Comparison of two- and three-dimensional full waveform inversion imaging using wide-angle seismic data from the Deep Galicia Margin

Bhargav Boddupalli1*, Tim A. Minshull1, Joanna Morgan2, Gaye Bayrakci3 and Dirk Klaeschen4

1School of Ocean and Earth Science, University of Southampton, SO14 3ZH, UK.
2Department of Earth Sciences and Engineering, Imperial College London, SW7 2AZ, UK.
3National Oceanography Centre Southampton, SO14 3ZH, UK.
4GEOMAR, Helmholtz Centre for Ocean Research Kiel, 24148, Germany.

*now at Institute for Geophysics, University of Texas at Austin, USA

Corresponding author: Bhargav Boddupalli

Email address: bhargav@ig.utexas.edu

In original form date: 07th August 2020
SUMMARY

Full waveform inversion (FWI) is a data-fitting technique capable of generating high-resolution velocity models with a resolution down to half the seismic wavelength. FWI is applied typically to densely sampled seismic data. In this study, we applied FWI to 3D wide-angle seismic data acquired using sparsely spaced ocean bottom seismometers (OBSs) from the Deep Galicia Margin west of Iberia. Our dataset samples the S-reflector, a low-angle detachment present in this area. Here we highlight differences between 2D, 2.5D and 3D-FWI performances using a real sparsely spaced dataset. We performed 3D FWI in the time domain and compared the results with 2D and 2.5D FWI results from a profile through the 3D model. When overlaid on multichannel seismic images, the 3D FWI results constrain better the complex faulting within the pre- and syn-rift sediments and crystalline crust compared to the 2D result. Furthermore, we estimate variable serpentinisation of the upper mantle below the S-reflector along the profile using 3D FWI, reaching a maximum of 45%. Differences in the data residuals of the 2D, 2.5D and 3D inversions suggest that 2D inversion can be prone to overfitting when using a sparse dataset. To validate our results, we performed tests to recover the anomalies introduced by the inversions in the final models using synthetic datasets. Based on our comparison of the velocity models, we conclude that the use of 3D data can partially mitigate the problem of receiver sparsity in FWI.

Keywords: Waveform inversion, Controlled source seismology, crustal imaging

1. Introduction

Seismic imaging is one of the most widely used tools for understanding crustal processes. Advances in seismic imaging are mainly driven by the hydrocarbon exploration industry. In academia, first arrival travel-time tomography (FATT) using long-offset refraction data is a widely used method to obtain smooth velocity models of the crust (e.g. Hammer et al. 1994; Zelt & Barton 1998). A drawback of the FATT method is its resolution, which is limited at best
to the width of the first Fresnel zone. In addition, the presence of low-velocity zones can pose
difficulties for identifying the first arrivals in the data. To deal with such problems, seismic
imaging studies are moving forward towards higher-resolution techniques like waveform
inversion, which utilise more information in the waveform record to derive finer resolution
models. The full waveform inversion (FWI) technique was proposed by Lailly (1983) and
Tarantola (1984) as a linearized inverse problem using adjoint operators. Advances in
computational facilities and seismic data acquisition techniques have greatly supported the
development of FWI. FWI is capable of resolving structures to half the shortest seismic
wavelength present in the data. Moreover, it accounts for information beyond the traveltimes
of the waveform depending upon the wave attributes inverted during the inversion. However,
higher resolution comes at a cost of large computational requirement and increased non-
linearity of the inverse problem. Obtaining the correct solution using FWI largely depends on
how well-posed is the inverse problem.

Wide-angle seismic data are well suited for performing FWI because they are rich in low-
frequencies. Furthermore, OBSs enable recording of long-offset marine seismic data given a
powerful source, which otherwise is not possible using a limited-offset streamer. Although
wide-angle acquisition using OBSs has been widely conducted in last few decades, FWI
techniques have rarely been applied to these data until recently. In academia, the main concern
in using a wide-angle dataset for FWI is the data sparsity. A sparse dataset may cause the
inversion to fail to converge to the right solution due to a poorly constrained inverse problem.
The first application of FWI to synthetic wide-angle data was demonstrated by Pratt et al.
(1996) in 2D in the frequency domain, while real data applications were demonstrated by Dessa
et al. (2004) and Operto et al. (2006). In the last few years, different variants of FWI have been
applied to OBS data from different tectonic settings, restricted mostly to 2D (e.g., Górsczyk
et al. 2017, 2019; Kamei et al. 2013). Morgan et al. (2016) demonstrated the application of 3D
acoustic, anisotropic FWI to a sparse dataset from the Juan de Fuca Ridge, revealing low-
velocity zones interpreted as magmatic-hydrothermal reaction zones. Using a similar approach,
Davy et al. (2018) presented isotropic high-resolution models using sparsely spaced and noisy
2D OBS data from the Deep Galicia margin west of Iberia. The dataset used by Davy et al.
(2018) is a subset of the 3D dataset used in this study.
The Galicia margin has been a testing ground for various hypotheses on rifting, that explain
the observed crustal thinning, reduced mantle velocities, and mantle exhumation (e.g., Boillot
et al. 1987; Reston 2009). This margin is a sediment starved and magma-poor margin allowing
optimal resolution of features using geophysical imaging. Seismic studies in this region have
revealed many interesting observations over last few decades (e.g., Boillot et al. 1980;
Whitmarsh et al. 2001). 3D seismic reflection imaging has enabled unique insights into the
rifting tectonics in the margin, including 3D interpretation of the fault structures, identification
of the coherent corrugations on the S-surface and identification of the S-interval (Lymer et al.
2019; Schuba et al. 2018). 2D and 3D tomographic models derived using wide-angle seismic
data were used to identify thin oceanic crust west of the margin and estimate serpentinisation
below the S-reflector (Bayrakci et al. 2016; Davy et al. 2016)
In this paper, we performed 3D FWI using wide-angle seismic data recorded by 40 OBSs in
the Deep Galicia Margin (DGM) assuming an acoustic and isotropic medium. We also
performed 2.5D and 2D FWI using subsets of the dataset and compared the 3D FWI result with
2D and 2.5D FWI results. 3D wide-angle data have the potential to improve the resolution of
subsurface velocity structure compared to 2D because of the extra data recorded from the shot
lines away from the receivers, but 2D wide-angle studies are more common than 3D studies.
Hence, we aim to highlight differences in imaging between 2D, 2.5D and 3D approaches using
a real sparse dataset, and illustrate the limitations in resolution of 2D and 2.5D FWI imaging
in a rifted margin setting compared to 3D. A detailed interpretation of full 3D FWI results will be presented elsewhere.

