuals are related to the first reflection at 1 m depth as well as deeper reflections (Figure 5.9, middle). After inversion, the misfit is largely reduced. Note that in the imaging algorithm, the shallowest part is not updated, indicating that the largest final residuals are associated to the first event (Figure 5.9, right).

It is interesting to analyse the snapshots in the initial, exact and inverted models (Figure 5.10). This is a representation of the wave field for a given propagation time, here  $t = 22.9$  ms. The snapshot for the exact model is not available on real data. After minimization, there is an excellent match between the up-going waves. The down-going waves differ (Figure 5.10, bottom left) as they are not recorded at the surface: the model is not constrained below the deepest reflectors.

Least-squares migration (LSM) is a linearized version of FWI. It only deals with reflected waves and ignores transmitted waves, refracted waves, multiples, etc.

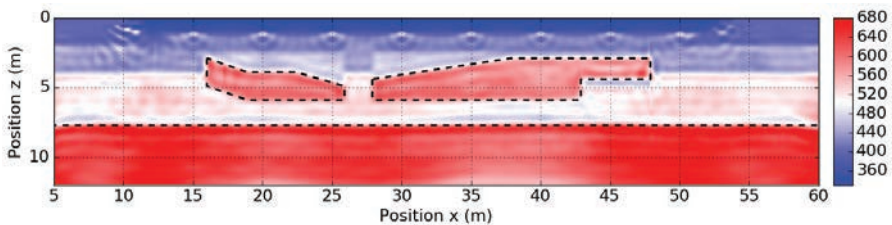
In practice, a pre-processing step should be applied to only select reflected waves from the observed data. We ran LSM in the same initial model as for the second FWI scheme (Figure 5.6, top). Here, the velocity perturbation is updated, not the initial macro-model. It means that the Green's functions are always computed in the same model. They cannot generally be saved in memory as their size depends on the number of sources, the spatial size of the model and the number of time samples. The final result is an oscillatory signal (Figure 5.11), here expressed in terms of velocity perturbations. We used a quantitative LSM approach, but migration algorithms are often qualitative, for a structural interpretation. Here, with the quantitative approach, it is possible to sum the macro-model (initial model, Figure 5.6, top) and the velocity perturbation, yielding a very similar result as the one provided by FWI. The reason for this is that most of the information is contained in the reflected energy, and not in the transmitted waves already explained by the initial model.

### Sensitivities

We tested the sensitivity of FWI with respect to the

- acquisition geometry;
- choice of the model parameter to be inverted.

We repeated the same process as in Figure 5.6, except that we selected a source every 5 m, instead of every 1 m. The final FWI is similar (Figure 5.12) and the differences are localized in the shallow part, around 1 to 2 m depth, every 5 m along the  $x$  axis. These zones with higher (white) velocities are the imprint of the acquisition design: for datasets that are too sparse, the model cannot be properly reconstructed because of aliasing effects (Gray, 2013). These effects would have been even stronger if a mask had not been applied to prevent any updates in the first meter.



**Figure 5.12** Same as for Figure 5.4, bottom, but for sources every 5 m instead of every 1 m as in all other examples. The differences are mainly visible in the shallow part.

In the previous example, the velocity was updated, while the density model remained fixed, even if the exact Earth (Figure 5.3) contained both velocity and density variations. We discussed the fact that the best reconstruction was obtained by the impedance (product of velocity by density). Here, we ran FWI for a fixed

velocity (from Figure 5.6, top): only the density model is updated. After 30 non-linear quasi-Newton FWI inversions, the final density model is highly oscillating and tends towards either low or high values (Figure 5.13). The cemented structure cannot be interpreted. As expected, this means that a density model cannot be replaced by a velocity model. Velocity has an influence on the kinematics (event positioning) and dynamics (amplitude) of wave propagation, whereas density only plays a role with regard to amplitude. Thus, a pure density-FWI scheme cannot update an incorrect kinematic part. The objective function did not reach a value as low as before (Figures 5.8 and 5.14). Note that the value of the objective function itself is not very significant; the data residuals should be analysed as in Figure 5.10.

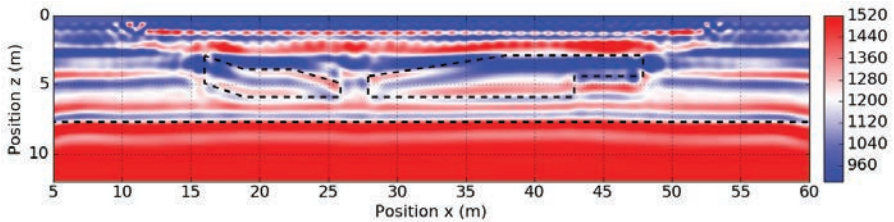


Figure 5.13 As for Figure 5.6 bottom, except for the density model (in  $\text{kg/m}^3$ ). The velocity model has not been updated (Figure 5.6, top).

