Coherence-enhancing anisotropic diffusion filter for 3D high resolution reconstruction of P-wave velocity and density using full waveform inversion: application to a North Sea Ocean Bottom Cable dataset

Ludovic Métivier, Romain Brossier, Alexandre Hoffmann, Jean-Marie Mirebeau, Giuseppe Provenzano, Alizia Tarayoun, Peng Yong

Coherence-enhancing anisotropic diffusion filter for 3D high resolution reconstruction of P-wave velocity and density using full waveform inversion: application to a North Sea Ocean Bottom Cable dataset

Ludovic Métivier[1,2], Romain Brossier[2], Alexandre Hoffmann[2], Jean-Marie Mirebeau[3], Giuseppe Provenzano[2], Alizia Tarayoun[2] and Peng Yong[2]

[1] *Univ. Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France*

[2] *Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France*

[3] *Univ Paris Saclay, Centre Borelli, F-91190 Gif-sur-Yvette, France*

(September 13, 2023)

Running head: **Coherence enhancing filter for FWI**

ABSTRACT

Regularization is a central topic in the study of the solutions of ill-posed inverse problems. High resolution seismic imaging using full waveform inversion belongs to this category of problems. Regularization through anisotropic diffusion, a technique which emerged in the field of image processing, is an interesting alternative to conventional regularization strategies. Exploiting the structural information of a given image, it has the capability to smooth this image along its main structures. The main difficulty is how to design the anisotropic diffusion operator. The concept of coherence enhancing proposed in 2D is extended in

3D and applied so as to filter and enhance the structural coherence of the model updates within a full waveform inversion algorithm. The benefits of this strategy are investigated on a 2D synthetic experiment before considering the multi-parameter inversion of a 3D field dataset from the North sea up to 10 Hz. From this data, the vertical velocity and the density are simultaneously reconstructed. Compared with a conventional nonstationary Gaussian regularization strategy, the models obtained using the coherence enhancing anisotropic diffusion strategy show an enhanced resolution, especially for the density model. The high resolution reflectivity image computed from the impedance volume clearly illustrates the benefit this filtering approach can deliver in terms of structural interpretation.

INTRODUCTION

Full waveform inversion (FWI) has become a prominent seismic imaging tool during the last decade. It is used at various scales, from global-scale (Bozdağ et al., 2016; Karaoğlu and Romanowicz, 2018; Lei et al., 2020; Thrastarson et al., 2022), lithospheric/continental scales (Yuan et al., 2014; Fichtner and Villaseñor, 2015; Lu et al., 2020) and deep crustal scales (Górszczyk et al., 2017), to exploration scale with dense acquisition and active sources, with depth of investigation reaching few kilometers (Sirgue et al., 2010; Plessix and Perkins, 2010; Warner et al., 2013; Stopin et al., 2014; Vigh et al., 2014; Operto et al., 2015; Raknes et al., 2015; Solano and Plessix, 2019). In the last years, smaller scale applications have also been increasingly investigated: from near surface scales (Koehn et al., 2018; Irnaka et al., 2022) to medical imaging (Guasch et al., 2020; Marty et al., 2021). A review on FWI and its application can be found for instance in Virieux et al. (2017).

The essence of the FWI methodology is to match observed data with synthetic data generated from numerical modeling (Lailly, 1983; Tarantola, 1984). This matching procedure is formulated as a local minimization problem. An initial estimate of the model is updated iteratively following descent directions computed from the gradient of the function calculating the data mismatch. From a mathematical stand-point, FWI is an ill-posed inverse problem in a large dimensional space, requiring the use of regularization techniques to reduce the size of the search space (Virieux and Operto, 2009; Virieux et al., 2017).

From this perspective, regularization can be interpreted as a way to inject prior information on the solution. This prior information can come from external sources, such as well logs or geological interpretation. It is also often injected in terms of prior assumptions

on the smoothness/roughness of the medium under investigation. From diffraction tomography analysis (Devaney, 1984; Wu and Toksöz, 1987; Sirgue and Pratt, 2004), the highest achievable resolution is equal to one-half of the local shortest wavelength. However, the numerical simulation of wave propagation underlying the FWI process requires a spatial discretization beyond this limit to avoid numerical dispersion.

For this reason, the common practice for most FWI applications consists in using smoothing operators to restrain the exploration of the solution space to models satisfying a certain level of smoothness. This can be enforced in several ways. A well-known strategy in the context of the solution of ill-posed inverse problems is the addition of a penalty term. This penalty term measures the departure of the solution from a prior model, and/or the norm of the spatial derivatives of the model. This strategy is known as the Tikhonov strategy (Tikhonov et al., 2013). It has the advantage of being easy to implement with the addition of one or several penalty terms to the misfit function whose gradients are straightforward to compute. One drawback is related to the need to set the weights of these penalty terms (hyperparameters) which is not trivial and is usually performed through trial and error. Another drawback, more fundamental, is that these penalty terms ensure the satisfaction of the constraints only at convergence, *i.e.* when the data misfit becomes comparable to the penalty term value.

An alternative approach consists in applying a smoothing operator to the descent direction along the course of iterations. This provides an immediate remedy to the aforementioned problem: with this technique the smoothness of the model is guaranteed at each iteration. From a mathematical stand-point, applying this smoothing operator can be inter-

preted as a modification of the scalar product used to express the Taylor's development of the misfit function at the core of local optimization (Zuberi and Pratt, 2017; Adriaens et al., 2021). This provides a nice mathematical framework to include different smoothing operators with different properties. A common choice is to use a Gaussian smoothing operator, potentially non-stationary, with coherent lengths based on an estimation of the expected local resolution. The latter is estimated from the local velocity and a reference frequency (for instance the peak frequency of the inverted data). This strategy has been implemented in early FWI applications to field data (Operto et al., 2006)

If the local structure of the model can be inferred from another source of information, for instance from geological interpretation or from migrated/reflectivity images, it is possible to inject it in the smoothing operators. The usual way to do this is to extract dip and strike angle maps and perform a directional smoothing, to smooth strongly along the structures and weakly across the structures. This oriented smoothing can be implemented in many ways, for instance through directional Laplacian filtering (Guitton et al., 2012) or directional Bessel filtering (Trinh et al., 2017). One difficulty is related to the fact that information on the structure is not always available when performing FWI, or not sufficiently reliable to extract relevant information. Also this information is fixed at the beginning of the inversion and can not be easily or automatically updated: the two processes, FWI on one side, extraction of the structural information on the other side, are separate processes, which leads to a lack of flexibility.

There is an exception to this situation, which should be noted here though. In the context of reflection oriented waveform inversion (RWI, Xu et al. (2012) or JFWI, Zhou et al.

(2015)) the subsurface is parameterized with a smooth macro-velocity model and a reflectivity model which are jointly updated. It is thus possible to extract structural information from the updated reflectivity model to inject it to the smoothing operator applied in the reconstruction of the macro-velocity model. This strategy has been applied in Provenzano et al. (2022) based on a Bessel filtering approach. The same could be performed in the frame of migration velocity analysis (Symes, 2008), which relies on the same scale separation between macro-velocity and reflectivity, and updates these two parameters in an iterative fashion. This has not yet been implemented in this framework to the best of our knowledge.

Edge preserving smoothing through Total Variation (TV) regularization is also a conventional technique applied in FWI, with a special interest for the reconstruction of high contrast bodies such as salt bodies in exploration case studies (Strong and Chan, 2003; Peters and Herrmann, 2017; Anagaw and Sacchi, 2018; Aghamiry et al., 2020). The boundary of these structures is sharp, while the mechanical properties are almost constant within them, making TV regularization an appropriate tool for their reconstruction.

We shall also mention that a generalization of the concept of smoothing to make the model consistent with respect to the expected resolution has been proposed in Capdeville and Métivier (2018), based on the homogeneization theory (*i.e.* the theory of equivalent media). Homogeneization theory for elastic wave propagation highlights the fact that a finite-frequency band wavefield behaves in a diffraction regime for subsurface heterogeneities down to the smallest propagated wavelengths (Capdeville and Marigo, 2013). For hetero-

geneities smaller than this reference scale, the effect on the wavefield is of an apparent anisotropy. A more general and elegant way to inject prior information on the smoothness of the medium is thus proposed in Capdeville and Métivier (2018) through homogeneization, where the solution of the FWI problem is searched in the space of equivalent media related to the bandwidth of the inverted data. This method is however currently at a more experimental stage and suffers from several drawbacks. First, the homogeneization theory and associated numerical tools are mainly developed for elastic media while most of the industrial applications of FWI are performed in the acoustic approximation. Homogeneization theory for acoustic media actually exists, and in the context of constant density media, it amounts to smooth the velocity model, which is consistent with the common practice of FWI (this is not true for variable density media however, an issue well explored in Cance and Capdeville (2015)). Second, implementing robust and efficient 3D homogeneization algorithms is still a challenge. An ill-conditioned elasto-static problem needs to be solved which incurs limitations for 3D applications (see Cupillard et al., 2020, for instance). Third, in the elastic approximation, the medium after homogeneization is fully anisotropic (21 coefficient stiffness tensors) and how to deal with such parameterization in the context of an inverse problem is not straightforward.

In this study, we are interested in an alternative smoothing approach coming originally from image processing (Weickert, 1998). The idea is first to rely on an anisotropic diffusion equation to smooth the gradient. Second, the anisotropic diffusion operator is built upon the structure tensor of the image. The diffusion is set to be strong in the direction of slow variations and weak in the direction of fast variations so as to preserve the main structures of the

image. This idea has been exploited for instance in Lewis et al. (2014) and Lee and Pyun (2018) and to a certain extent in Xue et al. (2020). In those studies, the design of the diffusion operator is not really detailed, although it relies on the structure tensor associated with the gradient and its spectral decomposition. What we propose here is to generalize to 3D the coherence-enhancing diffusion strategy introduced by Weickert (1998) which is a specific way to design the anisotropic diffusion operator. It uses a local measure of coherence along directions in the plane orthogonal to the direction of fastest variations, and a threshold function to control the transition from weak diffusion to strong diffusion depending on the orientation. We shall add that what makes this diffusion-type technique appealing over other aforementioned directional filtering techniques are the mathematical properties inherited from the partial differential equations formalism (preservation of the maximum, minimum and mean value, and the scale-space property) as well as the automation of the structure detection through the design of the anisotropic diffusion operator, without having to compute angle maps and correlation lengths prior the application of the filter, and the relatively limited number of tuning parameters (essentially the diffusion time). We have studied this technique already in the frame of 2D FWI, with the idea that it could be applied in the model space to filter the gradient, and in the data space to filter and denoise the data while enhancing its low frequency content (Métivier and Brossier, 2022). In the present study, after providing some illustrations on a 2D synthetic example, we apply it to the 3D multi-parameter inversion of a North Sea OBC dataset up to 10 Hz. We provide a comparison between applying a conventional non-stationary Gaussian smoothing and our coherence enhancing smoothing. We show that an isotropic Gaussian smoothing has a tendency to destroy the structure, especially at depth, which is detrimental to the final model

resolution and also penalizes the convergence of the whole FWI algorithm. Conversely, the FWI strategy coupled with the coherence enhancing smoothing makes it possible to recover sharper interfaces in the whole domain. Then this also enhances the reconstruction of a sharper density model. The coherence enhancing strategy is also efficient to remove the footprint of the acquisition in the shallow part of the model, which is all the more important when a source subsampling method is used to decrease the computational cost of FWI as is the case in our 3D OBC application. We believe these features are particularly valuable for high resolution FWI, a current trend in the industry to push FWI towards directly interpretable models inverted from data in frequency bands similar as what is used for migration (Shen et al., 2018; Wang et al., 2019; Huang et al., 2021; Kalinicheva et al., 2020).

The structure of the study is as follows. In the next Section we introduce the methodology of our 3D coherence enhancing smoothing. Then we illustrate it on a 2D simple synthetic experiment. We present its application to the inversion of 3D OBC field data. A discussion of our results follows, before we present conclusion and perspectives.

METHODOLOGY: A NONLINEAR ANISOTROPIC DIFFUSION APPROACH

Generalities

The structural smoothing technique we propose originates from the work of Weickert (1998) for image processing. Consider an image $I(\mathbf{x})$ with $\mathbf{x} \in \Omega \subset \mathbb{R}^d$, d being the

model dimension. In what follows, $d = 2$ for our synthetic example and $d = 3$ for the OBC data application.

The fundamental idea consists in considering the following nonlinear anisotropic diffusion process to generate a smooth image $I^s(\mathbf{x})$ from $I(\mathbf{x})$. Solve the partial differential equations

$$\begin{cases} \dfrac{\partial u}{\partial t} - \text{div}\left(D(u)\nabla u\right) &= 0, \text{ on } \Omega \times [0,T] \\ u(\mathbf{x},0) &= I(\mathbf{x}), \text{ on } \Omega \\ \langle D(u)\nabla u, \mathbf{n}\rangle &= 0, \text{ on } \partial\Omega \times [0,T], \end{cases} \tag{1}$$

where $I(\mathbf{x})$ is the initial condition for the nonlinear anisotropic diffusion described by the d-by-d diffusion matrix $D(u)$, $\mathbf{n}(\mathbf{x}) \in \mathbb{R}^d$ is the vector normal to the boundary at $\mathbf{x} \in \partial\Omega$, and $< ., . >$ denotes the usual Euclidean scalar product in $\mathbb{R}^d$. The smooth image $I^s(\mathbf{x})$ is defined as

$$I^s(\mathbf{x}) = u(\mathbf{x}, T), \tag{2}$$

for a chosen diffusion time T. The nonlinearity comes from the dependency of D to u. The anisotropy comes from the fact that D is a matrix. In the case where $D = I_d$, where I_d is the identity matrix, we recover the heat equation. In this case, there is an analytic solution for equation 1 which is expressed as the convolution of the initial condition with the Gaussian kernel $K_{\sqrt{2T}}(\mathbf{x})$:

$$I^s(\mathbf{x}) = K_{\sqrt{2T}}(\mathbf{x}) * I(\mathbf{x}), \tag{3}$$

with $*$ the convolution in space and

$$K_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma^2}\right). \tag{4}$$

This shows the equivalence between isotropic diffusion and Gaussian smoothing.

The originality of the approach developed by Weickert (1998) is to consider an anisotropic diffusion process based on a specific matrix $D(u)$ computed from the structure tensor matrix $S(u)$, defined by

$$S(u) = \nabla u \nabla u^T. \tag{5}$$

A Singular Value Decomposition (SVD) of $S(u)$ provides useful information. It has one positive eigenvalue $\lambda_1 = |\nabla u|^2$ and $d-1$ eigenvalues equal to 0. The eigenvector associated with λ_1 is parallel to the gradient ∇u, while the others are perpendicular to the gradient. Note that an image might not be differentiable everywhere, if it exhibits strong small scale variations, for instance due to the presence of noise. Thus a pre-smoothing is applied to u to remove spurious oscillations smaller than a given reference scale σ, yielding the image u_σ, such that

$$u_\sigma = K_\sigma * u. \tag{6}$$

Furthermore, the information embedded in $S(u)$ being strictly local, Weickert (1998) proposes to use a local average over specific scales to analyze the image structure. We thus consider the matrix $S_{\ell,\sigma}(u)$ defined by

$$S_{\ell,\sigma}(u) = K_\ell . * \left(\nabla u_\sigma \nabla u_\sigma^T \right), \tag{7}$$

where $.*$ denotes a term-by-term convolution operation (the convolution applies to each entry of the matrix). The scale ℓ is referred to as the coherence scale: it is a length over which the image is going to be averaged to perform the local analysis of its structure.

2D coherence enhancing filter

To fix ideas, in the 2D case ($d = 2$), $S_{\ell,\sigma}(u)$ has two eigenvalues $\lambda_1 \geq \lambda_2 \geq 0$. The eigenvector v_1 (resp. v_2) gives the direction in which $u(\mathbf{x})$ varies the most rapidly (resp. the most slowly). In what follows, the symbol $\gg$ indicates "significantly larger than". Different situations can occur, corresponding to particular structures in the image:

- $\lambda_1 \simeq \lambda_2 \simeq 0$ identify zones with almost constant values;
- $\lambda_1 \gg \lambda_2 \simeq 0$ identify zones with sharp edges;
- $\lambda_1 \geq \lambda_2 \gg 0$ identify corners.