### 2. Geological Setting

The West Iberia margin is a magma-poor rifted margin where rifting is thought to have occurred in several phases (e.g. Murillas et al. 1990; Péron-Pinvidic et al. 2007; Tucholke et al. 2007). In the early stages of rifting during the late Triassic to early Jurassic, the fault-bound Lusitanian and Galicia Interior Basins were formed (Figure 1; Murillas et al. 1990; Tucholke et al. 2007). During the Late Jurassic to Early Cretaceous, the Galicia Interior Basin developed further westward on the Galicia margin, thinning the continental crust to 15 km (Pérez-Gussinyé et al. 2003). In the final stages, the extension was primarily focused in the DGM, thinning the crust to less than 5 km (Murillas et al. 1990) and exhuming the subcontinental mantle lithosphere (Boillot & Winterer 1988; Tucholke et al. 2007). The DGM is a hyperextended zone characterised by west-dipping normal faults that sole into a detachment fault (Figure 1; Reston et al. 1996, 2007). Here, the continental crust comprises tilted fault blocks of crystalline basement rocks overlain by pre-, syn- and post-rift sediments. Various mechanisms are proposed to explain the faulting pattern, crustal thinning, asymmetry on the conjugate margins and role of the detachment fault in continental break-up. The detachment fault underlying the fault blocks is called the S-reflector, which is imaged as a bright reflector on seismic sections in this region (De Charpal et al. 1978; Lymer et al. 2019; Reston et al. 1996, 2007; Schuba et al. 2018). It is interpreted as a boundary separating the continental crust above from the serpentinised mantle below (Leythaeuser et al. 2005; Reston et al. 1996).

Westward of the thinned continental crust lies a narrow, north-south oriented ridge with 2-4 km of relief. Based on direct sampling, this ridge is interpreted as exhumed mantle and is known as the Peridotite Ridge (e.g. Beslier et al. 1993; Boillot et al. 1980). Similar ridges are identified along-strike on the West Iberia margin (Beslier et al. 1993). At the DGM, exhumed
mantle is postulated to extend seaward by at least 85 km from the Peridotite Ridge (Davy et al. 2016; Dean et al. 2015). Under the influence of continuous stretching during rifting, the entire crust was thinned to less than 10 km, inferred to have made it cooler and brittle and to have coupled the lower and upper crust (Pérez-Gussinyé et al. 2003; Pérez-Gussinyé & Reston 2001). The normal faults cut through the entire crust, acting as fluid conduits, transporting seawater to the upper mantle and serpentinising the upper mantle peridotites (Bayrakci et al. 2016; Pérez-Gussinyé & Reston 2001; Reston 2007). Using P-wave velocities derived from wide-angle seismic data, Bayrakci et al. (2016) and Davy et al. (2018) presented the evidence for preferential hydration of the upper mantle close to the normal faults, with maximum serpentinisation of 45%.

3. Seismic data

The Galicia-3D (G3D) experiment was conducted at the DGM from June to August 2013, to acquire a multichannel reflection and coincident wide-angle seismic data. Wide-angle data were recorded within the 3D box using 78 ocean bottom hydrophones/seismometers (Figure 2). The FS Poseidon deployed and recovered the OBSs, which concurrently recorded the signals from the seismic reflection experiment. OBSs were deployed in four lines with a spacing of ~6.5 km between the lines (Figure 2). The mean spacing between OBSs along each line is ~3.2 km. In this study, wide-angle data recorded by the hydrophones were used because of the absence of geophones on some instruments. The multichannel reflection seismic volume, acquired by the RV Marcus G. Langseth, was processed over an area of 68.5 km (east-west) by 20 km (north-south) (Figure 2). The 3D seismic reflection data were acquired using four 6 km streamers towed 200 m apart. The survey was carried out using two air-gun arrays each consisting of 20 airguns. The two arrays were towed 100 m apart with gun volumes between 40 cu.in. and 360 cu.in. at a depth of 9 m. The total individual volume of each airgun array was 3,300 cu.in.. Seismic shots were fired alternately using two source arrays every 37.5 m (shot
interval of ~16 s with ship speed of 4.5 knots). Processing of the reflection seismic data set was performed by Repsol, who produced a 3-D pre-stack Kirchhoff time migrated seismic volume.

4. Full-waveform inversion

FWI technique involves iteratively updating a starting model using a least-squares local inversion and the Born approximation (Lailly 1983; Tarantola 1984). The basic equation that is solved in the FWI is:

$$\tilde{m} = G^{-1} d_o$$  \hspace{1cm} (1)

where $d_o$ is the observed data; $G^{-1}$ is the inverse of the forward modelling operator ($G$); and $\tilde{m}$ is the inverted model. There is no formal method to estimate $G^{-1}$ from $G$. Hence, the solution to the inverse problem is achieved with an iterative approach using gradient based methods. In this study, we perform 3D FWI in the time domain using the FULLWAVE code (Warner et al. 2013). The starting model is updated in steps by reducing the residuals between the observed and predicted data generated with a forward modelling algorithm, which in our case uses the finite difference method (Warner et al. 2013). The root-mean square (RMS) amplitude of the predicted dataset is scaled to match the RMS of the observed data trace-by-trace, and a broad sliding time-window is applied during the modelling so that different phases are normalised independently (Warner et al. 2013). The misfit ($f$) between predicted and observed datasets is calculated by summing the squares of the differences between the trace-by-trace normalised predicted ($d_p$) and observed ($d_o$) data, for every time sample:

$$f = \frac{1}{2} \left( \frac{1}{\|d_p\|_2} - \frac{1}{\|d_o\|_2} \right)^2$$  \hspace{1cm} (2)
Assuming a linear relationship between model perturbation and data residual, the model update is given as

\[ \delta m = -H^{-1} \frac{\partial f}{\partial m} \]  

where \( H \) is the Hessian matrix containing all the second-order differentials of the functional \( f \) with respect to model parameters, and \( \frac{\partial f}{\partial m} \) is the gradient of the functional with respect to each of the model parameters. The gradient term is computed by taking a zero-lag cross-correlation between the source generated forward-propagated wavefield and residual generated back propagated wavefield using the adjoint source (Tarantola 1984). The computation of the Hessian matrix is avoided by replacing the Hessian with a scaling parameter using the steepest-descent algorithm for optimisation. Density is estimated from the velocities using Gardner’s law below the seafloor and is fixed during the inversion step (Gardner et al. 1974).

4.1 FWI Workflow

We followed the workflows of Warner et al. (2013) and Morgan et al. (2016):

1. Generating the source wavelet
2. Determining the starting model
3. Determining the starting frequency
4. Pre-processing the data
5. Defining the modelling and inversion strategy
6. Quality assurance

4.1.1 Generating the source wavelet

The source wavelet was generated by developing a Wiener filter. Building the wavelet involved the following steps: 1) extracting the nearest offset (<50 m) direct water-wave arrivals from a few selected OBSs, muting the arrivals after 1 second to minimise contamination by subsurface properties; 2) aligning the direct water-wave on every individual trace to the same time (say
3.5 seconds); 3) stacking the traces to obtain a mean trace for the direct-water wave at one OBS location and resampling the trace to 1 ms; 4) generating a synthetic trace using a random guess wavelet and water column velocities for the same selected OBS (the guess wavelet should be band-pass filtered in the same way as the data and resampled to 1 ms); 5) repeating steps 1 and 2 for the synthetic trace by extracting and adjusting the direct water arrival time; 6) matching the synthetic trace with the observed trace to obtain an inverse filter using the Wiener method; 7) convolving the inverse filter with the guess wavelet to build the source wavelet; and 8) verifying the source wavelet by re-generating synthetic data and comparing it with the observed OBS data.