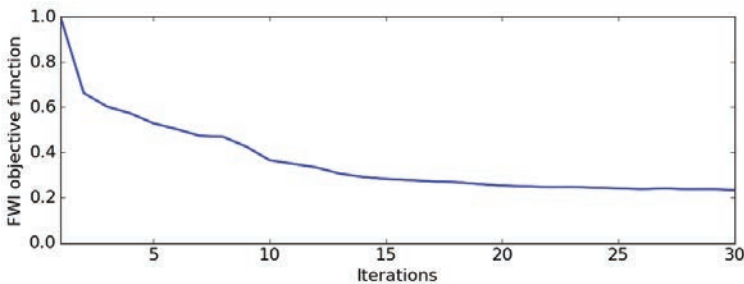


Figure 5.14 As for Figure 5.8, but where only the density model has been updated.

## Towards applications on real data

Before concluding, we discuss the additional steps required for applications on real data. Most of the published results are related to the marine environment. Water in the shallow part prevents energetic surface waves, which remains a challenging problem for FWI (Brossier *et al.*, 2009).

As FWI is a data misfit approach, the amplitudes play an important role. We discussed the importance of density contrasts in the synthetic example. Diving waves are sensitive to velocities but not to density. This is one of the main reasons why they are

classically used in FWI. In practice, a window is applied around the transmitted waves (direct and diving/refracted) so that only these waves are considered in the FWI, while removing the surface waves and reflected/diffracted waves. This means that large offsets should be recorded. The other reason to consider transmitted waves is that the objective function is more convex than for the reflected waves. This avoids the problem of falling into a local minimum.

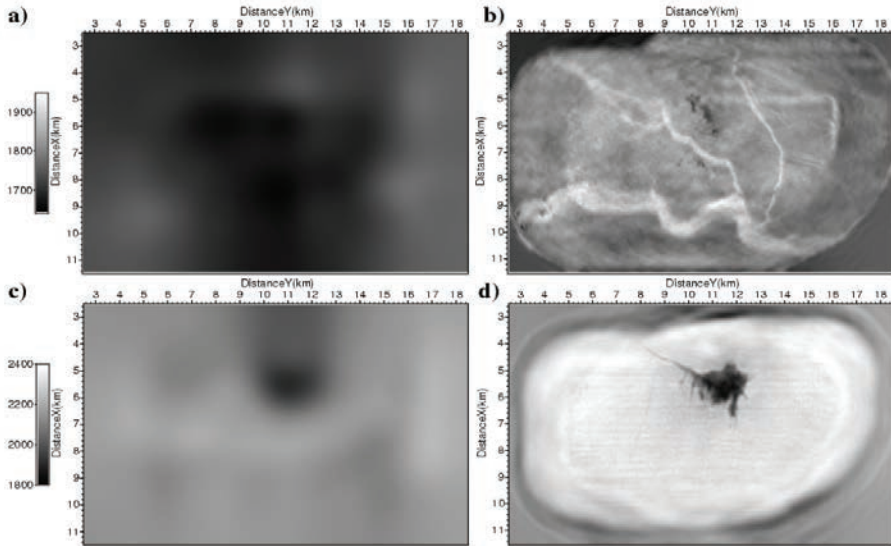


Figure 5.15 Depth sections in 3D velocity models on the marine Valhall OBC data set, obtained by ray-based reflection tomography (left), followed by FWI (right), for depths  $z = 150$  m (top) and 1050 m (bottom). In the FWI images, channels (top) and the presence of gas (bottom) are clearly visible (Sirgue et al., 2009).

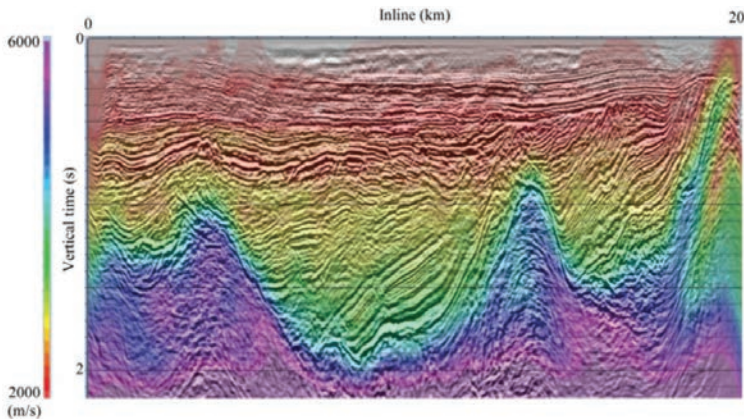


Figure 5.16 Migrated section superimposed on the FWI velocity model, in a land example (Baeten et al., 2008).