In addition the quantity $(\lambda_1 - \lambda_2)^2$ is a measure of the local coherence of the image, which becomes large as soon as anisotropic structures with sharp edges are involved. This quantity is key in the design of the diffusion operator $D(u)$.

Weickert (1998) proposes the following definition of $D(u)$ for a coherence-enhancing diffusion filter. Let $S_{\ell,\sigma}(u) = P(u)^T \Lambda(u) P(u)$ be the SVD decomposition of $S_{\ell,\sigma}(u)$. The diffusion matrix $D(u)$ is defined as

$$D(u) = P(u)^T \Sigma(u) P(u), \tag{8}$$

where $\Sigma(u)$ is defined as

$$\Sigma(u) = \begin{pmatrix} \alpha & 0 \\ 0 & g\left((\lambda_1(u) - \lambda_2(u))^2\right) \end{pmatrix}, \tag{9}$$

with g a threshold function from α to 1

$$g(x) = \begin{cases} \alpha & \text{if} \quad x = 0 \\ \alpha + (1-\alpha)\exp\left(\frac{-C}{x}\right) & \text{else,} \end{cases} \tag{10}$$

where α and C are user-defined scalar values. A profile of the threshold function $g(x)$ depending on C for $\alpha = 10^{-2}$ is given in Figure 1. Playing with the constant C amounts to playing with the steepness of the threshold (how fast the function moves from α to 1). Of note, in all the following numerical experiments with the diffusion filter, we use the steepest threshold function presented here, which corresponds to the parameter values $\alpha = 10^{-2}$ and $C = 10^{-8}$. After testing different values it turns out that a stiff transition provides qualitatively better results for our application.

[Figure 1 about here.]

The diffusion matrix $D(u)$ shares the same eigenvectors as $S_{\ell,\sigma}(u)$ but with different eigenvalues. Typically α takes small values while $g(x)$ increases rapidly toward 1. When zones of large coherence are detected the quantity $(\lambda_1 - \lambda_2)^2$ becomes large, and the diffusion weight along the slow variation direction v_2 increases rapidly toward 1, resulting in a stronger smoothing along v_2. In the opposite case, the smoothing remains isotropic (approximately same weight α in both direction) and relatively weak (α should be small). This adaptive diffusion makes it possible to follow the orientation of the main structures of the image, to smooth along the structures and not across them.

3D coherence enhancing filter

In this study, we generalize this concept to 3D. Considering that in 3D the matrix $S_{\ell,\sigma}(u)$ has three eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ and corresponding eigenvectors v_1, v_2, v_3, the following situations can occur:

- $\lambda_1 \simeq \lambda_2 \simeq \lambda_3 \simeq 0$ identify zones with almost constant values;
- $\lambda_1 \gg \lambda_2 \simeq \lambda_3 \simeq 0$ identify a planar structure in (v_2, v_3) with sharp discontinuity along v_1;
- $\lambda_1 \geq \lambda_2 \gg \lambda_3 \simeq 0$ identify a one dimensional structure in v_3 orthogonal to the plane (v_1, v_2);
- $\lambda_1 \geq \lambda_2 \geq \lambda_3 \gg 0$ identifies a corner.

The quantities $(\lambda_1 - \lambda_2)^2$ and $(\lambda_1 - \lambda_3)^2$ now measure the coherence along direction v_2 and v_3 respectively.

Based on this interpretation, we extend in this study the coherence-enhancing diffusion filter in 3D by defining the diffusion matrix $D(u)$ as

$$D(u) = P(u)^T \Sigma(u) P(u), \tag{11}$$

where $\Sigma(u)$ is defined as

$$\Sigma(u) = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & g\left((\lambda_1(u) - \lambda_2(u))^2\right) & 0 \\ 0 & 0 & g\left((\lambda_1(u) - \lambda_3(u))^2\right) \end{pmatrix}, \tag{12}$$

with g the same threshold function as in equation 10.

In the same spirit as in the 2D case, following this definition of $D(u)$, as soon as coherent structures are detected along v_2 or v_3 by an increase of the quantities $(\lambda_1 - \lambda_2)^2$ or $(\lambda_1 - \lambda_3)^2$, the diffusion weight increases rapidly to 1 while remaining small and equal to α in the direction of the main variation v_1. If no coherent structure is detected, the diffusion remains close to isotropic, with similar small α values in each direction. This extends the coherence-enhancing filter to the 3D case, with again the ability to smooth along detected structures and not across them. As for the 2D case, the detection of the structures is performed automatically thanks to the SVD of the averaged structure tensor $S_{\ell,\sigma}(u)$.

Discretization of anisotropic diffusion equations

Designing discretization schemes for anisotropic partial-differential equations is not a trivial task and is the matter of dedicated mathematical studies. In Weickert (1998), it is shown that the coherence enhancing anisotropic diffusion filter presented above enjoys very interesting mathematical properties at the continuous level, such as: conservation of the mean value, conservation of the minimum and maximum value (also known as min/max principle), invariance to translation and rotations, and scale-space properties. The latter could be summarized as the property to generate smoother and smoother images by repetition of the application of the filter, without introducing structures with a finer scale.

How to preserve these properties at the discrete level is a key question for designing discretization schemes for these equations. One central property is to be able to generate

non-negative discrete matrices $D(u)$, *i.e.* matrices with non-negative off-diagonal terms. In Weickert (1998) a 2D scheme is proposed, which guarantees this non-negativity for any diffusion matrix such that the ratio λ_1/λ_2 between the highest and lowest eigenvalue in $D(u)$ is bounded by $1 + \sqrt{2} \approx 2.41$. This is the scheme we have used in Métivier and Brossier (2022). However no extension to 3D is proposed in Weickert (1998).

In the current study, we rely on more sophisticated schemes proposed in Fehrenbach and Mirebeau (2014). These schemes are based on a Lattice-Basis-Reduction technique. They can be implemented in 2D and 3D, and can guarantee the non-negativity of the discretization of $D(u)$ for any value of λ_1/λ_2 or λ_1/λ_3. The stencils are adapted to finite-difference Cartesian meshes. The computational complexity to build the stencils is low, which makes the method very efficient. Describing in detail how these schemes are derived is outside the scope of this paper, for this we refer the interested reader to Fehrenbach and Mirebeau (2014) and references therein.

Implementation within a full waveform inversion algorithm

We formulate FWI as

$$\min_m f(m) = G\left(d_{cal}[m], d_{obs}\right), \tag{13}$$

where $G(d_1, d_2)$ is a general positive function measuring the misfit between two datasets d_1 and d_2, d_{obs} is the observed data, and $d_{cal}[m]$ is the calculated data obtained through

$$d_{cal}[m] = Ru[m], \quad A(m)u = s. \tag{14}$$

In equation 14, $u[m]$ is the modeled wavefield, solution of a wave propagation problem represented by the operator $A(m)$, which can span a variety of wave propagation physics, from constant density acoustic to the visco-elastic anisotropic approximation. The right-hand side s represents the seismic active source, and R is a restriction operator extracting the value of the wavefield at the receiver positions to build the calculated data. In equations 13 and 14, m represents the subsurface parameters we want to reconstruct through FWI, which depends on the choice of the operator $A(m)$. In this study, we focus on the visco-acoustic approximation, and we perform a multi-parameter inversion for the P-wave velocity model and the density model in the field data application.

The local optimization scheme for the solution of the FWI problem is: starting from an initial model estimation m_0, build the sequence

$$m_{k+1} = m_k + \alpha_k \Delta m_k, \tag{15}$$

where α_k is a linesearch parameter satisfying the Wolfe's conditions (Nocedal and Wright, 2006) to ensure convergence towards the nearest local minimum. The descent direction Δm_k is computed as

$$\Delta m_k = -Q_k \mathcal{F} \left(\nabla f(m_k) \right), \tag{16}$$

where the operator $\mathcal{F}$ stands for the filtering/smoothing operator we apply to the gradient and Q_k is a preconditioning matrix. We rely on the l-BFGS algorithm to estimate Q_k as an approximation of the inverse Hessian operator $\left(\nabla^2 f(m_k)\right)^{-1}$ from l previously computed filtered gradient values $\mathcal{F}\left(\nabla f(m_k)\right), \mathcal{F}\left(\nabla f(m_{k-1})\right), \ldots, \mathcal{F}\left(\nabla f(m_{k-l+1})\right)$ (Nocedal, 1980). Note that the $l-$BFGS method makes it possible to combine Q_k with any other preconditioning operator P approximating $\left(\nabla^2 f(m_k)\right)^{-1}$ (Métivier and Brossier,

2016).

In this study, we illustrate the benefits that can be obtained from using a coherence enhancing diffusion filter rather than a classical non-stationary Gaussian filter. As a reminder, the latter can be formulated as

$$I^s(\mathbf{x}) = \mathcal{G}(\mathbf{x}) * I(\mathbf{x}), \tag{17}$$

where

$$\mathcal{G}(\mathbf{x}) = \exp\left(-\frac{x^2}{2\sigma_x(\mathbf{x})^2}\right)\exp\left(-\frac{y^2}{2\sigma_y(\mathbf{x})^2}\right)\exp\left(-\frac{z^2}{2\sigma_z(\mathbf{x})^2}\right). \tag{18}$$

The nonstationarity of the filter comes from the functions $\sigma_x(\mathbf{x}), \sigma_y(\mathbf{x}), \sigma_z(\mathbf{x})$ which depend on the space variable $\mathbf{x}$. Based on diffraction tomography analysis (Devaney, 1984; Wu and Toksöz, 1987; Sirgue and Pratt, 2004), we relate these functions to a fraction of the local wavelength through the relations

$$\sigma_x(\mathbf{x}) = \frac{r_x v_P(\mathbf{x})}{f_0},\ \sigma_y(\mathbf{x}) = \frac{r_y v_P(\mathbf{x})}{f_0},\ \sigma_z(\mathbf{x}) = \frac{r_z v_P(\mathbf{x})}{f_0}, \tag{19}$$

where $v_P(\mathbf{x})$ is the P-wave velocity field, f_0 is a user-defined reference frequency, and r_x, r_y, r_z are three user-defined scalar parameters. The latter parameters determine the fraction of the local wavelength we want to take into account in the Gaussian filter.

To implement this filter in an efficient way, we exploit the separability of the kernel $\mathcal{G}(\mathbf{x})$ to perform a series of 1D convolutions in space. This makes the computational cost of the application of this nonstationary Gaussian filter negligible compared to the gradient computation as is shown in the numerical experiments presented in the sequel.

SYNTHETIC EXPERIMENTS

We start by illustrating the properties of our coherence enhancing diffusion filter on a 2D synthetic case study based on the Marmousi II model.

Data generation

We generate observed data in the 2D acoustic approximation with variable density using our 2D/3D time-domain full waveform modeling and inversion engine TOYxDAC_TIME, developed in the frame of the SEISCOPE project (Yang et al., 2018). This code uses a 4th order finite-differences spatial discretization and a 2nd-order finite-differences time discretization of the velocity/stress equations. The reference data is built from the velocity and density models presented in Figure 2(a,b). These models have been obtained by upscaling on a 5 m grid the original 1.25 m grid Marmousi II models (Martin et al., 2006). To generate the data, we use a fixed spread acquisition with 128 shot positions each 132.5 m apart, and 169 receivers each 100 m apart. The source wavelet, presented in Figure 2(c) is a Ricker wavelet centered on 5 Hz, low-cut below 2.5 Hz to mimic a realistic exploration scenario in which such low frequencies are not available. The spectrum of the wavelet is presented in Figure 2(d). A Gaussian white noise filtered in the frequency-band from 0 to 12.5 Hz is added to the data. Two shot gathers of this reference dataset are presented in Figure 3.

[Figure 2 about here.]

[Figure 3 about here.]

Initial model and source wavelet estimation

To generate the initial velocity model, we smooth the exact model on the 5 m grid using an isotropic Gaussian smoothing with a correlation length of 1250 m and we project it on a 25 m grid. This is the "coarse" grid that we use for gradient computation and inversion. We compute an initial density model using Gardner's law (Gardner et al., 1974)

$$\rho(\mathbf{x}) = 1741 \times \left(\frac{V_P(\mathbf{x})}{1000}\right)^{0.25}. \tag{20}$$

These initial models are presented in Figure 4(a,b). We estimate the source wavelet from these initial velocity and density models using the deconvolution approach of Pratt (1999). The resulting estimated source wavelet and its spectrum are presented in Figure 4(c,d). Due to the inaccuracy of the initial velocity and density models and the noise on the data, the estimated wavelet exhibits a lower amplitude than the true wavelet as well as a less smooth amplitude spectrum.

[Figure 4 about here.]

Gradient: comparison between Gaussian and anisotropic diffusion filter

We start by comparing the FWI gradient obtained with the conventional Gaussian smoothing described in equations 17 to 19 with the one obtained through the anisotropic diffusion smoothing considered in this study.

For the Gaussian smoothing, we set f_0 in equation 19 to 5 Hz, which corresponds to the

central frequency of the wavelet used to generate the data. In Figure 5 we present the gradient without smoothing together with the gradient after a Gaussian smoothing is applied, with $r_x = r_z$ and values for these two parameters varying from 0.1 to 0.5. The gradients are presented alone (left column), and with level set superimposed (right column) to emphasize the structural information. It can be seen that, as (r_x, r_z) increase, the smoothing effect is more and more pronounced. The structures at depth, below 2 km, are also progressively erased from the gradient.

In comparison, we present in Figure 6 the same progressive smoothing using this time the coherence enhancing diffusion filter, with diffusion iterations going from 50 to 800. In this experiment the noise and coherence scales σ and ρ are both set to 50 m (two discretization points on the coarse grid). As the number of iterations increases, the smoothing effect is also more pronounced however it appears that the structural information is preserved along the smoothing process, especially at depths below 2 km. This displays all the interest of using this filter for FWI: the spurious small oscillations due to noise and uneven illumination are removed, while the main structures of the gradient are preserved.

To further highlight this point, we compare in Figure 7 the gradient obtained with Gaussian smoothing and $r_x = r_z = 0.4$ with the gradient obtained with nonlinear anisotropic diffusion and 200 iterations. This provides approximately the same "level" of smoothing. However, one can appreciate how much of the structural information is preserved by using the diffusion filter. These observations are confirmed by k_x, k_z wavenumber spectra displayed in Figures 7(b,d). We see that the extent of the two spectra is approximately the same in the k_x dimension, while it is larger in the k_z dimension for the gradient obtained

by anisotropic diffusion. The loss of structural information using the Gaussian smoothing appears indeed mostly in the vertical direction, while the level of details in the horizontal direction is approximately the same for the two gradients.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

Computational cost: linear vs nonlinear filter, Weickert vs Fehrenbach stencil

What is the increase in computational cost due to the nonlinear anisotropic diffusion filter? We present some statistics in Table 1. The computational cost is driven here by the solution of the forward and adjoint wave equations. The smoothing time using the non-stationary Gaussian filter is negligible. When it comes to the coherence enhancing diffusion filter, together with the stencil of Fehrenbach and Mirebeau (2014), this cost increases to 25% of the total computational time for the gradient. This is non-negligible and might be problematic in the perspective of 3D applications. The computational cost here might be actually driven by two factors: the finite-difference stencil, which might be wider when using the one from Fehrenbach and Mirebeau (2014) instead of the one from Weickert (1998), and the nonlinearity of the diffusion filter, which requires us to rebuild the diffusion tensor, and to perform as many SVD as grid points, at each iteration of the diffusion process.

As can be seen in Table 1, using the same nonlinear anisotropic diffusion with the stencil from Weickert (1998) slightly reduces the computational cost to 23%. In comparison, switching to a linear anisotropic diffusion process drastically reduces the computational effort: the time for smoothing is reduced to 3.7% of the total time for building the gradient.

We compute the relative error $e(\mathbf{x})$ between the gradient $g_{nonlin}(\mathbf{x})$ obtained after nonlinear anisotropic diffusion filter and the gradient $g_{lin}(\mathbf{x})$ obtained after using a linear anisotropic diffusion filter as

$$e(\mathbf{x}) = 100\frac{g_{nonlin}(\mathbf{x}) - g_{lin}(\mathbf{x})}{||g_{nonlin}||_{L^2}}. \tag{21}$$

In Figure 8, we present this error map for the gradients presented in Figure 7 and the ones obtained with a linear anisotropic diffusion filter. The highest error value reaches locally 0.5%. The error is weaker and more localized for a small number of diffusion iterations (50) and simultaneously increases in amplitude and spreads over the model as the number of iterations grows (up to 800). This is expected: the nonlinear effect increases with the number of iterations as, on the one hand, the diffusion matrix $D(u)$ is recomputed at each diffusion iteration (nonlinear diffusion) while, on the other hand, it is kept the same for all diffusion iterations (linear diffusion).