4.1.2 Starting model

The starting velocity model is an important component of FWI. To qualify as a starting model for FWI, a velocity model should be able to predict the observed data within half a wavelength at the starting frequency. The starting model used for FWI in this study was derived from FATT of the wide-angle seismic data recorded by OBSs from Galicia3D experiment presented by Bayrakci et al. (2016). The authors used the First Arrival Seismic Tomography code, which solves the Eikonal equation using finite difference scheme and optimization using the LSQR variant of the conjugate gradient algorithm (Paige & Saunders 1982; Zelt & Barton 1998). The RMS misfit for the final FATT model is 73 ms and the chi-squared value is 1. We modify this model by replacing a constant water velocity of 1520 m/s with the water column velocity profile measured during data acquisition to improve the accuracy of the first arrival predictions in the forward modelling, and linearly interpolating the grid from 500 m spacing to 50 m, to ensure an accurate and smooth seafloor. Although the final inversion runs were performed using a 50 m grid spacing, most of the test runs initially were performed using a grid spacing of 100 m to reduce computational time. A 3D linear interpolation technique was used to interpolate the model from 500 m to 100 m and then further to 50 m after application of
smoothing filters described below. Cubic-spline, near-neighbour and linear interpolation techniques were all tested. The spline method introduced spurious values at the seafloor around the edges of the Peridotite Ridge, and the near-neighbour technique resulted in a pixelated model as it uses just the neighbouring points for calculating values at intermediate grid points. Use of a linear interpolation technique overcame these shortcomings and resulted in a smoother starting model compared to the near-neighbour technique. Smooth starting models without sharp velocity contrasts are preferred for FWI, unless such contrasts are well-constrained, such as the seafloor (Morgan et al. 2016). Therefore a further smoothing process was carried out using a 2D convolutional filter applied along each profile of the 3D velocity model. Prior to the application of the filter, the velocity model was converted into a slowness model and vertically resampled to 1 m. The filter was applied over a window with dimensions 200 m vertically and 1500 m horizontally to remove any sharp velocity changes.

The adequacy of the starting model was tested for its ability to predict the data within half a seismic wavelength for all instruments. The phase plots, plotting the phases of the first arrivals, assured the adequacy of the starting model by indicating that the starting model is not cycle-skipped (Figure 3). The phases are computed by applying a Fourier transform to each windowed trace to include just the first arrival and extracting the phase at a particular frequency (Warner et al. 2013). The concentric nature around the OBS of the phase changes of the first arrival with increasing offset indicates a good signal-to-noise ratio to maximum offset of ~15 km. At ~15 km offset, the phase plots for the observed data show more chaotic behaviour indicating poor signal to noise ratio (Figure 3a and b). If the starting model, compared to the observed data, is not cycle-skipped, then the phase change is smooth and consistent with offset whereas if it is cycle-skipped there will be sudden jumps in the phase (Shah et al. 2012). The phase residuals were calculated by subtracting phases of the first arrivals of the observed data from the synthetic data. The phase of the first arrivals of the predicted and observed data were
compared to check if the predicted data were within an acceptable limit (-π to +π) (Figure 3e and f). Further verification was carried out by plotting sets of 10 traces from each dataset – observed and predicted – alternating to manually check for poor fit (Figure 4).

### 4.1.3 Starting frequency

Non-linearity of the inverse problem can be partially mitigated by starting the inversion from the lowest frequency available in the data and progressing towards higher frequencies (Bunks et al., 1995). The choice of the starting frequency for our inversion is based on examining the phase of the first arrivals in the observed data with increasing offset (Figure 3a and b). The RMS values of the phase residuals at 2.7 and 3 Hz are 1.42 and 1.44 radians, respectively. RMS values well below π radians clearly indicate that the starting model is able to predict the phase of the first arrivals within the acceptable limits, and the starting frequency for the inversion can be 2.7 or 3 Hz. Test inversion results performed using a subset of the dataset starting from 2.7 Hz and 3 Hz did not show significant differences. The RMS amplitude of the difference between the results of these test runs is equal to 36 m/s with the largest perturbation value equal to 236 m/s (Figure 5). Hence, we started the final inversion run by applying low-pass filter rolling off at 3 Hz to reduce the computational time, rather than following the fundamental notion of starting the inversion from the lowest possible frequency (Sirgue, 2006).

### 4.1.4 Pre-process the data

A minimum-phase Ormsby bandpass filter with corner frequencies of 2-3 to 5-7 Hz was applied to reduce the random noise in the input data. The limits for the band-pass filter were chosen after trying different frequency ranges. The same filter was also applied to the source wavelet. Manually picked mute gates were applied to each OBS to include only the first arrivals. The unmuted window lengths were customized depending on the match of the observed traces with the synthetic traces generated using the starting model (Figure 4). The window lengths varied between 700 ms and 1200 ms for all the instruments. Data up to maximum offset of 15 km
were included into the inversion, beyond which the data quality was degraded by arrivals from the previous shot, making it unsuitable for FWI.

### 4.1.5 Modelling strategy

We inverted only for P-wave velocities, assuming isotropic and acoustic approximations for the medium. These assumptions are based on the previous tomographic and FWI studies in this region using the same wide-angle dataset, which indicated an absence of strong elastic and anisotropic effects (Bayrakci et al. 2016; Davy et al. 2018). In this work, we developed a 2D velocity model using only two nearby shot lines from the instruments along line 2, a 2.5D velocity model using all shots from the instruments along line 2, and a 3D velocity model using wide-angle data from OBSs along lines 1, 2 and 3 (Figure 2).

Wide-angle refraction data from 40 OBSs were considered for the 3D inversion (Figure 2). The number of profiles recorded by each instrument was 100. Data from the northernmost line (line 4) of OBSs were not included because of larger uncertainty in their re-located positions, which is partly due to the lack of a shooting profile directly above the OBSs. Profiles with low signal-to-noise ratio on a particular OBS were excluded from the inversion. Only data that were predicted by the starting model within a half seismic wavelength were included in the 2.5D and 3D inversions (Appendix A). Data from OBS 51 were not included in the 2D FWI because the data along the closest shot lines were not recorded by this instrument, but were included in the 2.5D and 3D FWI. Data from OBSs 80 and 85 were not used in our inversions because the starting model was unable to predict the data accurately due to time shifts that we were unable to correct.