We present here two examples of impressive FWI results, based on marine (Figure 5.15) and land (Figure 5.16) datasets. In the former, the starting model was obtained by a conventional ray-based reflection travel time tomography (Bishop *et al.*, 1985). FWI was then used to refine the model, leading to clear channels in the shallow part (in white in Figure 5.15-b) and to the presence of gas in the deeper part (in black in Figure 5.15-d). It is possible to see gas leakage along the main faults.

The second example, in a land acquisition context (Inner Mongolia, China), was more challenging due to the presence of highly energetic surface waves (Baeten *et al.*, 2008; Brossier *et al.*, 2009). Here, they were filtered out in a pre-processing step. The initial velocity was derived from travel time tomography and is mainly a 1D model (not represented here). Specific attention was paid to the preservation of energy in the dataset between 1.5 to 2 Hz: this was a crucial step in the FWI construction of the velocity model.

In both cases, the acoustic FWI largely outperforms standard travel time tomography. More work is needed in future to consider higher frequencies and more complex physics.

## 5.6 Conclusions

Full Waveform Inversion is a technique to obtain seismic quantitative images of the subsurface. However, there are a number of difficulties in terms of its applicability. In particular, a low to high frequency strategy may need to be applied. A carefully considered initial model as well as suitable pre-processing steps must be determined. Finally, the multi-parameter estimation (beyond P-wave estimation) is still an active area of research.

In future, FWI will hopefully become “full”: currently, windows are applied to the data to select transmitted waves, for example, or to remove surface waves. These waves contain interesting information on the subsurface (Pérez Solano *et al.*, 2014). The question is to know how to efficiently extract it.

There have been a number of possible alternatives to FWI, for example the Adaptive Waveform Inversion strategy (Warner and Guasch, 2016). Usually, the objective is to remove the cycle-skipping effects or to be able to consider the reflected waves more easily (as in the Reflection Waveform Inversion approach (Zhou *et al.*, 2015)). Another alternative is to split the problem into two parts: the estimation of the macro-model containing the main structures (migration velocity analysis or tomography), and of the model perturbation (migration) (Symes, 2008; Chauris and Cocher, 2017).

Despite larger computation capabilities and memory allocation, FWI can only process limited frequencies in 3D and usually follows a deterministic approach: starting from an initial model, the model is iteratively updated. In future, uncertainties around the final solutions should be properly evaluated, taking into account

uncertainties regarding the data as well as the non-perfect knowledge of the physics of wave propagation through the Earth. On the geotechnical scale, FWI is still in its early development stage, but it has great potential for the future.

## References

- Baeten G., De Maag J.W., Plessix R.-E., Klaassen R., Qureshi T., Kleemeyer M., Ten Kroode F., Rujie Z., 2013, The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study, *Geophysical Prospecting*, 61 (4), 701–711, DOI: 10.1111/1365-2478.12010.
- Basker B., Rüger A., Deng L., Jaramillo H., 2016, Practical considerations and quality control for an FWI workflow, *The Leading Edge*, 35 (2), 151–156, DOI: 10.1190/tle35020151.1.
- Billette F., Brandsberg-Dahl S., 2005, The 2004 BP velocity benchmark.: *67th Annual Conference and Exhibition*, EAGE, Extended Abstracts, B035.
- Bishop T.N., Bube K.P., Cutler R.T., Langan R.T., Love P.L., Resnick J.R., Shuey R.T., Spindler D.A., Wyld H.W., 1985, Tomographic determination of velocity and depth in laterally varying media, *Geophysics*, 50 (6), 903–923, DOI: 10.1190/1.1441970.
- Brossier R., Operto S., Virieux J., 2009, Seismic imaging of complex on-shore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, 74, WCC105–WCC118, DOI: 10.1190/1.3215771.
- Bunks C., Salek F.M., Zaleski S., Chavent G., 1995, Multiscale seismic waveform inversion, *Geophysics*, 60 (5), 1457–1473, DOI: 10.1190/1.1443880.
- Chauris H., Cocher E., 2017, From migration to inversion velocity analysis, *Geophysics*, 82 (3), S207–S223, DOI:10.1190/GEO2016-0359.1
- Claerbout J.F., 1968, Synthesis of a layered medium from its acoustic transmission response, *Geophysics*, 33 (2), 264–269, DOI: 10.1190/1.1439927.
- Deidda G.P., Balia R., 2001, An ultrashallow SH-wave seismic reflection experiment on a subsurface ground model, *Geophysics*, 66, 1097–1104, DOI: 10.1190/1.1487057.
- Fichtner A., 2010, *Full Seismic Waveform Modelling and Inversion*, Springer Verlag.
- Gao F., Levander A., Pratt R.G., Zelt C.A., Fradelizio G., 2007, Waveform tomography at a groundwater contamination site: surface reflection data, *Geophysics*, 72, G45–G55, DOI: 10.1190/1.2752744.
- Gélis C., 2005. Inversion des formes d'onde élastique dans le domaine espace-fréquence en deux dimensions. Application à la caractérisation de la subsurface