Overall, the error remains small and relatively localized for the number of iterations considered here. This indicates that the nonlinearity of the filter does not play a significant role in this example. In the remainder of the study, we will thus rely on a linear anisotropic filtering, for which the anisotropic diffusion matrix $D(u)$ is computed at the first diffusion iteration and is not updated throughout the diffusion iterations. We will also use the Fehrenbach and Mirebeau (2014) scheme, which presents better numerical properties and

is developed for 2D and 3D filtering.

[Figure 8 about here.]

[Table 1 about here.]

Inversion: comparison between Gaussian and linear anisotropic diffusion filter

To conclude this 2D synthetic test, we compare the results obtained after 50 FWI iterations using the non-stationary Gaussian filter and the linear anisotropic diffusion filter, starting from the P-wave velocity and density models presented in Figure 4(a,b). We invert only for the P-wave velocity, the density is not updated (passive parameter). We use the estimated wavelet presented in Figure 4(c,d). The results are presented in Figure 9. While the final models (Fig. 9(a,b)) are similar, especially in the shallowest part above 2 km depth, the resolution of the final model obtained with the anisotropic diffusion filter is higher. The bent layering structure below 3 km depth is better reconstructed especially in the central part between 6 and 12 km in horizontal distance. The central structure, between 1.5 and 2.5 km depth and between 6 and 12 km in horizontal distance is also better resolved, with higher contrasts between fast and slow layers. This is confirmed in Figure 9(c,d) where we compare the decrease of the data misfit and the model misfit along the FWI iterations, depending on the choice of the filter. Using the anisotropic diffusion filter, both the data and model errors decrease faster.

[Figure 9 about here.]

APPLICATION TO THE 3D OBC FIELD DATA

We apply in this section our coherence enhancing diffusion filter and compare it to the standard nonstationary Gaussian convolution filter in the framework of time-domain FWI of a 3D OBC field data. Unlike most of the studies performed on a similar environment (Sirgue et al., 2010; Operto et al., 2015; Kamath et al., 2021; Pladys et al., 2022), we consider here a multi-parameter P-wave velocity (v_P), density (ρ) inversion. The reconstruction of the density together with the vertical velocity from a similar dataset has been tackled only in the frequency-domain in the work of Operto and Miniussi (2018).

Field data presentation and pre-processing

The data we consider has been recorded in a shallow water environment, in the North sea. Several FWI studies have focused on similar data since the seminal paper of Sirgue et al. (2010), which presents the first 3D FWI application to industrial field data. The particular dataset we use is a 4-components ocean bottom cable (OBC) dataset, with 2048 receivers deployed on the seabed and 50824 shots at 5 m depth. The area covered by the shots is around 145 km^2. We present in Figure 11 the acquisition layout. The field is characterized by a shallow water environment (the water column is approximately constant equal to 70 m in the whole domain) and an anticlinal in chalk in the Upper Cretaceous Hod and Tor formations. An apparent low velocity anomaly in the upper part of the model from 1 km

to 2 km depth in the center of the volume is the main target in terms of imaging, as well as the structure below it, the imaging of which is made difficult by this low velocity anomaly. The Tertiary overburden is relatively simple and free of complex structures.

[Figure 10 about here.]

[Figure 11 about here.]

We focus on the hydrophone component of the data. We apply source-receiver reciprocity to reduce the computational cost. Therefore, in the following, what is called common shot gather (CSG) refers actually to a common receiver gather (CRG). The preprocessing of the data is minimal: we apply a simple despiking, and apply a minimum-phase band-pass filter to the data to generate three distinct datasets in the 2.5 - 5 Hz, 2.5 - 7 Hz and 2.5 - 10 Hz frequency band respectively. For each dataset, we compute the total amplitude of all CSG and remove the ones with anomalous amplitudes. Four of them are excluded in bands 2.5 - 5 Hz and 2.5 - 7 Hz, and one only is excluded in the band 2.5 - 10 Hz (the signal over noise ratio increases with the frequency bandwidth). Finally, the visco-acoustic approximation we use makes it impossible to model the Schölte waves propagating at the fluid-solid interface. These waves are thus muted from the observed data, using a simple time-windowing based on a linear velocity of approximately 400 $\mathrm{m.s}^{-1}$ for these waves. As an illustration, we present in Figure 12 two CSG corresponding to the positions of cables A and B, before and after the mute is applied, filtered in the 2.5 - 10 Hz band. We choose these CSG because cable A intersects the low velocity anomaly while cable B is further away from the low velocity anomaly. We reproduce a data anatomy anal-

ysis similar to the one presented in Operto et al. (2015). The red, white, black arrows point to the reflection from a shallow reflector, the top of the low velocity anomaly and the top of the structure below the low velocity anomaly, respectively. The solid arrows point to pre-critical reflections, while the dashed ones point to post-critical reflections.

[Figure 12 about here.]

FWI setup

The inversion results presented here are obtained using the same 2D/3D time-domain visco-acoustic code we have used for the 2D Marmousi experiment (TOYxDAC_TIME, Yang et al. (2018)). We use a 3D visco-acoustic anisotropic modeling under the vertically transverse isotropy (VTI) approximation. The most significant anisotropy effect can be approximated as a VTI anisotropy and occur mostly from 0.6 km to 3 km depth approximately. It is due to the presence of shale whose specific crystalline structure generate different wavespeeds for vertically and horizontally propagating waves (intrinsic anisotropy) and also interbedding of shale and sandstone in the shallow part, and claystone and limestone in the deeper part (extrinsic anisotropy). A similar situation is described in the work of Gholami (2012). Introducing attenuation is also found crucial to properly reproduce the field data: the low velocity anomaly exhibits a significant attenuation effect on the wavefield propagation.

Five parameters are used to describe the subsurface mechanical properties: the vertical P-wave velocity v_P, the density ρ, the quality factor Q_P, and the Thomsen parameters ϵ and

δ (Thomsen, 1986). During the inversion, we will invert only for v_P and ρ, while keeping Q_P, ϵ and δ constant, equal to their initial values (passive parameters). The initial model for v_P has been provided to us courtesy of AkerBP and is displayed in Figure 13. It is obtained through reflection travel-time tomography. On top of a sharp reflector between 2.7 and 2.8 km depth, a smooth dark blue blob is visible, indicative of the low velocity anomaly . The inline and crossline vertical sections in Figures 13(b,c) show traversing and adjacent slices to this low velocity anomaly .

The initial model for ρ is derived from the initial v_P model based on the following Gardner's law (Gardner et al., 1974)

$$\rho = 309.6 V_p^{0.25}. \tag{22}$$

It is displayed in Figure 14 and exhibits the same structure as v_P. The Q_P model is taken constant, equal to 1000 in the water layer, and equal to 200 everywhere else, as in Operto et al. (2015) and Kamath et al. (2021). It is shown in Kamath et al. (2021) that in the considered frequency band, inverting for the Q_P model and deriving a more refined Q_P estimation does not provide a substantial improvement in the data fit using a time-domain FWI algorithm. In terms of implementation, standard linear solid (SLS) approximation is used in TOYxDAC_TIME to account for the attenuation. We use three mechanisms as in Kamath et al. (2021). The ϵ and δ models are obtained by tomography and well log information respectively, and have been provided courtesy of AkerBP as well. We display the parameter η in Figure 15, computed as

$$\frac{\epsilon - \delta}{1 - 2\delta}, \tag{23}$$

which can be roughly interpreted as a percentage of VTI anisotropy (Fig.15). The VTI anisotropy is pronounced starting from above 1 km depth at the top of the low velocity anomaly down to the reflector delineating the structure below. Anisotropy is absent further below.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

We use a conventional multi-scale inversion (Bunks et al., 1995), where the model inverted in band $i-1$ serves as initial model for the model in band i. The spatial discretization step h is set respectively to 70, 50 and 35 m for frequency bands 2.5 -5 Hz, 2.5 - 7 Hz, and 2.5 - 10 Hz. This satisfies the usual criterion to have at least 4 points per wavelength using our 4-th order finite difference discretization scheme in space to minimize numerical dispersion while ensuring minimal computational cost (Igel, 2017). To decrease the computational cost of the inversion further we rely on a source subsampling strategy (Warner et al., 2013; Kamath et al., 2021; Pladys et al., 2022). The initial ensemble of shots is decomposed in 16 groups, the first 15 being composed of 128 shots, and the 16th containing 124 sources for the two bands 2.5 - 5 Hz and 2.5 - 7 Hz and 127 sources for the band 2.5 - 10 Hz. One inversion over a given frequency band then consists in performing 3 l-BFGS iterations over each group, leading to 48 iterations in total. The 16 groups do not overlap (one shot belongs to a single group) to ensure each shot is used once per inversion.

The computational savings are substantial compared to the same inversion ran with all the sources at the same time (reduction from a factor of 16) with a very weak imprint on the inversion result (Kamath et al., 2021).

As the receivers are approximately all at the same shallow water depth (70 m), with similar coupling conditions, we consider a single source wavelet for all shots after reciprocity. To estimate this common source wavelet, we rely on the conventional frequency-domain deconvolution introduced by Pratt (1999). We use a time-windowing strategy to isolate the direct waves from the data, focusing on short offsets and limited propagation time. This estimation is performed once per frequency band, over 128 randomly selected shots. Note that this subgroup does not correspond to any of the 16 groups used in the inversion.

The preconditioner we use is similar to the one used in Kamath et al. (2021, Appendix B). It applies both to v_P and ρ gradient, and consists in an illumination compensation through the wavefield values. The Hessian approximation is diagonal where each element is computed as an integration over time and over shots of the source wavefield. For the shallow part, from $z = 0$ to $z = 800$ m, these wavefield based values are replaced with a linear function of depth to avoid a strong imprint of uneven illumination associated with the source subsampling strategy. We also use a unity-based normalization for both parameters to recast the values of the inverted parameters between 0 and 1, based on a given authorized range of variation for both parameters (Yang et al., 2018). The velocity is bounded between 1280 m.s^{-1} and 3650 m.s^{-1}. The density is bounded between 1000 kg.m^{-3} and 2800 kg.m^{-3}.

As is documented in Pladys et al. (2022), outside a zone delineated by the envelope

of the shot positions, the subsurface model is weakly updated. The reason is the poor illumination hence a poor sensitivity to the seismic data in this zone. As a result, FWI has a tendency to create a fictitious boundary along this envelope that might finally result in artificial reflections in the synthetic data. To prevent this effect, we implement an extrapolation strategy using a nearest neighbor strategy. The points in the gradient outside the well illuminated zone are affected with the value of the nearest point within the well illuminated zone. We reduce the size of the extrapolation stencil progressively with depth to mimic the geometrical spreading of the wavefield and the resulting weaker illumination at depth. This is a systematization of what is implemented in Pladys et al. (2022), where the same extrapolation scheme is used on the final model after each inversion.

To illustrate the benefit we can obtain from using an anisotropic diffusion filter for FWI, we perform a twin experiment. With the exact same configuration (multi-scale strategy, source subsampling groups, source wavelet estimation, preconditioner, extrapolation) we perform two inversions, one with the nonstationary Gaussian filter, the second with the anisotropic diffusion filter. For the Gaussian filter, the reference wavelength is computed from equation 18 where the reference frequency f_0 is set to 3.75, 5.5 and 6.5 Hz respectively for the frequency bands 2.5 - 5 Hz, 2.5 - 7 Hz, and 2.5 - 10 Hz. The parameter r_x, r_y and r_z from equation 19 are chosen as

$$r_x = 0.4, \quad r_y = 0.4, \quad r_z = 0.3, \tag{24}$$

to account for the layered structure of the medium (smaller smoothing in the z direction), while remaining efficient to remove the acquisition footprint in the shallowest part of the model.

For the anisotropic diffusion filter, we rely on its linear version, and we use the 3D adaptive discretization stencil from Fehrenbach and Mirebeau (2014). The noise and coherence scale σ and ℓ are adapted to the spatial direction x, y or z to reflect the layered structure of the medium as for the Gaussian filter design. The convolution steps described in equation 7 to build the diffusion tensor are thus performed using a Gaussian kernel as in equation 18 where we would have

$$\sigma_x = 2h, \quad \sigma_y = 2h, \sigma_z = h, \tag{25}$$

with h varying from 70, 50 and 35 m depending on the frequency band as already stated. We use the same smoothing both from the noise and coherence scale ($\ell_x = \sigma_x, \ell_y = \sigma_y, \ell_z = \sigma_z$). In addition, the number of iterations for the anisotropic diffusion is set respectively to 80, 40 and 20 depending on the frequency band. These values have been set by trial and error on an initial gradient estimation for each band.

Finally, we note here that in both cases (Gaussian and diffusion filters), the same filter is applied to the P-wave velocity and density gradients.

Multi-parameter FWI results

We present in Figures 16 and 17 several 2D slices of the final 3D velocity models estimated by FWI using the Gaussian and anisotropic diffusion filters respectively. As is documented in previous studies (Sirgue et al., 2010; Operto et al., 2015; Operto and Miniussi, 2018), we can see in the constant depth sections (Fig.16(a) and Fig. 17(a)) glacial sand channels deposits at z = 175 m, a low-velocity zone intersected by scrapes left by drifting icebergs

on the paleo-seafloor at z = 500 m, and the refined shape of the low velocity anomaly at z = 1000 m. These structures are absent from the initial tomography model in Figure 13. The layering horizontal structure of the low velocity anomaly is revealed by FWI from the smooth blob of the initial model in the traversing and adjacent slices presented in Figures 16(b,c) and 17(b,c). A chimney connecting the low velocity anomaly to the top part of the model also appears (Fig. 16(b) and 17(b)). Deeper, the base cretaceous reflector between 3.5 and 3.7 km depth is revealed.

A comparison between the v_P reconstruction results with the Gaussian (Fig. 16) and anisotropic diffusion (Fig.17) filters shows an enhancement of the resolution using the diffusion filter. This enhancement seems to vary with depth. In the shallow part, down to 1 km, the results remain very similar, even if some differences can be noted, especially a sharper delineation of the top of the low velocity anomaly using the diffusion filter. Between 1 km and the top of the structure below, the resolution enhancement is more visible. The low velocity anomaly layered structure is made finer and more coherent thanks to the diffusion filter. More details are also visible on the chimney displayed in the second inline section. Finally, the resolution enhancement is striking in the deeper part of the model. The reconstruction of the base cretaceous reflector between 3.5 and 3.7 km depth is much clearer using the diffusion filter compared with the Gaussian filter results.

[Figure 16 about here.]

[Figure 17 about here.]

The same trend is visible when comparing the density reconstruction results (Fig.18

and 19). Of note, both density reconstructions are quite stable, without the need to rely on specific data-weighting workflow or complicated preconditioning strategy. The density reconstruction appears to be less simple in the frequency-domain FWI study of Operto and Miniussi (2018) where dedicated frequency grouping strategies have to be implemented to stabilize its reconstruction. We might benefit here from the large redundancy of the time-domain medium sampling compared with the frequency-domain approach. The density values which are attained seem compatible with geological interpretation, with low density values at the core of the low velocity anomaly . The density models are also richer in high wavenumbers, giving information on the structure of the model and making more evident the horizontal and sub-horizontal layers traversing the low velocity anomaly, and those below. Comparing the results obtained with the anisotropic diffusion filter with those obtained with the Gaussian filter, the resolution increase obtained thanks to the diffusion filter is striking. All the aforementioned features appear much clearer in the result obtained with the anisotropic diffusion filter, this resolution improvement, as for v_P, increasing with depth.

[Figure 18 about here.]

[Figure 19 about here.]