The maximum and minimum velocities in the starting model are 1497 and 8000 m/s, respectively. The grid spacing in the velocity model is 50 m in each of the three dimensions. To ensure that the wavefield is computed in at least two locations in every cell given the maximum velocity in the starting model, the sampling interval of the input data should be less
than 3.125 ms (0.5 * grid spacing/maximum velocity; Lines et al., 1999). Hence, the data were resampled to 3 ms to satisfy this condition. The grid dimensions in the inline and crossline directions were chosen to include all the receivers and shots within the grid and to avoid edge effects in the wave propagation, which requires shots falling within 1 km of the edge of the grid to be removed. The dimensions of the grid along the inline (along the shooting profiles) and crossline (across the shooting profiles) directions are 78.5 km and 22.1 km, respectively, and the vertical dimension is 12 km. To facilitate the implementation of the code, a data reciprocity rule was applied to treat the shots as receivers and OBSs as shots. The inversion was performed following the multi-scale approach starting from low frequencies and progressing towards higher frequencies in steps (Bunks et al. 1995). During the inversion, the program applies low-pass filters, in steps rolling off at frequencies 3, 3.3, 3.9, 4.5 and 5.2 Hz, to the data and the source wavelet (Warner et al. 2013). Thirty iterations were performed at each filter setting to allow the misfit functional to reduce and flatten. For the final inversion run, smoothing was applied to the gradient using a Gaussian filter with correlation lengths of 2 x seismic wavelength and 1 x seismic wavelength in the horizontal and vertical directions, respectively. The seismic wavelength for gradient smoothing was calculated using the local velocity in the model and the rolling off frequency specified for the current iteration. These smoothing parameters were adjusted to suppress short-wavelength artefacts in the result. 2.5D and 2D inversions were performed using same parameterisation as 3D inversion but included data only from instruments along line 2. The total number of shots for the 3D, 2.5D and 2D inversions are 84768, 55591 and 1792, respectively. For the 2D inversion, two shot lines were used from 16 OBSs along line 2 (Figure 2) and projected onto an imaging line closest to line 2. The velocity model dimensions for the 2D inversion were 78.5 km in the horizontal direction and 12 km in the vertical dimension.
FWI is a computationally demanding imaging technique, so forward modelling and inversion runs were performed on two large Linux clusters: Iridis 5 (University of Southampton) and Mobilis (National Oceanography Centre). Mobilis comprises 72 computing nodes each with 64 GB of memory (DDR3 memory) and 2 x Intel Xeon E5-2650 2.6 GHz eight-core processors, giving a total of 1152 processor cores. Iridis 5 comprises of 464 compute nodes with dual 2.0 GHz Intel Skylake processors. Each compute node has 40 CPUs per node with 192 GB of DDR4 memory. Initial testing and 2D inversions were performed on Mobilis because of its smaller queue time, whereas the final 3D inversion runs were performed on Iridis 5. The 2D inversion run was performed by assigning 4 OBSs to each node, while for 3D inversion 2 OBSs were assigned to each node. In 3D inversions, no extra CPUs/nodes were assigned for the OBSs that have just two shot lines recorded (Figure 2). Runs on Mobilis were performed using 4 nodes and the computation time for each iteration was ~1 minute in the final 2D inversion run. For the final 3D inversion run, 18 nodes were employed on Iridis 5, with one extra node reserved for communication between the nodes. The compute time for each iteration for the 3D run was ~45 minutes.

4.1.6 Quality assurance

We tracked the progress of the 3D inversion by monitoring the phase residuals, obtained by subtracting the phases of the predicted and observed data (Figure 6). These phase residuals clearly have reduced after the first iteration at both frequencies (Figure 6c and d). Abrupt changes from one extreme phase to other (\(-\pi\) to \(+\pi\)) indicate cycle-skipping, which is not observed except for a few noisy traces at the far and near offsets. Including a substantial number of cycle-skipped traces into the inversion could cause it to converge to a local minimum that is not the global minimum. However, if the number of such traces is small, the effects from cycle-skipped traces are nullified by the rest of the data driving the model towards the global minimum (Shah et al. 2012). For most of the data the phase change is gradual at 3 Hz, except
for a very few traces at far offsets that have poorer signal-to-noise ratio (Figure 6). At 4.5 Hz, there are regions that change their phases abruptly from red to blue. However, these same regions show small phase changes at 3Hz. There is a gradual improvement as the inversion progresses without cycle-skipping, and the phase difference for most of the traces is close to zero for the final model (Figure 6e and f). This is a clear advantage of progressing from lower frequencies to higher frequencies during the inversion. At lower frequencies, the probability of cycle-skipping is lower because the misfit function near the global minimum is smoother and broader, hence reducing the chance of getting trapped in a local minima. To further check the convergence of the FWI, we plotted the predicted data from the travel-time model and FWI inverted model with the observed data to check the improvements resulted from FWI (Figure 7; Appendix B). The result shows clear improvement in the alignment of the FWI model synthetics with the observed data compared to the travel-time model synthetics.

5. Results and discussion

We compare the results from 2D, 2.5D and 3D FWI along the seismic line 420 (OBS line 2) in both depth and time domains (Figure 2, 8 and 9). In our comparison, we have also included the 2D FWI result from Davy et al. (2018), which was derived along the same line using the wide-angle data from the OBSs, but from a different starting model (from Davy et al. 2016). Our 2D result is derived using just two shot profiles closest to the instruments along line 2 and the starting velocity model was obtained from the 3D travel-time model by extracting a profile closest to the shots and instruments along line 2 (Bayrakci et al. 2016). In order to highlight the differences between 2D, 2.5D and 3D imaging, we overlaid the time-converted velocity profiles on a time-migrated seismic section obtained from 3D seismic volume (Figure 9; Lymer et al. 2019). The interpretations used in our comparison were developed through the entire 3D reflection seismic volume and are independent of our FWI velocity models (Lymer et al. 2019). In the text that follows, we use the following abbreviations: 2DA corresponds to 2D FWI result
of Davy et al. (2018), 2DB corresponds to our 2D FWI result with shots and receivers projected onto an imaging line, 2DC corresponds to the 2D FWI result with shots and receivers at their actual locations, 3DA corresponds to the 2.5D inversion using data from instruments along line 2 and 3DB correspond to full 3D results.

5.1 2D FWI

5.2.1 2DA vs 2DB:

The inversion strategy of the 2DA FWI result differs from that of the 2DB FWI result in the length of the average unmuted time window applied to the input seismic data, the maximum offset of data used for the inversion and the smoothing parameters applied to the models during the inversions. These differences are likely to be the primary reason for the differences observed in the two sections, especially the short wavelength structures (Figures 8a-b and 9a-b). In addition, the starting models used for 2DA and 2DB are derived from different studies (Figures 8a-b, 9a-b; Bayrakci et al. 2016; Davy et al. 2016). In 2DA, the starting model was derived from 2D travel-time tomography, in which the thickness of the sediment column is also constrained using the top basement depths from the multichannel images. Such a modelling approach resulted in improved fit of the top of the basement with the multichannel images, which is evident at the Peridotite Ridge compared to the result from 2DB (Figure 8). In both the results, the long wavelength structure fits the interpreted top of the pre- and syn-rift sediments and crystalline crust well, indicating that different starting models generated a similar background model. However, the finer details within the fault blocks and the undulations at the S-reflector differ between the images.