- dans le cadre de la détection de cavités souterraines. PhD thesis: Université de Nice-Sophia Antipolis.
- Gray S.H., 2013, Spatial sampling, migration aliasing, and migrated amplitudes, *Geophysics*, 78 (3), S157-S164, DOI: 10.1190/geo2012-0451.1.
- Lailly P., 1983. The seismic inverse problem as a sequence of before-stack migrations, in J. Bednar, ed., *Conférence on inverse scattering: Theory and application*, Society for Industrial and Applied Mathematics, 206–220.
- Louboutin M., Witte P., Lange M., Kukreja N., Luporini F., Gorman G., Herrmann F.J., 2017, Full-waveform inversion, Part 1: Forward modelling, *The Leading Edge*, 36 (12), 1033–1036, DOI: 10.1190/tle36121033.1.
- Louboutin M., Witte P., Lange M., Kukreja N., Luporini F., Gorman G., Herrmann F.J., 2018, Full-waveform inversion, Part 2: Adjoint modelling, *The Leading Edge*, 37 (1), 69–72, DOI: 10.1190/tle37010069.1.
- Mora P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, 52 (9), 1211-1228, DOI: 10.1190/1.1442384.
- Mora P., 1989. Inversion = migration + tomography, *Geophysics*, 54 (12), 1575-1586, DOI: 10.1190/1.1442625.
- Nocedal J., 1980. Updating quasi-Newton matrices with limited storage, *Mathematics of Computation*, 35, 773–782, DOI: 10.1090/S0025-5718-1980-0572855-7.
- Operto S., Gholami Y., Prioux V., Ribodetti A., Brossier R., Metivier L., Virieux J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice, *The Leading Edge*, 32 (9), 1040-1054, Doi: 10.1190/tle32091040.1.
- Pérez Solano C.A., Donno D., Chauris H., 2014, Alternative waveform inversion for surface wave analysis in 2-D media, *Geophysical Journal International*, 198 (3), 1359–1372, DOI:10.1093/gji/ggu211.
- Plessix R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, 167(2), 495–503, DOI: 10.1111/j.1365-246X.2006.02978.x.
- Pratt R.G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: *Geophysics*, 64 (3), 888-901, DOI: 10.1190/1.1444597.
- Raknes E.B., Arntsen B., 2017, Challenges and solutions for performing 3D time-domain elastic full-waveform inversion, *The Leading Edge*, 36 (1), 88–93, DOI: 10.1190/tle36010088.1.
- Sajeva, A., Aleardi M., Stucchi E., Bienati N., Mazzotti A., 2016, Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the Marmousi model, *Geophysics*, 81 (4), R173-R184, DOI: 10.1190/geo2015-0198.1.

- Sirgue L., Pratt R.G., 2004, Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies: *Geophysics*, 69 (1), 231-248, DOI: 10.1190/1.1649391.
- Sirgue L., Barkved O.I., Van Gestel J.P., Askim O.J., Kommedal J.H., 2009. 3D Waveform Inversion on Valhall Wide-azimuth OBC: Session WEM and WAZ Velocity Model Building, *71st EAGE Conference and Exhibition incorporating*, DOI: 10.3997/2214-4609.201400395.
- Symes W.W., 2008, Migration velocity analysis and waveform inversion, *Geophysical Prospecting*, 56 (6), 765-790, DOI: 10.1111/j.1365-2478.2008.00698.x.
- Tarantola A., 1984, Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, 49 (8), 1259-1266, DOI: 10.1190/1.1441754.
- Virieux J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method, *Geophysics*, 51 (4), 889-901, DOI: 10.1190/1.1442147.
- Virieux J., Operto S., 2009, An overview of full waveform inversion in exploration geophysics, *Geophysics*, 74 (6), WCC1-WCC26, DOI: 10.1190/1.3238367.
- Warner M., Guasch L., 2016, Adaptive waveform inversion: Theory, *Geophysics*, 81 (6), R429-R445, DOI: 10.1190/geo2015-0387.1.
- Witte P., Louboutin M., Lensink K., Lange M., Kukreja N., Luporini F., Gorman G., Herrmann F.J., 2018, Full-waveform inversion, Part 3: Optimization, *The Leading Edge*, 37 (2), 142-145, DOI: 10.1190/tle37020142.1.
- Zhou W., Brossier R., Operto S., Virieux J., 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation, *Geophysical Journal International*, 202 (3), 1535-1554, DOI: 10.1093/gji/ggv228.