As we have access simultaneously to P-wave velocity and density, we build impedance cubes $I(\mathbf{x})$ as

$$I(\mathbf{x}) = \rho(\mathbf{x}) v_P(\mathbf{x}). \tag{26}$$

From it we derive a reflectivity image by computing the impedance derivative across the main interfaces, following

$$\frac{\partial I}{\partial n} = \frac{\partial I}{\partial x}\sin\theta\cos\phi + \frac{\partial I}{\partial y}\sin\theta\sin\phi + \frac{\partial I}{\partial z}\cos\theta, \tag{27}$$

where $\theta(\mathbf{x})$ and $\phi(\mathbf{x})$ are dip and azimuth angles normal to the subsurface reflectors. To obtain $\theta(\mathbf{x})$ and $\phi(\mathbf{x})$, we proceed similarly as what we do to design the anisotropic diffusion tensor. We perform a SVD of the pre-smoothed structure tensor matrix associated with the impedance volume. This SVD provides us with the direction of main variation associated with the larger eigenvalues and the plane orthogonal to this direction of slower variations. From this orientation we can deduce the angles $\theta(\mathbf{x})$ and $\phi(\mathbf{x})$. This image building strategy has been popularized by Huang et al. (2021) under the name of full wavefield imaging. Its purpose is to derive interpretable images directly from high resolution FWI results.

We present in Figures 20 and 21 a comparison between the reflectivity images obtained using the Gaussian filter and the anisotropic diffusion filter respectively. The resolution increase noted already on the velocity and density reconstruction is even more striking. Thin layers within and around the low velocity anomaly are made apparent in the shallow part around 1 km depth which are not visible using the Gaussian filter. The main reflector delineating the interface between the low velocity anomaly and the structure below is made flatter and thinner. The deeper structure of the model, below 3 km and down to 4 km is revealed, with a clearly visible base cretaceous reflector between 3.5 and 3.7 km.

[Figure 20 about here.]

[Figure 21 about here.]

The overall resolution of the reflectivity images obtained using the diffusion filter appears higher. To further analyze this point, we compare in Figures 22 and 23 the wavenumber spectra of the reflectivity images vertical profiles presented in Figures 20(b,c) and 21(b,c). The spread of the wavenumber spectrum of the images derived from the FWI results using the diffusion filter is broader than their counterpart obtained using the Gaussian filter, which is indicative of a higher resolution.

[Figure 22 about here.]

[Figure 23 about here.]

We summarize this comparison with the 3D views of the velocity, density, and reflectivity volumes computed using the Gaussian (Fig.24) and anisotropic diffusion (Fig.25) filters. We have chosen to cut the cube in sections making clear the connection between what is identified as a chimney in the vertical section and the low velocity anomaly visible at 1 km depth in the constant depth section. In this 3D representation, the constant depth section at 3.5 km depth is also made visible to highlight the gain in resolution at this depth obtained using the anisotropic diffusion filter. This is particularly clear in the density and reflectivity image models.

[Figure 24 about here.]

[Figure 25 about here.]

Quality control

To assess the quality of the results we start by investigating the decrease of the misfit function over the course of iterations, for the three frequency bands 2.5 - 5 Hz, 2.5 - 7 Hz, 2.5 - 10 Hz (Fig. 26). The misfit function is normalized to 1 at the beginning of each band, and 48 iterations are performed on each band. As we rely on a source subsampling strategy, we can identify jumps in the misfit function, each 3 iterations, associated with the inversion of data associated with a new subgroup of sources. This pattern is especially visible in the first and last frequency band, while the convergence appears smoother in the second frequency band. Interestingly, we can observe that FWI using the anisotropic diffusion filter achieves a faster decrease of the misfit function on all of the three frequency bands. This is particularly pronounced on the 2.5 - 7 Hz band and at the beginning of the 2.5 - 10 Hz but is also visible in the 2.5 - 5 Hz band. This confirms what has been observed on the 2D Marmousi synthetic experiment: enhancing the coherent features in the gradient can accelerate the convergence of the FWI machinery.

[Figure 26 about here.]

We compare the fit to the data on the two CSG aligned with cable A and cable B presented in Figure 12. We adopt a mirror display, where the calculated data is presented surrounded by the observed data on its left and right sides. The calculated data is mirrored so as to highlight the fit at zero offset on the left and at far offset on the right. We perform this comparison for the data calculated in the initial model and the two final models obtained using the two filters. This comparison is presented in Figures 27 and 28. We can

observe that both FWI results achieve a remarkable fit at far offset, which indicates a good match of diving events and post-critical reflections. This shows that the long wavelength structures of the model have been correctly resolved. As mentioned in Operto et al. (2015), the presence in the initial model of the main reflector makes it possible to predict the reflections. No other reflected events are predicted by the initial model. In the two FWI models, the reflection on the shallow reflector on top of the low velocity anomaly and below are correctly matched. However, one can note a significant difference in the amplitude match of the reflections between the data predicted using the model obtained with the Gaussian filter and the data predicted using the model obtained using the anisotropic diffusion filter. In the latter, the amplitude of all the predicted reflected events is much stronger and closer to the energy of the reflected events in the field data.

[Figure 27 about here.]

[Figure 28 about here.]

This is confirmed in Figure 29, where we superimpose the predicted data in blue/red colorscale to the field data in black/white colorscale. A good fit is indicated by the absence of white and red in the final image. We see that the predicted data in both Gaussian filter and anisotropic diffusion filter FWI models is satisfactory for the post-critical reflection part, however for the pre-critical reflection, the fit in amplitude of the reflected event achieved by the anisotropic filter FWI model is better, especially for late events ($t > 4$ s).

[Figure 29 about here.]

To complement the quality control, we have three sonic logs, whose positions are displayed in Figure 11. The fit to the sonic logs is presented in Figure 30. We can observe that both FWI models nicely follow the trend of the logs, correcting from the initial guess, as achieved in previous studies. The differences between the two FWI models are subtle but exist. In the low velocity anomaly (well log 1), stronger variations can be observed in the anisotropic diffusion filter FWI model, which is closer to the log data. Stronger variations are also visible in the well log 2. Well log 3 is the only one providing information at depths, below 3 km. It appears that only the anisotropic diffusion filter FWI model presents variations in agreement with the log data at this depth and below. The Gaussian filter FWI model remains constant in this zone.

[Figure 30 about here.]

DISCUSSION

The results presented in the previous section bear different mark of interest. First, it appears that the use of the linear anisotropic diffusion filter significantly improves the resolution of the estimated models, especially in their deeper parts, compared with the use of a conventional Gaussian filter. In the multi-parameter settings we are exploring, this is especially true for the density model, which is much sharper in the estimation using the coherence enhancing diffusion filter. From this sharper density model, it is possible to build, together with the vertical velocity model, a high resolution impedance model, and derive from it a sharp reflectivity image volume which brings valuable structural information down to 4 km depth.

The reason why the coherence enhancing diffusion filter outperforms the Gaussian filter could be related to the crude estimation of the expected resolution which serves to determine the correlation length used in the Gaussian filter. First, as the velocity increases at depth, the correlation length of the Gaussian filter naturally increases with depth, which provides a tendency to oversmooth the deeper parts of the model. This is also seen in the Marmousi synthetic experiment. Second, more accurate expected resolution maps could be inferred by integrating information on the illumination angles. From diffraction analysis (Devaney, 1984; Wu and Toksöz, 1987; Sirgue and Pratt, 2004), it is well known that the expected resolution is driven by the local velocity and mean frequency, multiplied by a cosine of half the illumination angle, where the illumination angle is the angle formed by the rays connecting the source and the receiver to the imaged diffraction point. In other words, by neglecting the information on the illumination angle, the resolution power of FWI is underestimated. In addition, the same smoothing is applied to both the velocity and density models, while the latter is sensitive only to short angle reflection data. This leads to sever oversmoothing of the density model and can explain the loss of resolution observed by using the Gaussian filter.

On the other hand, the coherence enhancing diffusion filter only tries to enhance coherent features already present in the gradient, therefore avoiding conflict between expected resolution and actual features present in the gradient. This is to us the main interest for using this filter instead of the Gaussian filter.

The tuning parameters for the anisotropic diffusion filter appear to be relatively light. The parameters to tune are: the constants α and C in the threshold function g in equation 10,

the noise and coherence scale σ and ℓ, and the number of iterations to solve the anisotropic diffusion equation. Regarding α, a value of 10^{-2} ensure a total anisotropy ratio of a factor 100 which seems sufficient to capture most of the structure in the gradient. The value of 10^{-8} for C ensures a fast transition from weak to strong diffusion. For the noise and coherence scale σ and ℓ, it appears that selecting the same values for both parameters is efficient. This reduces the number of parameters to set. We have adapted them to the finite-difference grid used to solve the wave equation in our multi-scale approach, using a size of one or two discretization steps h depending on the direction. Given the fact that the model under investigation exhibits more regularity in the horizontal directions x and y we have taken σ and ℓ as respectively $2h, 2h$ and h in the x, y, z directions. This is a light injection of prior information on the model structure. Finally, regarding the number of diffusion iterations, we set it manually by trial-and-error, but an initial guess can be found benefiting from the analogy between Gaussian smoothing and isotropic diffusion (equations 3 and 4). Using this analogy, one can find a diffusion time T based on a given smoothing length L and devise the corresponding number of diffusion iterations based on the diffusion time-step, controlled by the CFL of the scheme of Fehrenbach and Mirebeau (2014). On top of that, the same filter is applied here to the velocity and density model without the need to design a specific filter for each parameter.

We summarize the discussion regarding computational aspect in the Table 2. The results reported here have been obtained on an Intel Skylake CPU machine with 32 cores at 2.1 GHz per node for the 2.5 - 5 Hz band (GRICAD/CIMENT local facility, Grenoble, France) and on Irene-ROME (TGCC, French national computing center), an AMD Epyc

CPU machine with 64 cores at 2.6 GHz per processors and 2 processors per node. Our code TOYxDAC_TIME implements a double level of parallelization. The outer one is at the shot level. The inner one is an OpenMP acceleration for the computation of the incident and adjoint wavefields for each source. We activate it only for the 2.5 - 7 Hz and 2.5 - 10 Hz frequency bands. In these bands, we use respectively 8 and 64 OpenMP threads per source. As we use batches of 128 shots with our subsampling strategy, FWI thus runs on 128 cores on the 2.5 - 5 Hz band, then 1024 cores on the 2.5 - 7 Hz band, and 8192 cores on the 2.5 - 10 Hz. We see in Table 2 that the elapsed time for the incident and adjoint field computation (plus the cost of recomputing the incident field backward in time, see Yang et al. (2016)) remains almost constant along the frequency bands thanks to the OpenMP acceleration. We also see that the computational time for the Gaussian filter is always negligible, less than 1% of the total time for the gradient computation. In counterpart, the computational time for the anisotropic diffusion filter, despite we use its linear version, is non negligible, and reaches almost 25 $\%$ of the computation time for one gradient in the last frequency band. This is due to the fact that the anisotropic diffusion filtering is performed on a single core and does not take advantage of the OpenMP acceleration. One shall keep in mind that the computational complexity of the diffusion process is roughly in $O(N^3)$ with N being the number of discrete points in one direction of space, while the complexity for solving the wave equation is in $O(N^4)$. This favors a better ratio between the time spent in filtering and the time spent in computing the wavefields at higher frequency. This is the reason why despite we lack the OpenMP parallelization for the anisotropic diffusion filter the increase in computational cost is maintained at a reasonable fraction of the global gradient computation time.

[Table 2 about here.]

One clear improvement to bring to the current implementation is to take benefit from the inner level of parallelization within our FWI algorithm. While a first option is to implement OpenMP directive in the solution of the anisotropic diffusion equation, a more interesting alternative we are currently investigating is to move to a domain decomposition algorithm within the TOYxDAC_TIME framework for the inner parallellization instead of OpenMP. This would make it possible to accelerate both the incident and adjoint field computation with small communications at each time steps, but also to solve the diffusion equation through this domain decomposition machinery. This is the strategy which is currently implemented in our (visco-)elastic full waveform modeling and inversion code SEM46 (Trinh et al., 2017; Cao et al., 2022).

This development would bring additional value to the current TOYxDAC_TIME package making it possible to target much larger acoustic FWI problems, as the scalability of the OpenMP directives is limited and depends on the number cores per node sharing the same memory. Moving to a domain decomposition paradigm could also alleviate the gradient computation by making it possible to store the incident wavefield on the subdomains and avoid recomputing it as we are currently doing. This recomputation strategy avoids memory bottleneck but is time consuming especially for viscous media such as the one we consider here (Yang et al., 2016). In the perspective of the field data application, such a domain decomposition algorithm would make it possible to invert for higher frequency bands up to 20 Hz or higher. We are interested in performing such investigation as the results presented here are promising, especially regarding the multi-parameter aspect. We could

expect that on such a wider frequency-band, the data becomes more sensitive to attenuation variations. We would be interested in assessing if we can perform a high resolution inversion for the vertical velocity, the density, and the quality factor simultaneously, using our time-domain algorithm, as the attempt presented in Kamath et al. (2021). This could bring interesting discussion also regarding the current trend toward very high resolution FWI, which seems to be done with acoustic mono-parameter modeling, and sometimes in the limit of numerical dispersion regarding the modeling schemes which are used.

On a side note regarding computing efficiency, we would like to mention that a porting of the coherence enhancing diffusion filter on GPU architecture has been performed. It is made available as a standalone open-source package[1]. This work differs from previous implementations of similar PDEs by the use of wide adaptive stencils, which allows to handle arbitrary anisotropy while preserving solution positivity, but could be expected to be less suited to GPUs due to data non-locality. Two points raise specific implementation difficulties. First, the computation of the eigenvalues of the structure tensors, for which a custom method was used, since the standard libraries were found to have excessively high memory usage and to be significantly slower. Second, the time step iterations, for which a custom kernel using atomic-addition operations is used, rather than the standard sparse matrix libraries, for the same reasons. Other routines could be used as is from the standard libraries, such as the convolutions involved in the structure tensor construction, or are embarrassingly parallel hence could be implemented in a straightforward manner, such as a key ingredient of the chosen numerical method known as Selling's matrix decomposition. Filtering the

[1] `https://github.com/Mirebeau/AdaptiveGridDiscretizations`

gradient for the field data application in the 2.5 - 10 Hz frequency band ($131 \times 259 \times 459$ ≈ 15 million points) was run on a laptop equipped with an NVIDIA RTX 4000 GPU. After a preprocessing step of 0.0046 s (construction of the diffusion tensors and of the scheme coefficients), a sequence of 20 time steps takes 0.179 s. This is a speed up by two orders of magnitude over the CPU implementation, which opens promising perspectives, despite its current limitation to a specific subtask of the seismic imaging pipeline.

Finally, the next leap forward regarding the inversion of such shallow environment OBC field data relies on interpreting the multi-component data, which is currently disregarded. This should be done taking into account a more accurate visco-elastic modeling of the wave propagation, with the potentiality to recover both P-wave and S-wave velocities. This is another line of investigation we are currently following (Cao et al., 2022).

CONCLUSION

We present in this study a novel filtering technique for the gradient smoothing step in FWI. This filtering technique is based on the solution of an anisotropic diffusion equation. The diffusion tensor is built automatically from the structure tensor associated with the gradient, using a coherence enhancing technique. Compared to other directional filtering methods, this diffusion based method inherits from the stability properties of a partial-differential-equations based smoothing technique: preservation of the minimum, maximum and mean values, independence to translation and rotations, scale-space property. Compared to previous work on diffusion based smoothing in FWI, what we propose here is a systematic 3D extension of the coherence-enhancing technique of Weickert (1998) which provides

a natural and efficient way to build the anisotropic diffusion operator. We illustrate the behavior of this filtering technique first on a 2D synthetic experiment on Marmousi. By comparison with a conventional isotropic Gaussian smoothing, we show the interest of a structure-oriented smoothing, which makes it possible to accelerate the convergence of the FWI scheme. We then study the interest for such a filtering technique in the framework of a 3D field data application on a North Sea OBC dataset. The results obtained using the coherence enhancing diffusion filter are compared with those obtained using the conventional nonstationary Gaussian technique. The resolution of the former is improved at depth, making more visible deep reflectors at the reservoir level. The 3D density volume is much better resolved, and a 3D reflectivity image volume deduced from the vertical velocity and density models clearly illustrate the improvement of the overall resolution of the estimated models. Standard quality controls are applied to show that the fit of the reflected energy is better achieved using the coherence enhancing diffusion filter, which we relate to the resolution improvement we observe.