5.2.2 2DB vs 2DC:

Traditionally in 2D imaging the shots and receivers locations are projected onto an imaging line, which results in errors in wave propagation. However, if the projections are within an acceptable distance from their actual locations to the closest imaging line passing through the
shots and receivers, the resulting errors would be insignificant. Performing synthetic tests with the projected locations can validate these approximations, in which anomalies introduced artificially into the velocity model are recovered using synthetic data. Such a synthetic test was performed for the 2DB case, and the introduced anomalies were recovered during the inversion using the same starting model used for 2DB inversion (Figure 14a-b; Bayrakci et al. 2016). The result from this test is discussed in detail in the section 6. Although the synthetic test validated the approximations considered in the 2DB case, we performed another 2D inversion (2DC) using the actual shot and receiver locations to highlight the differences in 2D imaging due to the projected locations. In order to consider their actual locations, we performed the 2DC inversion using a narrow 3D velocity grid of width 2 km in the crossline direction encompassing the shots and receivers along the line 2 (Figure 2). The other dimensions of the velocity grid were the same as the 2DB case and the starting model was extracted from the 3D travel-time tomography (Bayrakci et al. 2016). Also, the number of shots and receivers used in 2DC were the same as the 2DB case. The models 2DB and 2DC compare well up to a depth of 8 km and show large differences in the deeper regions of the model (Figure 10). At the S-reflector, 2DB shows velocities greater than 6.5 km/s, whereas the velocities in 2DC are less than 6.5 km/s (Figure 10). In principle, these two inversion runs should generate similar images, considering that the errors from shots and receiver locations are insignificant. However, predominantly large differences below 8 km depth suggest that the 2D imaging has resulted in a non-unique model of the deeper sections (Figure 10). It is arguable that the 2DC inversion results in more accurate 3D amplitudes than the 2DB inversion, therefore giving rise to the difference in the images. Seismic data generated using 3D and 2D codes produce synthetic data with a factor of $\sqrt{t}$ difference in amplitudes (where $t$ is travel-time). Since the 2DC inversion uses a narrow 3D grid the amplitudes of the synthetic arrivals will better match the observed, and this may lead to differences in the 2DC and 2DB images. However, in the
FWI code used here, the input data are normalised using a broad sliding window, and this has a net effect of minimising the effect of amplitude differences due to a 2D approximation, attenuation, elasticity etc (Warner et al. 2013).

Although 2D FWI results show a reasonable match with the multichannel seismic image (Figure 8a-b, 9a-b), differences in the short-wavelength structures are noticeable. Compared to the 3D starting model, all the 2D results show an improved fit to the multichannel images especially along the top basement blocks (Figure 9). Our preferred 2D FWI model for this study is 2DB because it has been developed following the conventional 2D procedure by projecting shots and receivers unto the imaging line. Also, the parameters used in developing 2DB are the same as the 3D inversions making it more suitable for comparison.

5.2 3D FWI

Both 3D velocity models have resolved the structures that are geologically relevant, but there are differences between them (Figure 8). To highlight the differences between 3D models, we carried out a detailed comparison of the images based on the match between the velocity models and multichannel seismic images, focusing on the pre- and syn-rift sediments, the basement blocks and the S reflector. We also include the 2D FWI models in the following discussion.

5.2.1 Pre- and syn-rift sediments:

Pre- and syn-rift sediments are interpreted by Lymer et al. (2019) between the boundaries of the base of the post-rift sediments (dashed blue line) and top of the crystalline crust (dashed red line) (Figures 8 and 9). All of the 2D and 3D FWI results resolve the long wavelength structure in this zone well, as indicated by their match with the multichannel interpretations. The velocities of pre- and syn-rift sedimentary packages range from ~3.5 km/s to ~5.25 km/s overlying the crystalline crust along this profile. The 3.5 km/s velocity contour tends to oscillate in the FWI results compared to the starting model, and the pattern of the oscillations
is inconsistent between different 2D and 3D results beyond 25 km (Figure 8 and 9). This oscillatory pattern is likely to be the result of complex faulting within the pre- and syn-rift packages which has been previously interpreted in multichannel seismic images from the Deep Galicia margin (Lymer et al. 2019; Ranero & Pérez-Gussinyé 2010; Reston 2005). Such a fine-scale faulting pattern is beyond the resolution limits of our inversions, primarily because of the absence of high-frequency content in our dataset. However, the oscillations are smaller in the 3DB result than in the 3DA and all 2D results. Although the contour shows short-wavelength perturbations (oscillations) within the pre- and syn-rift sediments, it shows an improved alignment with the faults compared to the starting model, especially along faults F5, F4 and F3 (Figure 9).

Our 3DA model shows artefacts within the sedimentary column, which are vertical in nature and lie beneath the OBS positions (Figure 8 and 9). These artefacts are mainly observed in the western region of the model between the Peridotite Ridge and the crustal blocks. They arise from the inclusion of shot profiles away from the OBS locations. Arrivals from the far-offset profiles travel through deeper parts of the model that are not well constrained in the 3DA inversion, because in 3DA there are no other OBSs located between the far-offset shot profiles and the OBS location to constrain deeper parts of the model. Further, the fact that these artefacts are not observed in the 2DB and 3DB models supports the above explanation. Another contributing factor for these artefacts may be the high uncertainty of the starting model in the western part (Bayrakci et al. 2016).

5.2.2 Crystalline crust:

For all the FWI models, the crystalline crustal section shows a significantly better match than the starting model to the multichannel seismic image (Figure 9). The velocity range for crystalline basement inferred from our results is ~5.25 km/s to ~6.5 km/s. All of the 2D models show circular velocity contours in some locations within the crust, which are geologically
unrealistic. Such structures can be a result of a poorly constrained inverse problem or from the assumed approximation of the wave equation as acoustic, isotropic and non-attenuating. However, these circular features would also be observed in 3D imaging if the assumed approximations of the wave equation do not apply to the real Earth model in the study region. A good match of the 3DB velocities within the crystalline crust with multichannel images supports the 3D nature of crustal fault-blocks.

5.2.3  S-reflector and serpentinisation:

Previously, the 3D travel-time model has shown that velocity 6.5 km/s contour follows well the S-reflector (Bayrakci et al. 2016). However, FWI results have introduced undulations in the 6.5 km/s contour in both 2D and 3D. 2DA and 2DB show different velocity patterns at S that are highlighted by the 6.5 km/s contour. The reason for this difference may be the different inversion strategies for the two cases, particularly the choice of the window lengths, smoothing parameters and offsets for the input data. Although the 2DA and 2DB models predict higher velocities than 6.5 km/s at S compared to the starting model, it is observed that 2D imaging has generated non-unique models at S in the section 5.1. To explore the velocities below the S-reflector, we plotted average velocity values below the S-reflector over a window of 100 ms along line 2 for all the models using the interpretations from the multichannel seismic image (Figure 11). It can be seen clearly that the 2D results show poor correlation with each other except at ~43 km where both models show unrealistically high velocities (Figure 11). However, the pattern of 2DA shows a match with the 3DB result in certain locations and the 2DB result shows a poor match with 3DB at all locations along the profile. None of the 2D results correlates well with the 3DA and 3DB results (Figure 11). The correlation between 3DA and 3DB is poor until ~27 km, beyond which the curves tend to follow each other. This mismatch between 3DA and 3DB in the western parts of the models may be due to inclusion of very limited data from OBSs 38 and 40. Moreover, the data from OBSs 38 and 40 were
limited to ~11 km offset because our starting model was not able to predict the data sufficiently well beyond these offsets for these OBSs. Only 10 shot profiles closest to their locations were included from these OBSs in 3DA and 3DB inversion runs.