ACKNOWLEDGMENT

This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKER BP, CGG, EXXONMOBIL, GEOLINKS, JGI, PETROBRAS, SHELL, SINOPEC and TOTALENERGIES. This study was granted access to the HPC resources provided by the GRICAD infrastructure (https: //gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, the HPC resources of Cray Marketing Partner Network (`https://partners.cray.com`), and those of IDRIS/TGCC under the

allocation 046091 made by GENCI. The authors would like to address special thanks to Aker BP and Espen Birger Raknes and Ross Milne for sharing the data and permission to present there results.

REFERENCES

Adriaens, X., L. Métivier, and C. Geuzaine, 2021, A trust-region newton method for frequency-domain full-waveform inversion: First International Meeting for Applied Geoscience & Energy Expanded Abstracts, 757–761.

Aghamiry, H., A. Gholami, and S. Operto, 2020, Compound regularization of Full-Waveform Inversion for imaging piecewise media: IEEE Transactions on Geoscience and Remote Sensing, **58**, 1192–1204; doi: 10.1109/TGRS.2019.2944464.

Anagaw, A. Y., and M. D. Sacchi, 2018, Edge-preserving smoothing for simultaneous-source full-waveform inversion model updates in high-contrast velocity models: Geophysics, **83**, A33–A37.

Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire, 2016, Global adjoint tomography: first-generation model: Geophysical Journal International, **207**, 1739–1766.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Cance, P., and Y. Capdeville, 2015, Validity of the acoustic approximation for elastic waves in heterogeneous media: Geophysics, **80**, T161–T173; doi: 10.1190/geo2014-0397.1.

Cao, J., R. Brossier, A. Górszczyk, L. Métivier, and J. Virieux, 2022, 3D multi-parameter full-waveform inversion for ocean-bottom seismic data using an efficient fluid-solid coupled spectral-element solver: Geophysical Journal International, **229**, 671–703.

Capdeville, Y., and J.-J. Marigo, 2013, A non-periodic two scale asymptotic method to take account of rough topographies for 2D elastic wave propagation: Geophysical Journal

International, **192**, 163–189.

Capdeville, Y., and L. Métivier, 2018, Elastic full waveform inversion based on the homogenization method: theoretical framework and 2-d numerical illustrations: Geophysical Journal International, **213**, 1093–1112; doi: 10.1093/gji/ggy039.

Cupillard, P., W. Mulder, P. Anquez, A. Mazuyer, and J. Barthélémy, 2020, The apparent anisotropy of the SEG-EAGE overthrust model: Presented at the 82^{th} Annual EAGE Meeting (Amsterdam), European Association of Geoscientists & Engineers.

Devaney, A., 1984, Geophysical diffraction tomography: Geoscience and Remote Sensing, IEEE Transactions on, **GE-22**, 3–13.

Fehrenbach, J., and J.-M. Mirebeau, 2014, Sparse Non-negative Stencils for Anisotropic Diffusion: Journal of Mathematical Imaging and Vision, **49**, 123–147; doi: 10.1007/s10851-013-0446-3.

Fichtner, A., and A. Villaseñor, 2015, Crust and upper mantle of the western mediterranean – constraints from full-waveform inversion: Earth and Planetary Science Letters, **428**, 52 – 62; doi: https://doi.org/10.1016/j.epsl.2015.07.038.

Gardner, G. H. F., L. W. Gardner, and A. R. Gregory, 1974, Formation velocity and density–the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770–780.

Gholami, Y., 2012, Two-dimensional seismic imaging of anisotropic media by full waveform inversion: PhD thesis, Université Sophia-Antipolis.

Górszczyk, A., S. Operto, and M. Malinowski, 2017, Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited: Journal of Geophysical Research: Solid Earth, **122**, 4601–4630; doi: 10.1002/2016jb013891.

Guasch, L., O. C. Agudo, M.-X. Tang, P. Nachev, and M. Warner, 2020, Full-waveform

941 inversion imaging of the human brain: NPJ digital medicine, **3**, 1–12.
942 Guitton, A., G. Ayeni, and E. Díaz, 2012, Constrained full-waveform inversion by model
943 reparameterization: Geophysics, **77**, R117–R127.
944 Huang, R., Z. Zhang, Z. Wu, Z. Wei, J. Mei, and P. Wang, 2021, Full-waveform inversion
945 for full-wavefield imaging: Decades in the making: The Leading Edge, **40**, 324–334.
946 Igel, H., 2017, Computational seismology: a practical introduction: Oxford University
947 Press.
948 Irnaka, T. M., R. Brossier, L. Metivier, T. Bohlen, and Y. Pan, 2022, 3D Multi-component
949 Full Waveform Inversion for Shallow-Seismic Target: Ettlingen Line Case Study: Geo-
950 physical Journal International, **229**, 1017–1040.
951 Kalinicheva, T., M. Warner, and F. Mancini, 2020, Full-bandwidth fwi: SEG Technical
952 Program Expanded Abstracts 2020, 651–655.
953 Kamath, N., R. Brossier, L. Métivier, A. Pladys, and P. Yang, 2021, Multiparameter full-
954 waveform inversion of 3D ocean-bottom cable data from the Valhall field: Geophysics,
955 **86**, B15–B35; doi: 10.1190/geo2019-0705.1.
956 Karaoğlu, H., and B. Romanowicz, 2018, Inferring global upper-mantle shear attenua-
957 tion structure by waveform tomography using the spectral element method: Geophysical
958 Journal International, **213**, 1536–1558; doi: 10.1093/gji/ggy030.
959 Koehn, D., D. Wilken, T. Wunderlich, D. De Nil, W. Rabbel, L. Werther, J. Schmidt, C.
960 Zielhofer, and S. Linzen, 2018, Seismic SH full waveform inversion as new prospec-
961 tion method in archaeogeophysics: Conference Proceedings, 24th European Meeting of
962 Environmental and Engineering Geophysics, European Association of Geoscientists &
963 Engineers, 1–5.

964 Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations:
965 Conference on Inverse Scattering, Theory and application, Society for Industrial and
966 Applied Mathematics, Philadelphia, Conference on Inverse Scattering, Theory and ap-
967 plication, Society for Industrial and Applied Mathematics, Philadelphia, 206–220.
968 Lee, D., and S. Pyun, 2018, Adaptive preconditioning of full-waveform inversion based on
969 structure-oriented smoothing filter: SEG Technical Program Expanded Abstracts 2018,
970 1048–1052.
971 Lei, W., Y. Ruan, E. Bozdağ, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill,
972 N. Podhorszki, and D. Pugmire, 2020, Global adjoint tomography—model glad-m25:
973 Geophysical Journal International, **223**, 1–21; doi: 10.1093/gji/ggaa253.
974 Lewis, W., D. Amazonas, D. Vigh, and R. Coates, 2014, Geologically constrained full-
975 waveform inversion unsing an anisotropic diffusion based regularization scheme: appli-
976 cation to a 3d offshore brazil dataset: Presented at the SEG Expanded Abstracts.
977 Lu, Y., L. Stehly, R. Brossier, A. Paul, and A. W. Group, 2020, Imaging Alpine crust using
978 ambient noise wave-equation tomography: Geophysical Journal International, **222**, 69–
979 85; doi: 10.1093/gji/ggaa145.
980 Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for
981 Marmousi: The Leading Edge, **25**, 156–166; doi: 10.1190/1.2172306.
982 Marty, P., C. Boehm, and A. Fichtner, 2021, Acoustoelastic full-waveform inversion for
983 transcranial ultrasound computed tomography: Medical Imaging 2021: Ultrasonic Imag-
984 ing and Tomography, International Society for Optics and Photonics, 1160211.
985 Métivier, L., and R. Brossier, 2016, The SEISCOPE optimization toolbox: A large-scale
986 nonlinear optimization library based on reverse communication: Geophysics, **81**, F11–

F25.

———, 2022, On the use of nonlinear anisotropic diffusion filters for seismic imaging using the full waveform: Inverse Problems, **38**, 115001; doi: 10.1088/1361-6420/ac8c91.

Nocedal, J., 1980, Updating Quasi-Newton Matrices With Limited Storage: Mathematics of Computation, **35**, 773–782.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea: Geophysical Journal International, **213**, 2037–2059; doi: 10.1093/gji/ggy103.

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

Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough: Journal of Geophysical Research, **111**, doi:10.1029/2005JB003835.

Peters, B., and F. J. Herrmann, 2017, Constraints versus penalties for edge-preserving full-waveform inversion: The Leading Edge, **36**, 94–100; doi: 10.1190/tle36010094.1.

Pladys, A., R. Brossier, N. Kamath, and L. Métivier, 2022, Robust FWI with graph space optimal transport: application to 3D OBC Valhall data: Geophysics, **87**, 1–76.

Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom

seismometer dataset: First Break, **28**, 71–78.

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Provenzano, G., R. Brossier, and L. Métivier, 2022, Robust and efficient waveform-based velocity-model-building by optimal-transport in the pseudotime domain: methodology: Geophysics, **in press**.

Raknes, E. B., B. Arntsen, and W. Weibull, 2015, Three-dimensional elastic full waveform inversion using seismic data from the sleipner area: Geophysical Journal International, **202**, 1877–1894.

Shen, X., L. Jiang, J. Dellinger, A. Brenders, C. Kumar, M. James, J. Etgen, D. Meaux, R. Walters, and N. Abdullayev, 2018, High-resolution full-waveform inversion for structural imaging in exploration: SEG Technical Program Expanded Abstracts 2018, 1098–1102.

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

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

Solano, C. P., and R.-É. Plessix, 2019, Velocity-model building with enhanced shallow resolution using elastic waveform inversion — an example from onshore oman: Geophysics, **84**, R977–R988; doi: 10.1190/geo2018-0736.1.

Stopin, A., R.-E. Plessix, and S. Al Abri, 2014, Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman: Geophysics, **79**, WA69–

WA77.

Strong, D., and T. Chan, 2003, Edge-preserving and scale-dependent properties of total variation regularization: Inverse Problems, **19**, S165–S187.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 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**, 1259–1266.

Thomsen, L. A., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

Thrastarson, S., D.-P. van Herwaarden, L. Krischer, C. Boehm, M. van Driel, M. Afanasiev, and A. Fichtner, 2022, Data-adaptive global full-waveform inversion: Geophysical Journal International, **230**, 1374–1393; doi: 10.1093/gji/ggac122.

Thurin, J., 2020, Uncertainties estimation in full waveform inversion using ensemble methods: Theses, Université Grenoble Alpes.

Tikhonov, A., A. Goncharsky, V. Stepanov, and A. Yagola, 2013, Numerical methods for the solution of ill-posed problems: Springer Science & Business Media.

Trinh, P. T., R. Brossier, L. Métivier, J. Virieux, and P. Wellington, 2017, Bessel smoothing filter for spectral element mesh: Geophysical Journal International, **209**, 1489–1512; doi: 10.1093/gji/ggx103.

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, R63–R77.

Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An introduction to Full Waveform Inversion, *in* Encyclopedia of Exploration Geophysics:

Society of Exploration Geophysics, R1–1–R1–40.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, P., Z. Zhang, J. Mei, F. Lin, and R. Huang, 2019, Full-waveform inversion for salt: A coming of age: The Leading Edge, **38**, 204–213.

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

Weickert, J., 1998, Anisotropic diffusion in image processing, treubner verlag, stuttgart: Treubner Verlag.

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

Xu, S., D. Wang, F. Chen, G. Lambaré, and Y. Zhang, 2012, Inversion on reflected seismic wave: SEG Technical Program Expanded Abstracts 2012, 1–7; doi: 10.1190/segam2012-1473.1.

Xue, Z., Z. Zhang, F. Lin, J. Mei, R. Huang, and P. Wang, 2020, Full-waveform inversion for sparse obn data: SEG Technical Program Expanded Abstracts 2020, 686–690.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016, Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method: Geophysics, **81**, R349–R362; doi: 10.1190/geo2016-0082.1.

Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018, A Time-Domain Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion: SIAM Journal on Scientific Computing, **40**, B1101–B1130; doi:

10.1137/17M1126126.

Yuan, H., S. French, P. Cupillard, and B. Romanowicz, 2014, Lithospheric expression of geological units in central and eastern north america from full waveform tomography: Earth and Planetary Science Letters, **402**, 176–186; doi: https://doi.org/10.1016/j.epsl.2013.11.057. (Special issue on USArray science).

Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, 1535–1554.

Zuberi, M. A., and R. G. Pratt, 2017, Mitigating nonlinearity in full waveform inversion using scaled-sobolev pre-conditioning: Geophysical Journal International, **213**, 706–725; doi: 10.1093/gji/ggx549.

LIST OF FIGURES

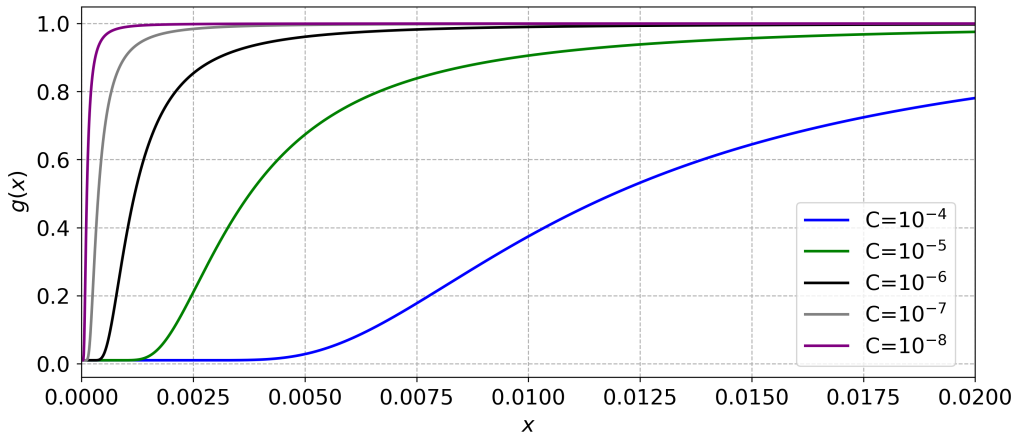

Figure 1: Threshold function $g(x)$ depending on the constant C for $\alpha = 10^{-2}$.

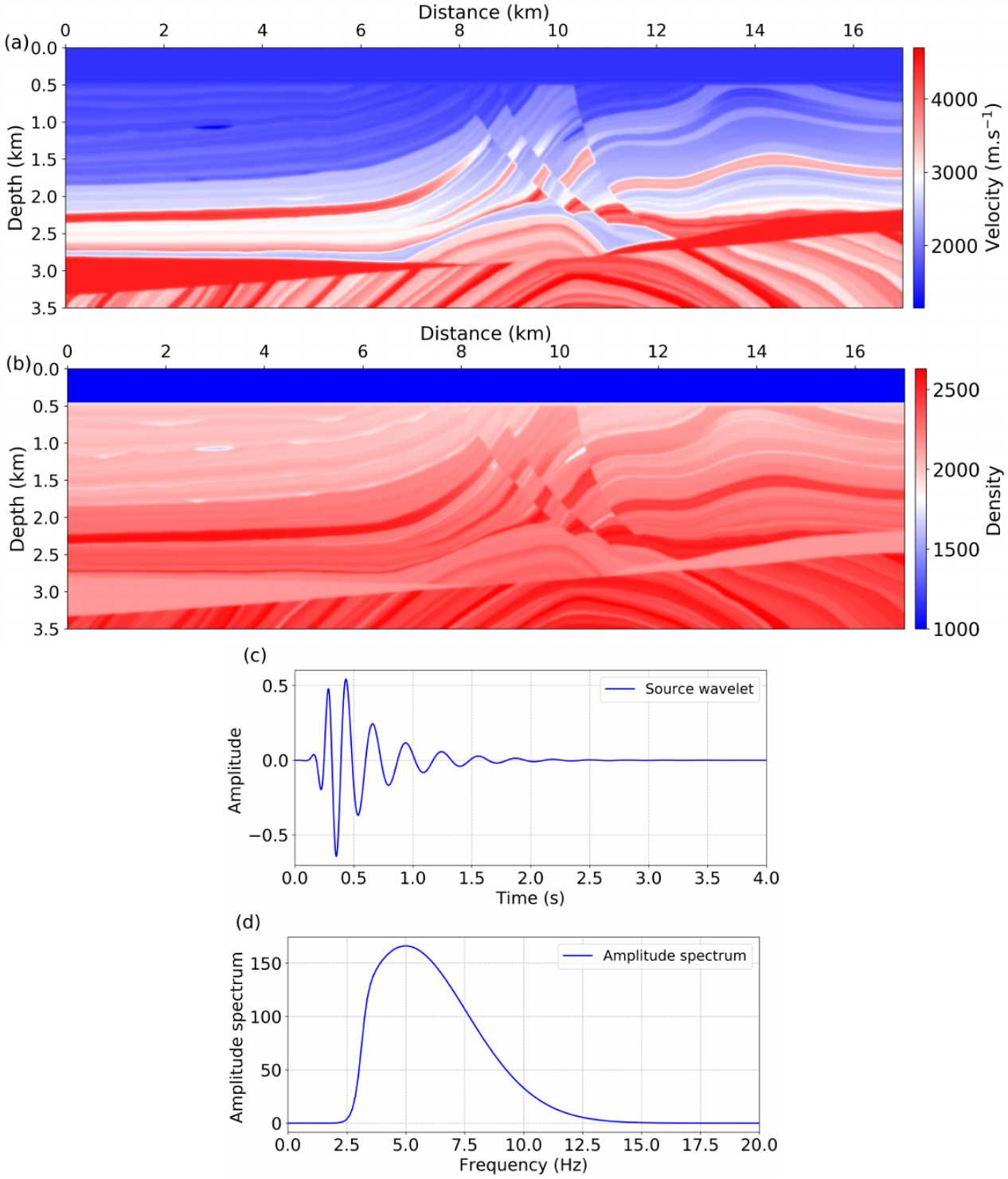

Figure 2: (a) Velocity and (b) density 5 m grid models used as reference to build the data for the Marmousi II synthetic experiment. (c) Source wavelet used to generate the data using these models and (d) its spectrum.

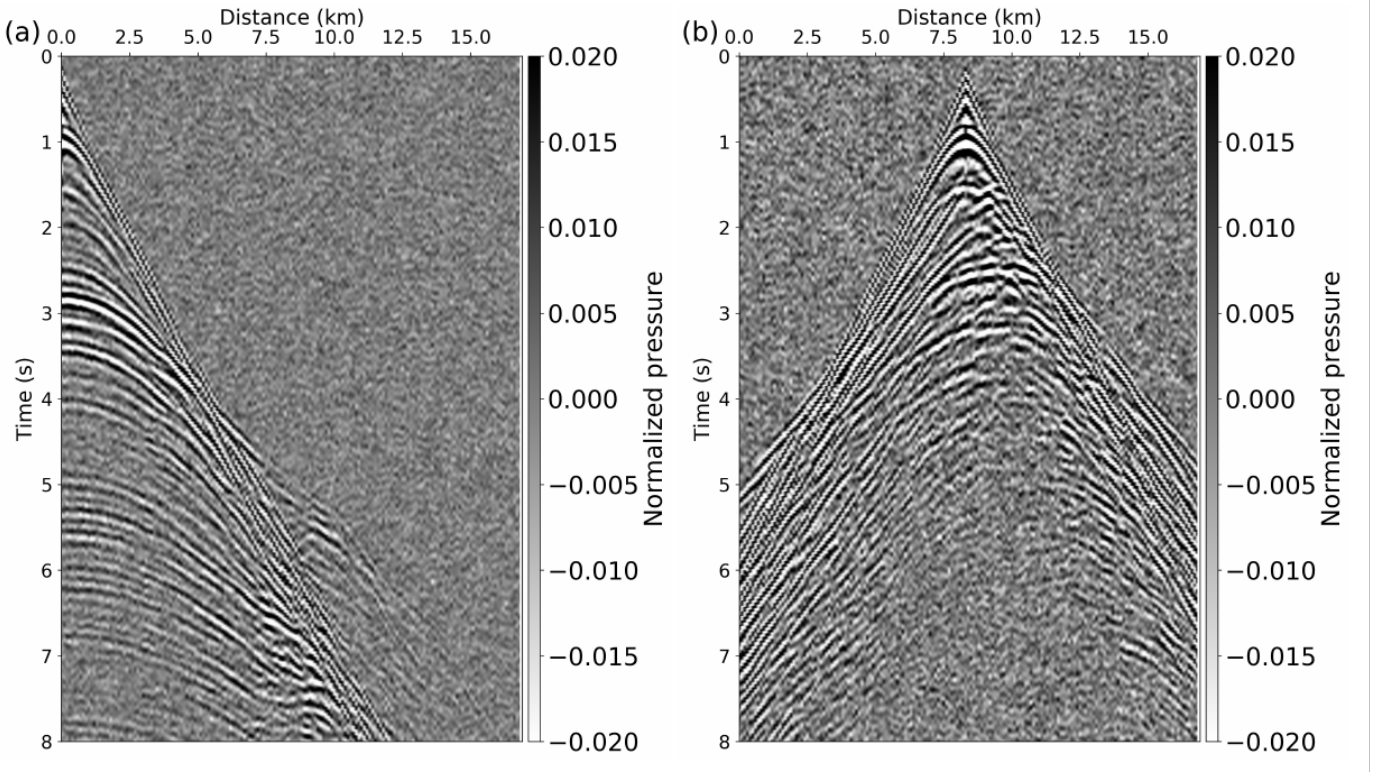

Figure 3: (a) Left most and (b) central shot gather computed using the models and the source wavelet presented in Figure 2.

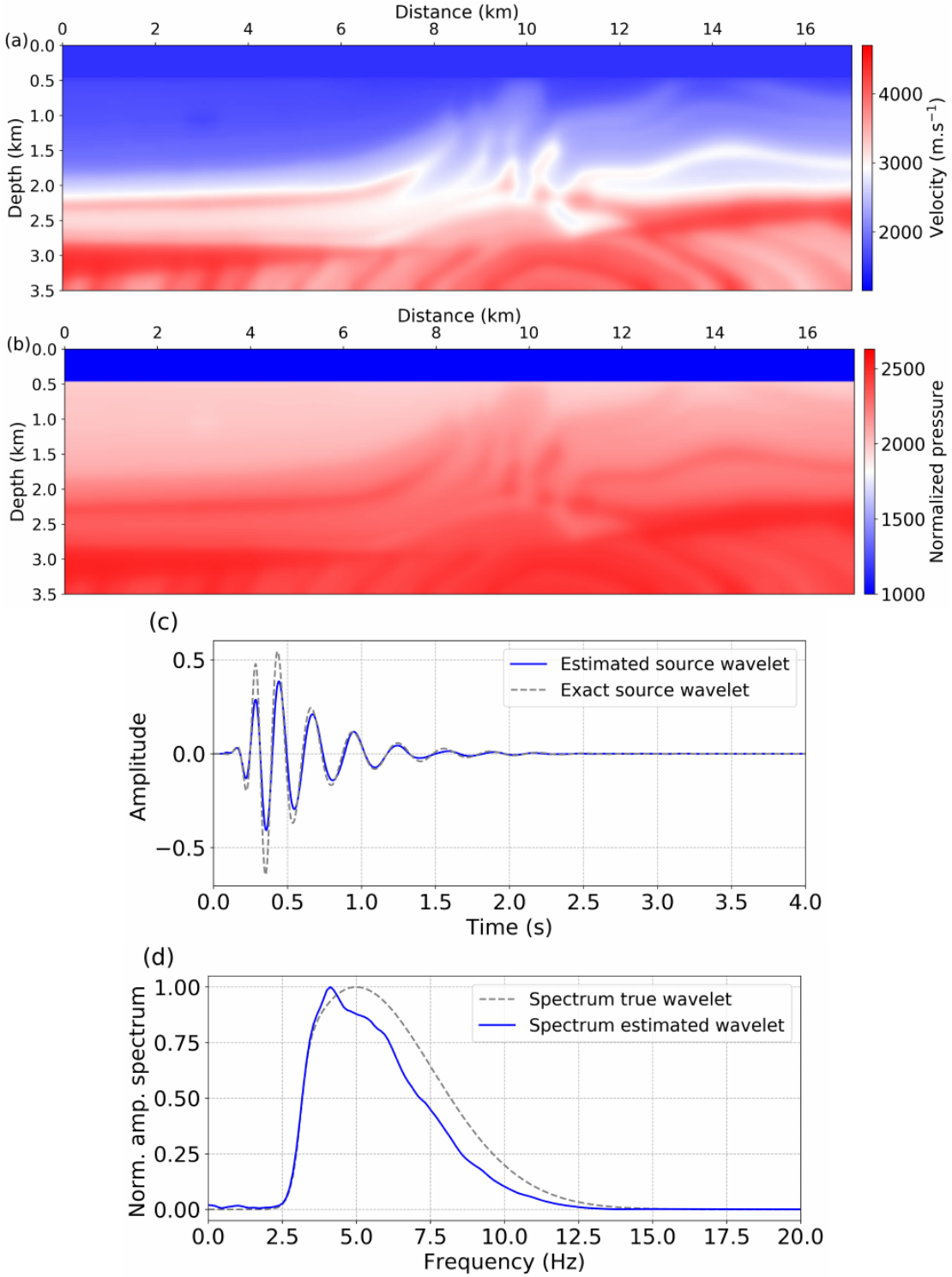

Figure 4: (a) Initial velocity and (b) initial density. (c) Estimated source wavelet compared to the true source wavelet. (d) Comparison of the corresponding two amplitude spectra. Due to the inaccuracy of the initial velocity and density models and the noise on the data, the estimated wavelet exhibits a lower amplitude shift than the true wavelet as well as a less smooth amplitude spectrum.

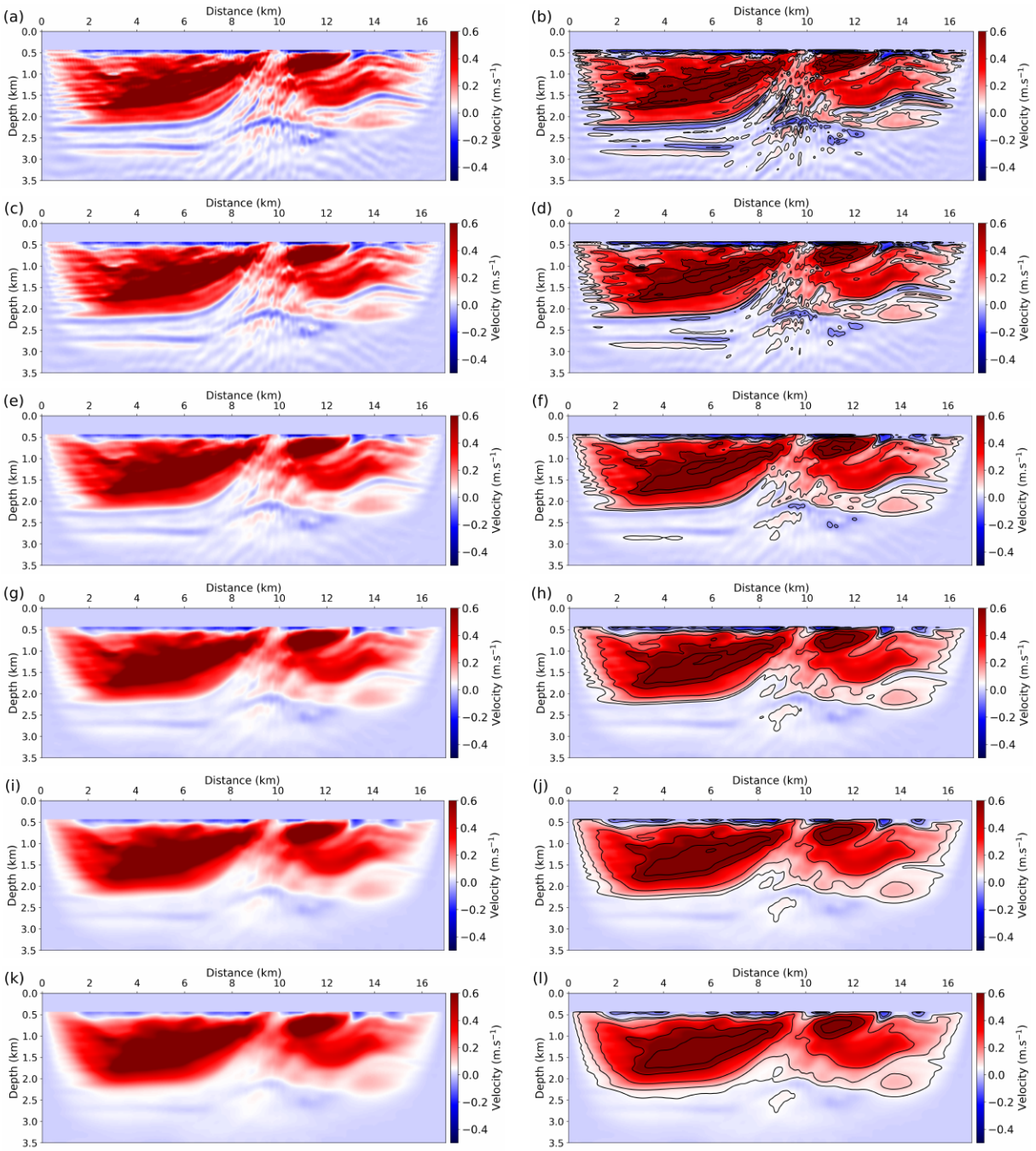

Figure 5: Evolution of the gradient for different Gaussian smoothing with $r_x = r_z$ from 0.1 to 0.5. On the left column the gradient alone is presented, while on the right column, the level set of the gradient are superimposed to better delineate its main structures. Gradient with (a,b) no smoothing, (c,d) $r_x = r_z = 0.1$, (e,f) $r_x = r_z = 0.2$, (g,h) $r_x = r_z = 0.3$, (i,k) $r_x = r_z = 0.4$, (k,l) $r_x = r_z = 0.5$.

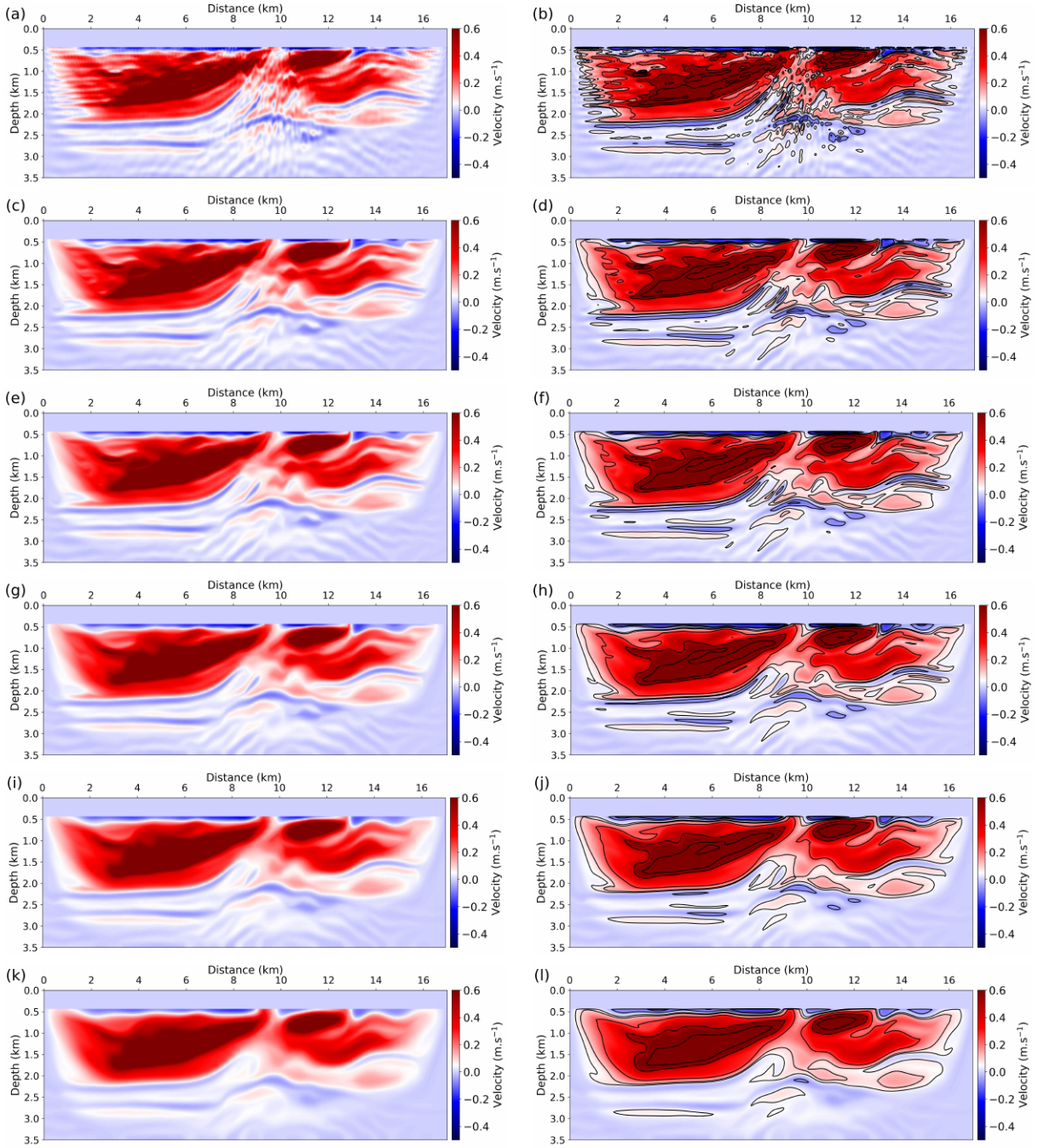

Figure 6: Evolution of the gradient along the nonlinear anisotropic diffusion iterations using the stencil from Fehrenbach and Mirebeau (2014). On the left column the gradient alone is presented, while on the right column, the level set of the gradient are superimposed to better delineate its main structures. Gradient with (a,b) no smoothing, after (c,d) 50 iterations, (e,f) 100 iterations, (g,h) 200 iterations (i,j) 400 iterations (k,l) 800 iterations.