Overlaying the 3DB velocity model over a time slice through the 3D seismic volume at 100 ms below an average depth of the S-reflector shows a good match between the velocity model and the time slice, suggesting that the model is well-resolved at these depths. The good match also suggests that, at the resolution of our models, the S-reflector is a velocity boundary (Figure 12). These observations suggest that 2D imaging has failed to generate a realistic model in the deeper parts. Furthermore, the range of velocities resolved by 3DB below the S-reflector indicates that 3DB has constrained the problem well with 8 km/s corresponding to unaltered peridotites with 0% serpentinisation (Carlson & Miller 2003). Velocities higher than 8.3 km/s are unrealistic at the S-reflector and lower velocities correspond to some degree of serpentinisation of the mantle (Bayrakci et al. 2016; Davy et al. 2018). In 3DB, the region below the S-reflector has been well resolved because of the data from the instruments located in the other OBS lines of the survey which were not included in the 2D and 2.5D inversion runs. Our 3D modelling has clearly enhanced resolution of the pattern of serpentinisation below the S-reflector compared to the starting model which predicted velocities mostly close to 6.5 km/s. Below the S-reflector the mantle peridotites have undergone serpentinisation, with an inferred maximum serpentinisation of 45% occurring around the fault intersections based on the relationship of Carlson & Miller (2003).

Based on 2DA, Davy et al. (2018) supported the concept of preferential serpentinisation of the upper mantle rocks close to faults. The result from 3DB along the same profile shows a better correlation between the fault intersections with the S-reflector and occurrence of lower velocities compared to the other models along this line.
5.3 Data fit

Another interesting aspect to compare between 2D and 3D FWI is the fit between the observed and predicted datasets. Our forward modelling scheme does not account for medium properties like anelastic attenuation and anisotropy, nor elastic and density effects that affect wave propagation in the real earth. Therefore, our modelling does not completely predict the field data, especially the amplitudes of the waveforms. Although technically it is possible to invert for these effects during an inversion, such an approach can make the inversion an ill-posed problem with many unknowns, particularly in case of a sparse OBS dataset with limited data coverage.

Wave propagation in 3D is more accurate than in 2D. Therefore, it can be expected that 3D waveforms fit the real data better than 2D waveforms. However, our data fitting results suggest that this expectation is not always met. We plotted data residuals for 2DB, 3DA and 3DB (2D, 2.5D and 3D, respectively) using data only used for the 2DB inversion (Appendix C). We applied trace-by-trace normalisation to the real and synthetic data before calculating the residuals. We calculated the data residuals for each model by subtracting the final model predicted data from the observed data and the RMS value of the difference was computed to quantify the fit in all three cases. For majority of the OBSs the fit between the predicted and observed data for the 2D and 3D cases is similar, and there are a few cases where 2D fits the data better than 3D (Figure 13). OBSs 38, 40, 81 and 82 show significantly better fits in 2D than in 3D. To explain these observations, we explore the cases of instruments 45, 46 and 38 in detail.

For OBS 45, the 2D inversion provides the best fit, while for OBS 46, the 2.5D inversion provides the best fit (Figure 14). The differences are small, making it difficult to establish clearly which inversion is better fitting the data. We performed 30 iterations at each frequency in all the inversions and increasing the number of iterations may have further narrowed the
differences without significant improvements in the velocity models. However, for OBS 38, the differences between 2D, 2.5D and 3D cases are significant compared to other OBSs (Figure 14). OBS 38 has high uncertainty in its location due to the need to correct large time shifts in the raw data, so only very limited data from this OBS, that our starting model was able to predict, were used in the inversion. The significantly smaller RMS value of 2D residuals compared to 2.5D and 3D residuals may be due to overfitting. It is possible that there is overfitting in all the three cases for all the OBSs. However, the degree of overfitting varies from OBS to OBS depending on various factors that affect the constraints on the inversion such as the dimensions of the problem, data sparsity, OBS location uncertainties and unaccounted wave propagation effects. We suspect higher degrees of overfitting of data in the 2D case in OBSs 38, 40, 81 and 82 which show larger differences in the RMS values of the residuals (Figure 13). It is also possible that the circular features observed in the 2D imaging partially resulted from overfitting. The difference between 2.5D and 3D RMS values is due to the inclusion of the data from the OBSs along lines 1 and 3 resulting in lesser constraints on the 2.5D inversion compared to the 3D inversion (Figure 2).

In our comparison of data residuals, it is important to note is that we have suppressed the amplitude effects during the inversion by performing trace normalisation of the datasets. Inverting for amplitude effects would have probably resulted in larger data residuals, fundamentally due to the different 2D and 3D forward modelling operators. Even with such subtle differences in the data residuals, an interesting observation is the occurrence of brighter residuals at the near offsets in the 3D data residuals (Figure 14). This is probably due to poor sampling of the subsurface by OBS data in the near offsets in general, perhaps sparse spacing between the OBS may have further reduce the constraints at the near offsets. However, the same near-offset arrivals in the 2D data residuals are matched due to overfitting by the inversion (Figure 14).
6. Recovering the anomalies

The most common way to estimate the resolution of the final inverted model is through checkerboard tests (e.g. Morgan et al. 2016). In our case, we did not perform checkerboard tests because introducing even small anomalies in the form of checkers resulted in changing of the arrival times, particularly far-offset arrivals, beyond the range of our mute gates applied to the input data for the actual inversion. Instead, we added the anomalies introduced by the FWI into the starting model and recovered the anomalies by performing synthetic tests. To check whether our inversion strategies produced consistent results and to explore the size of the anomalies introduced by the FWI, and hence the resolution of the inversions, we performed synthetic inversion runs using the 2DB, 3DA and 3DB models. In these tests, we generated synthetic data using the final models and used the synthetic data as observed data in the inversion runs. The synthetic data were generated using the same forward modelling scheme as used in the actual inversion runs. The starting model was the same as used in the actual runs (Bayrakci et al. 2016). The results from these inversions, when subtracted from the starting should match the anomalies recovered in the actual inversion runs 2DB, 3DA and 3DB (Figure 14). All three tests successfully recovered most of the anomalies introduced by the FWI into the final models, indicating that the anomalies recovered in the actual inversion runs are real. However, the amplitudes of the recovered anomalies are smaller in magnitude than the actual anomalies. This discrepancy may be because of employing trace-by-trace normalised misfit function in the inversion which in effect only inverts for phase information in the data. The RMS amplitudes of the differences between the anomalies introduced by FWI and the anomalies recovered are 45 m/s, 68 m/s and 83 m/s for the 2DB, 3DA and 3DB inversions, respectively.