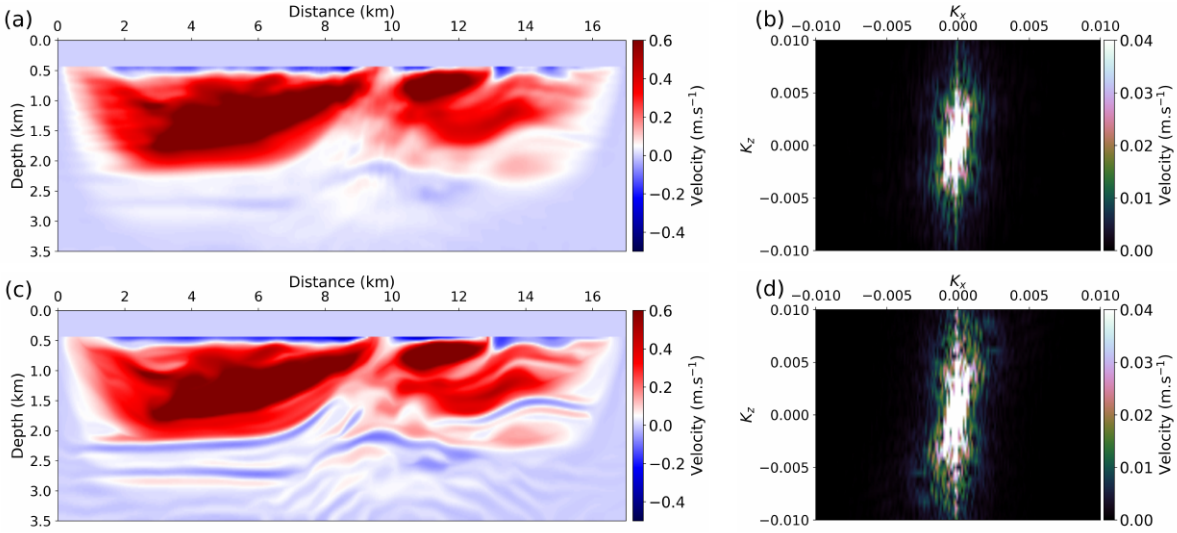

Figure 7: Comparison between the initial gradient obtained after (a,b) a Gaussian smoothing (c,d) 200 iterations of nonlinear anisotropic diffusion using the stencil from Fehrenbach and Mirebeau (2014). In (a) and (c) the gradients are presented, while in (b) and (d) their corresponding k_x, k_z spectrum are plotted. The gradient (a) corresponds to the one in Figure 5i while the gradient (c) corresponds to the one in Figure 6g.

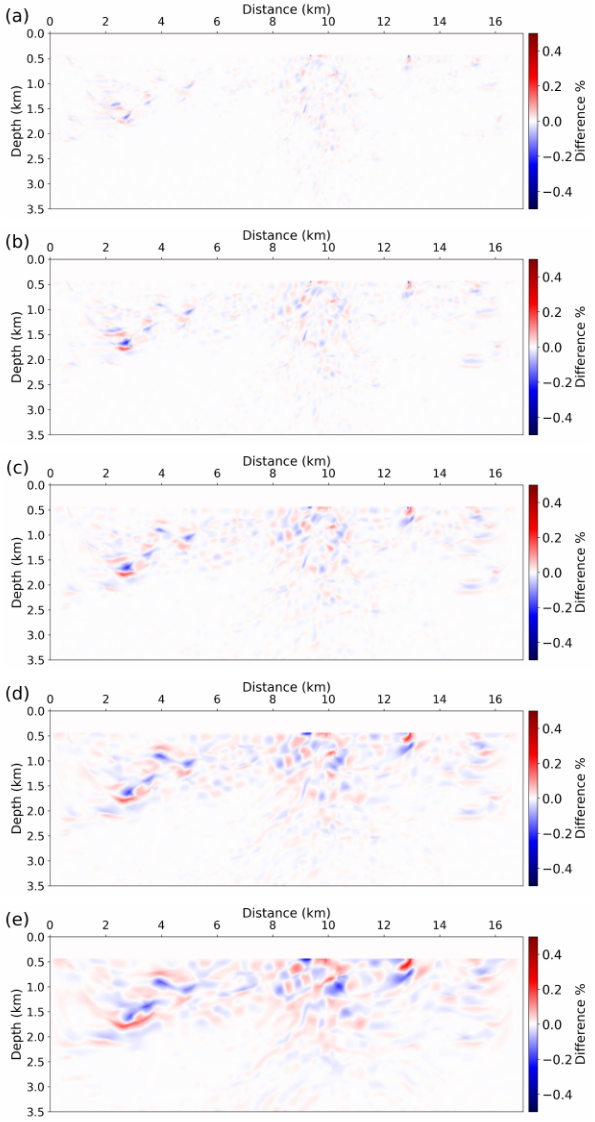

Figure 8: Difference in percentage (following equation 21) between a gradient obtained after *nonlinear* anisotropic diffusion and a gradient after *linear* anisotropic diffusion. Comparison for (a) 50 diffusion iterations, (b) 100 diffusion iterations, (c) 200 diffusion iterations, (d) 400 diffusion iterations, (e) 800 diffusion iterations. The maximum error in percentage reaches 0.5 $\%$.

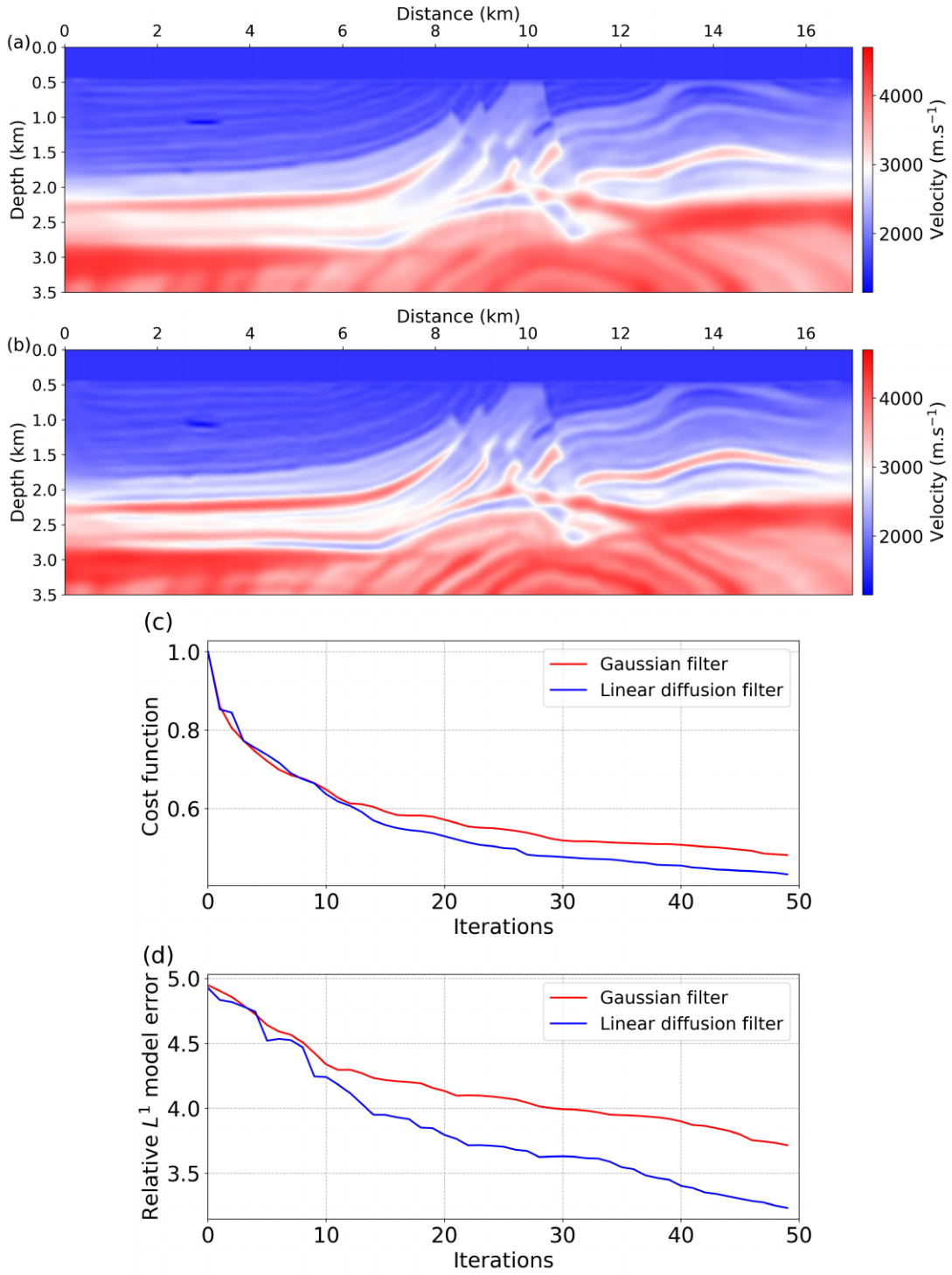

Figure 9: Final P-wave velocity model after 50 FWI iterations using (a) Gaussian smoothing (b) linear anisotropic diffusion smoothing. (c) Model error depending on the iteration number. The use of the linear anisotropic diffusion filter accelerates the convergence towards the exact model.

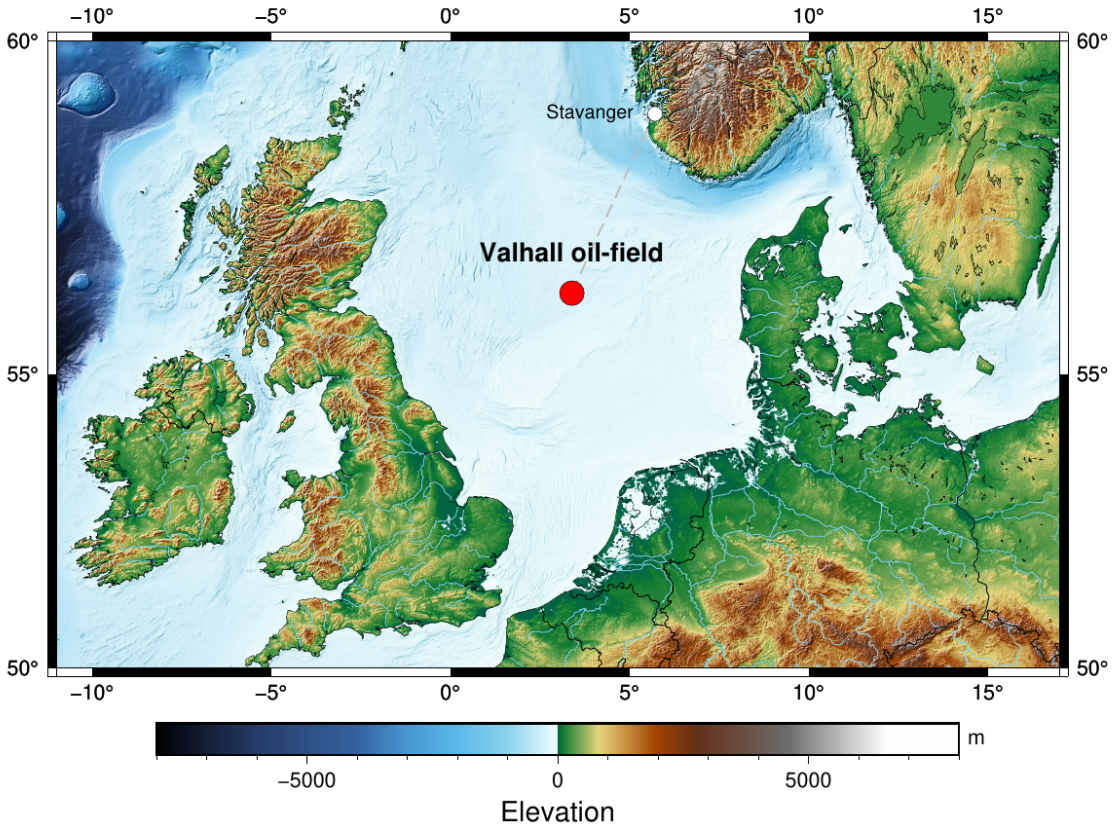

Figure 10: Location of the Valhall field on the North Sea (from Thurin (2020)).

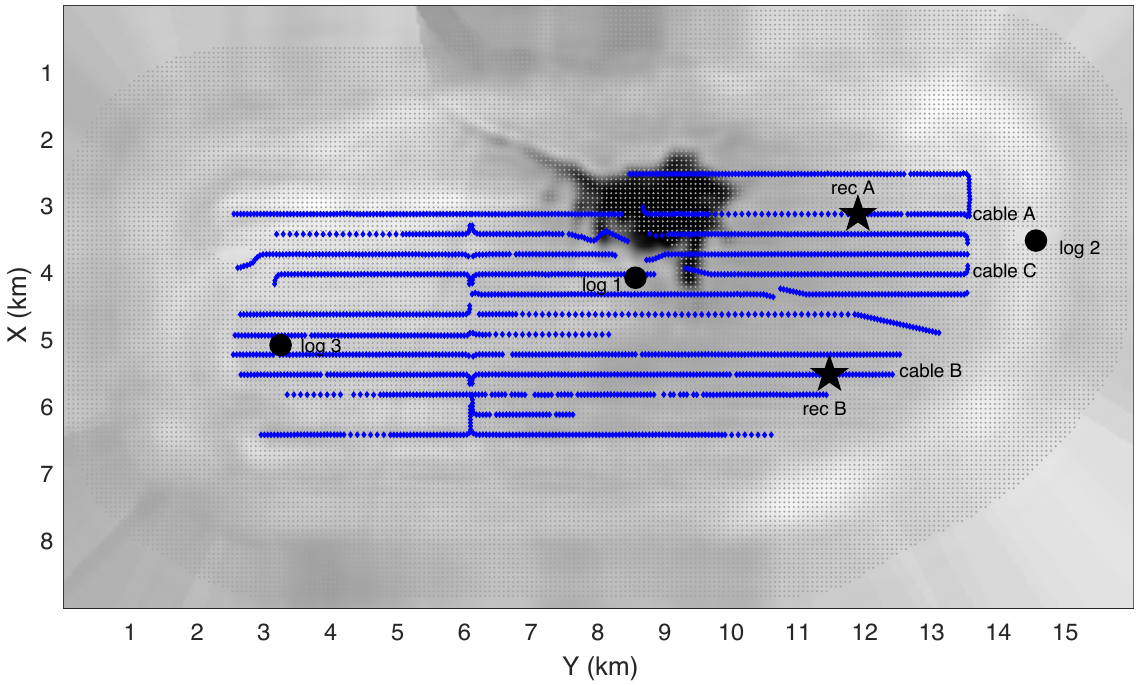

Figure 11: Layout of the acquisition overlapped on an horizontal P-wave velocity slice at 1 km depth obtained by FWI. Location of sources (gray dots) and receivers (blue diamonds). Two receivers positions (A and B) are located with black stars. Cables A ($x = 2950$ m), B ($x = 5530$ m) and C ($x = 3080$ m) are identified. Black dots denote the position of three available P-wave velocity sonic-logs.