Analysing the anomalies introduced by 2D, 2.5D and 3D inversion, we observe that the shapes of anomalies introduced by 2DB differ from 3DA and 3DB. This difference is due many factors
affecting the constraints on the inverse problem with major contribution coming from the difference in the amount of data. An interesting change in the anomalies from 2D to 3D is the extent of the blue anomaly below the S-reflector (Figure 15). In 2DB, the blue anomaly below S is mainly observed between ~35 and 50 km, while in 3DA and 3DB the anomaly extends to ~27 km and ~20 km, respectively, covering the full length of the S in 3DB (Figure 15). This progressive extension of the anomaly demonstrates that 2DB failed to resolve the velocities below the S-reflector along its full length. The reason for the poor performance of 2DB below the S-reflector may be that fewer first arrivals from these regions are included in the inversion. This problem may be a more general one for performing FWI on sparsely recorded datasets. Previously, most 2D FWI studies using OBS datasets have had advantage of close spacing of the OBS and long offset recording, enabling strong constraints on the deeper part (e.g. Dessa et al. 2004; Operto et al. 2006). Jian et al. (2017) demonstrated the application of 2D elastic FWI to a sparsely recorded OBS dataset to image an axial magma chamber-like structure below the Southwest Indian Ridge. In that study, data from only three OBSs were used and the spacing between the OBSs were 8 km and 16 km, which is greater than in our study. However, in spite of such sparse spacing, FWI successfully converged using data from offsets up to ~50 km. Combining the experience from this and previous studies of FWI implementation using OBS data, it can be inferred that OBS spacing is an important factor for successful application of FWI, but other factors are also important, such as shot spacing, the maximum offset with usable data and the target depth of investigation. In our case, the 3D dataset has partly mitigated the problem of data sparsity, which is evident from the differences between the 2D and 3D FWI results.

7. Conclusions

We have successfully performed 3D FWI in the time domain using wide-angle data recorded using sparsely spaced OBSs on the Deep Galicia margin. We compared our results with a 2D
FWI result derived using a subset of the dataset through the 3D model. Our comparison revealed:

1. Both 2D and 3D FWI applications show improved alignment of the structures with coincident multichannel seismic reflection images, compared to the travel-time model.

2. Within the pre- and syn-rift sediments, the 2D model shows an oscillatory behaviour, which is reduced in the 3D model indicating 3D nature of the fine-scale faulting.

3. In the crystalline crust, 2D imaging fails to fully recover the alignment of structures due to the complex 3D nature of the faults observed on the multichannel reflection images.

4. Poor constraints on the inverse problem due to sparse data and/or our assumed approximation of the wave equation results in circular velocity contours which are geologically unrealistic. 2D FWI images contained more such features than 3D images, indicating that they arise mainly from the data sparseness.

5. The 3D FWI model has enhanced the resolution of the pattern of serpentinisation below the S-reflector compared to the starting model. The occurrence of lower velocities below the S-reflector correlates with overlying fault intersections at the S-reflector.

6. Data residual plots suggest that 2D inversion can be prone to overfitting in the case of a sparse OBS dataset, while 2.5D and 3D inversions are better constrained with more data to avoid overfitting. Overfitting may vary from OBS to OBS depending on factors like location uncertainty and poor coverage by the data.

7. Anomaly recovery tests recovered the anomalies introduced by the FWI in both 2D and 3D. The progressive extension of an anomaly below the S from 2D to 3D results highlights that the 2D inverse problem is poorly constrained in the deeper parts of the model.
8. Acknowledgements

Data acquisition was supported by the U.S. National Science Foundation, the UK Natural Environment Research Council (NERC; grant NE/E016502/1), and GEOMAR. Ocean bottom instrumentation was provided by the NERC UK Ocean Bottom Instrumentation Facility and by GEOMAR. T.A.M. was supported by a Wolfson Research Merit award. The 3D ProMAX/SeisSpace package, supplied by Halliburton under a university software grant, was used to preprocess and analyse the field data within University of Southampton. We thank Jeff Blundell at the University of Southampton for assisting in setting up the code on supercomputing facilities. We thank the members of the Fullwave research consortium at Imperial College London who funded the development of the FWI software used in this study. We thank Richard Davy for providing his 2D FWI result and Gael Lymer for providing his interpretations of the 3D reflection volume. We thank Milena Marjanović, Andrzej Górszczyk and other anonymous reviewers for providing constructive reviews and comments that greatly improved the manuscript. We thank the editor Michal Malinowski for suggesting improvements to the manuscript.

Data Availability:

The data underlying this article are being uploaded to https://www.pangaea.de. Prior to availability there, they will be shared on a reasonable request to the corresponding author.

9. References


the ultraslowspreading Southwest Indian Ridge. *Geology, 45*(2), 143–146.

https://doi.org/10.1130/G38356.1


https://doi.org/10.1029/2005GL024026


https://doi.org/10.1190/1.1444605


Ranero, C. R., & Pérez-Gussinyé, M. (2010). Sequential faulting explains the asymmetry and

https://doi.org/10.1038/nature09520


https://doi.org/10.1016/j.epsl.2005.06.019


https://doi.org/10.1029/95JB03466


Figure 1 | Upper panel: map showing the location of the Galicia 3D (G3D) experiment in the DGM (black box). East of the DGM is the Galicia Bank, Galicia Interior Basin (GIB), Porto Basin (PB) and Lusitania Basin (LB). To the south of the study region is the Southern Iberia Abyssal Plain. Colours in the map indicate bathymetry and topography. Lower panel: time migrated seismic profile across G3D area highlighting the Peridotite Ridge (PR), pre- and syn-rift sediments (below dashed blue), crystalline crust (below dashed red) and the S-reflector (dashed black).
Figure 2 | Map showing the location of the OBSs, and multichannel seismic shooting in white. The OBSs locations are indicated along four profiles from south to north: line1, line2 line3 and line4. Locations of multichannel seismic line 420 through the instruments in line2 is shown in orange. The red-dashed box marks the area for the 3D FWI and the dashed brown box marks the multichannel seismic volume. Green coloured instruments are used for 3D FWI, whereas the yellow ones have not been used in this work. The instruments in pink have only two shot profiles recorded directly above them.
Figure 3 | Phase residuals for OBS 46 for the starting model at frequencies 2.7 Hz and 3 Hz. Each dot represents the phase of the first arrivals extracted using a Gaussian window centred on these arrivals. a) & b) are the phase plots of the observed data at 2.7 and 3 Hz, respectively. c) & d) are the phase plots of the synthetic data predicted using the starting model at 2.7 and 3 Hz, respectively. The residual plot (e & f) indicate that most of the data is predicted within half a cycle accuracy by the starting model. Noisy traces at far offsets indicated by red arrows are out of phase.
Figure 4 | Observed (highlighted) and predicted data using the starting model plotted in sets of 10 traces from OBS 46, alternating to show the match between the arrivals along two profiles. a) Profile 1 is the closest profile to the OBS 46 location and b) Profile 2 is at a distance of ~10 km from this location. The data shown are band-pass filtered and trace-by-trace normalised.
Figure 5 | Difference between the starting model and final model along a profile through the instruments in line 2 using the 3D inversion. a) is the difference plot for inversion starting from 2.7 Hz and b) is the difference plot for inversion starting from 3Hz. c) is the difference between a) and b).
Figure 6 | Phase plots residuals for starting model, model after 1$^{st}$ iteration and final models at 3 and 4.5 Hz. The phases show gradual change towards zero phase difference except at two locations highlighted in dashed red boxes at 4.5 Hz. A few lines have been excluded from the inversion because of their poor fit or poor quality, resulting in gaps in the refracted arrivals in the above plots.
Figure 7 | Observed and synthetic data along the same two profiles from OBS 46 as shown in Figure 4. a) and b) Left-hand profiles are alternating sets of 10 traces from observed data (highlighted) and synthetics from the starting model. c) and d) Right-hand plots show the observed data (highlighted) and synthetics from the final model.
Figure 8 | Velocity profiles along line 2: Profile a is the result from 2D FWI – 2DA (Davy et al. 2018), profile b is the 2D FWI from the 3D starting model – 2DB, profile c is the 3D FWI using only the instruments along line 2 – 3DA, profile d is the result from the full 3D FWI using instruments coloured in green and pink in figure 2, and profile e is the starting model from 3D travel-time tomography.
(Bayrakci et al. 2016). Comparing the 2D FWI and 3D FWI in depth domain. Interpretations from multi-
channel seismic images (Lymer et al. 2019) are converted from time to depth using respective velocity 
models and overlaid on the sections along with the OBS locations in yellow circles. Dashed blue line 
marks top of the syn-rift sediments, dashed red line is the top of the crystalline crust and dashed black 
line is the S-reflector. Velocity contours for 3.5, 5.25, 6.5 and 7.5 km/s are marked in solid black lines. 
The red arrow indicates imaging artefacts. The black solid box in 2DA around peridotite ridge highlights 
a good match between the multichannel interpretation (dashed blue) and the velocity contour for 3.5 
km/s compared to the other models.
Figure 9 | Comparison between 2D and 3D FWI results and multichannel time sections, which are overlaid with interpretations from Lymer et al. (2019). Profiles and velocity contours correspond to their respective profile and contour labels in Figure 8. Fault numbers (F-series) are marked on the FWI result (d). Red arrow marks vertical artefacts.
Figure 10 | 2D FWI inversion results a) 2DB obtained by projecting shots and receivers onto the imaging line b) 2DC performed using a narrow 3D grid encompassing the shots and receivers at their actual locations c) difference between 2DB and 2DC indicates that the differences are larger below 8 km depth.
Figure 11 | Velocities below the S-reflector averaged over a window of 100 ms for each of the FWI results. The intersection of the faults with the S-reflector are indicated by vertical dashed lines.
Figure 12 | Time slice extracted at 9.1 s two-way traveltme a) time slice from the multichannel volume b) 3DB FWI model overlaid on the time slice. Black arrows point to the boundary between the crust and upper mantle. North is shown by orange arrow. Line 2 is shown in orange.
Figure 13 | RMS percent of the residuals for 2D, 2.5D and 3D (2DB, 3DA and 3DB) for each OBS along line 2 within oval shaped circles in black with OBS numbers on their top.
Figure 14 | Data residual plots of the normalised observed and synthetic datasets for the 2DB, 3DA and 3DB models for OBS 45 (a), 46 (b) and 38 (c) along the closest shooting profile to the instruments. Each panel shows the OBS number, number of dimensions of the inversion, and RMS value of the residuals. The white arrows indicate the nearer offsets arrivals.
Figure 15 | a) Difference plot of the 2DB model and the starting model. b) Difference between the recovered anomaly model and the starting model. c) Difference plot of the 3DA model and the starting
model. d) Difference between the recovered 3DA anomaly model and the starting model. e) Difference
plot of the 3DB model and the starting model. f) Difference between the recovered 3DB anomaly model
and the starting model. The black arrows indicate a blue anomaly that progressively extends below the
S from 2D to 3D. The interpretations are same as in figure 8 for respective profiles.
### Appendix A:

<table>
<thead>
<tr>
<th>OBS number</th>
<th>Number of shots</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>3917</td>
</tr>
<tr>
<td>57</td>
<td>6445</td>
</tr>
<tr>
<td>59</td>
<td>6314</td>
</tr>
<tr>
<td>61</td>
<td>8193</td>
</tr>
<tr>
<td>62</td>
<td>8899</td>
</tr>
<tr>
<td>63</td>
<td>3961</td>
</tr>
<tr>
<td>64</td>
<td>9835</td>
</tr>
<tr>
<td>66</td>
<td>6921</td>
</tr>
<tr>
<td>67</td>
<td>5751</td>
</tr>
<tr>
<td>68</td>
<td>4899</td>
</tr>
<tr>
<td>70</td>
<td>12693</td>
</tr>
<tr>
<td>71</td>
<td>6651</td>
</tr>
<tr>
<td>37</td>
<td>6930</td>
</tr>
<tr>
<td>38</td>
<td>375</td>
</tr>
<tr>
<td>40</td>
<td>247</td>
</tr>
<tr>
<td>41</td>
<td>6120</td>
</tr>
<tr>
<td>43</td>
<td>6129</td>
</tr>
<tr>
<td>45</td>
<td>11618</td>
</tr>
<tr>
<td>46</td>
<td>12381</td>
</tr>
<tr>
<td>47</td>
<td>12659</td>
</tr>
<tr>
<td>48</td>
<td>7797</td>
</tr>
<tr>
<td>51</td>
<td>8818</td>
</tr>
<tr>
<td>52</td>
<td>6924</td>
</tr>
<tr>
<td>53</td>
<td>5675</td>
</tr>
<tr>
<td>54</td>
<td>4032</td>
</tr>
<tr>
<td>19</td>
<td>4237</td>
</tr>
<tr>
<td>20</td>
<td>2745</td>
</tr>
<tr>
<td>21</td>
<td>10230</td>
</tr>
<tr>
<td>25</td>
<td>18881</td>
</tr>
<tr>
<td>26</td>
<td>16603</td>
</tr>
<tr>
<td>27</td>
<td>4823</td>
</tr>
<tr>
<td>29</td>
<td>15640</td>
</tr>
<tr>
<td>30</td>
<td>18379</td>
</tr>
<tr>
<td>31</td>
<td>6461</td>
</tr>
<tr>
<td>34</td>
<td>8566</td>
</tr>
<tr>
<td>35</td>
<td>14345</td>
</tr>
<tr>
<td>79</td>
<td>265</td>
</tr>
<tr>
<td>81</td>
<td>298</td>
</tr>
<tr>
<td>82</td>
<td>285</td>
</tr>
<tr>
<td>86</td>
<td>201</td>
</tr>
</tbody>
</table>
Appendix B: Observed and predicted data using the final 3D model (3DB) plotted in sets of 10 traces from all OBSs, alternating to show the match between the arrivals along the closest profile to the OBSs. The data shown are band-pass filtered (observed data) and trace-by-trace normalised.
Appendix C: Data residual plots of the normalised observed and synthetic datasets for the 2D, 2.5D and 3D models for OBSs of line 2 along the closest shooting profile to the instruments. Each panel indicates the OBS number and the RMS value of the residuals.