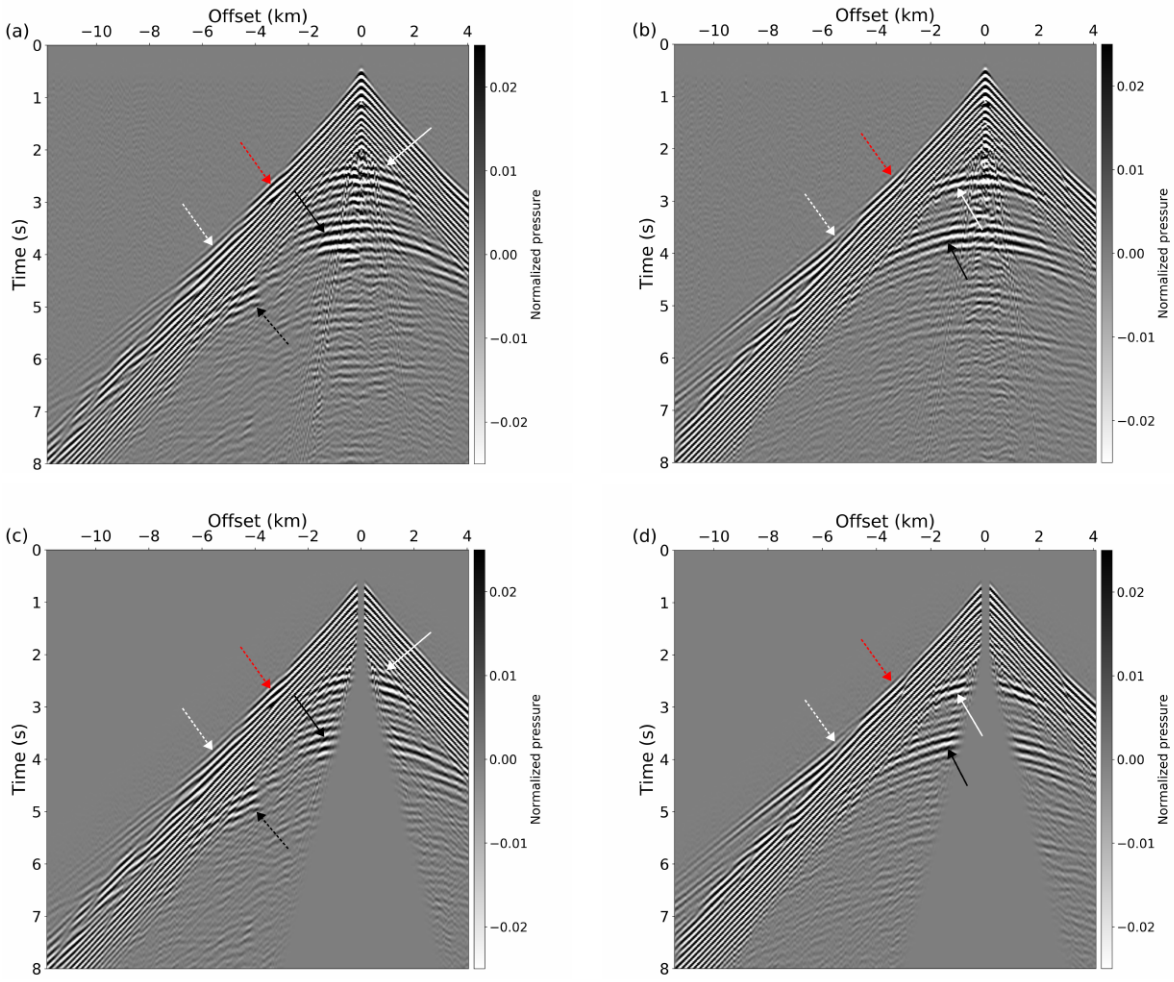

Figure 12: (a) CSG aligned with the positions of cable A before muting the Schölte. (b) CSG aligned with the positions of cable B before muting the Schölte waves. (c) CSG aligned with the positions of cable A after muting the Schölte waves. (d) CSG aligned with the positions of cable B after muting the Schölte waves. We reproduce the data anatomy analysis presented in Operto et al. (2015). The red, white, black arrows point on the reflection from a shallow reflector, the top of the low velocity anomaly and the top of the reservoir, respectively. The solid arrow points on the pre-critical reflections, while the dashed ones points on the post-critical reflections.

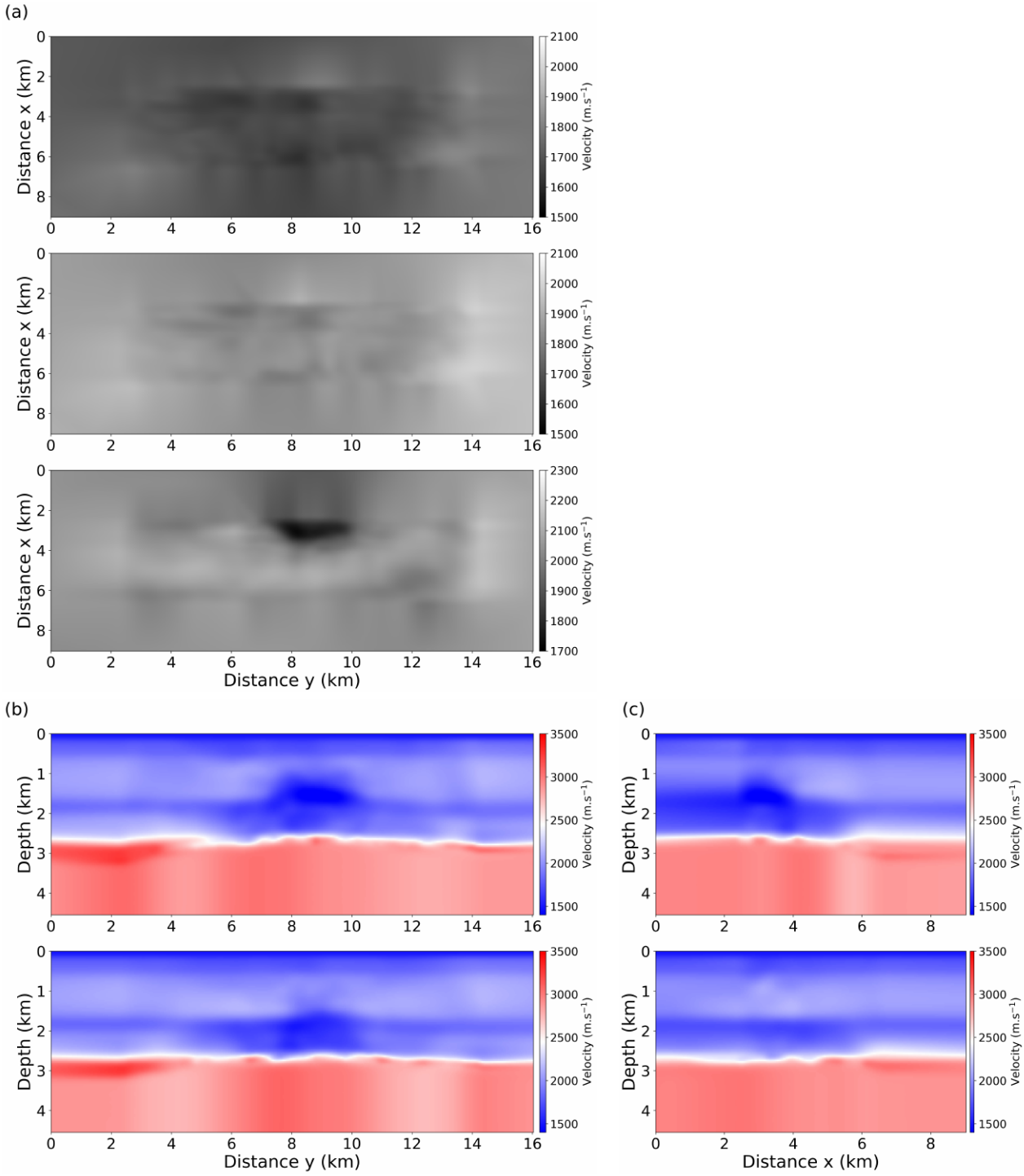

Figure 13: Initial v_P model. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

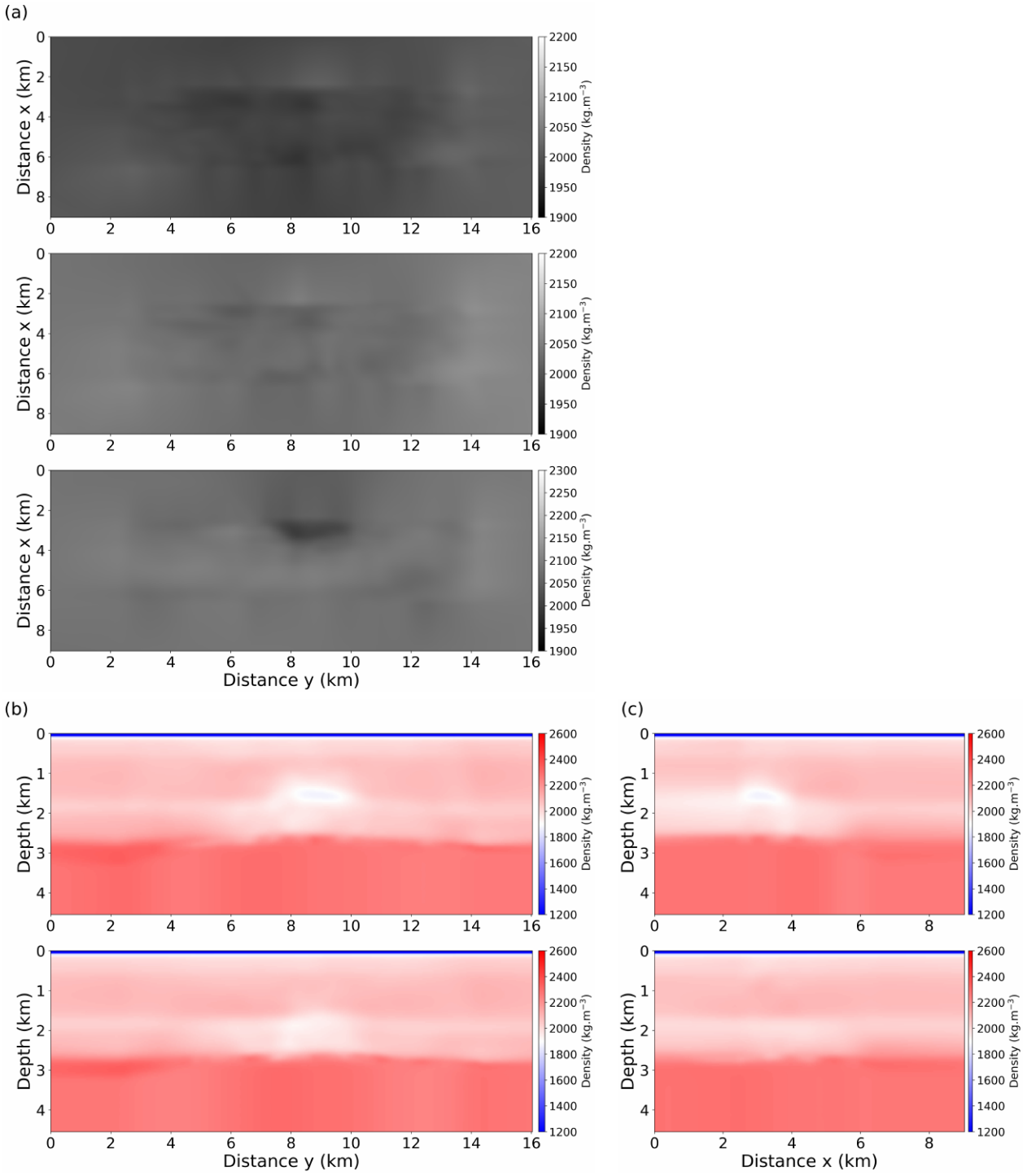

Figure 14: Initial density model. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

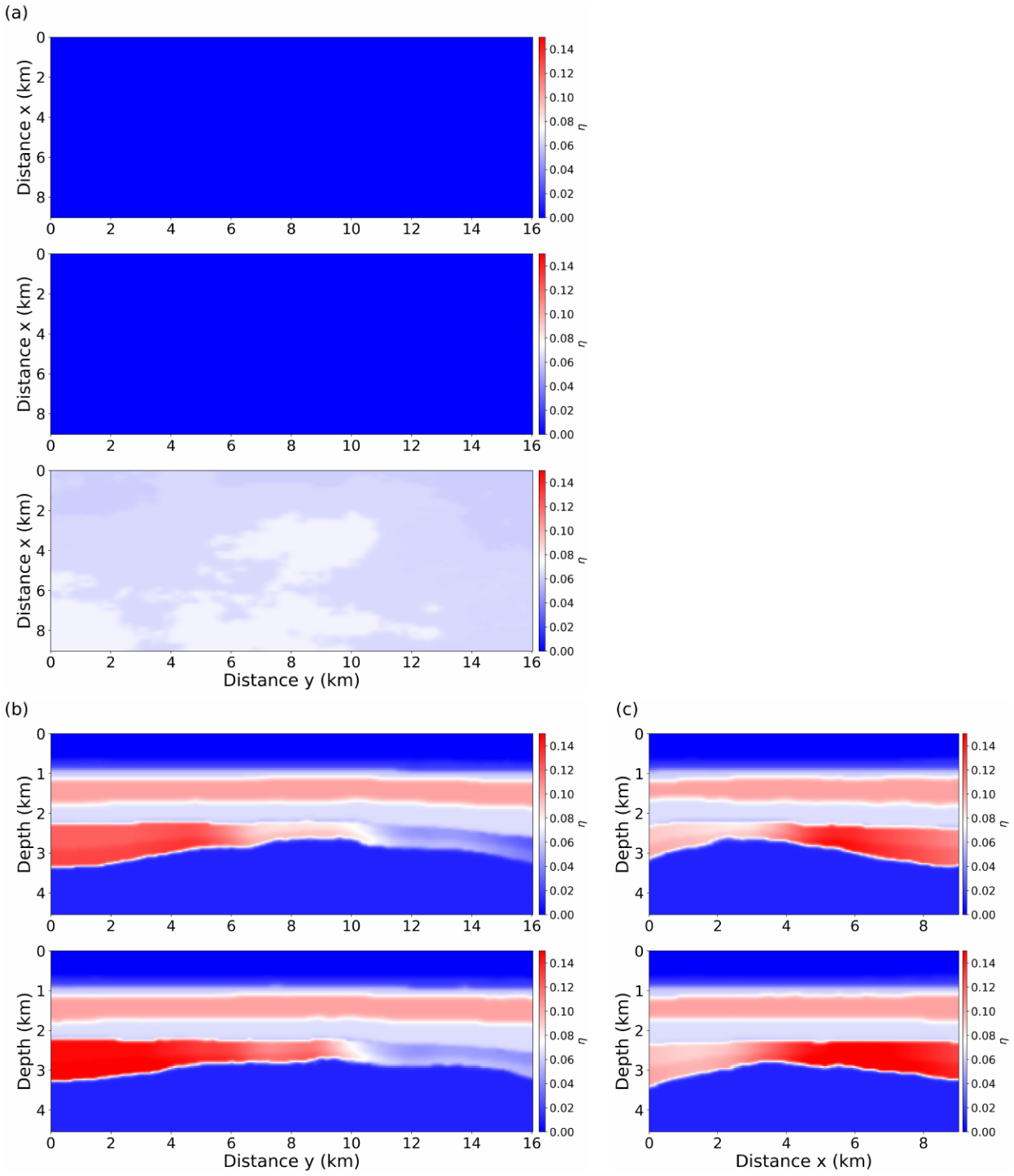

Figure 15: Initial anisotropy model η. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

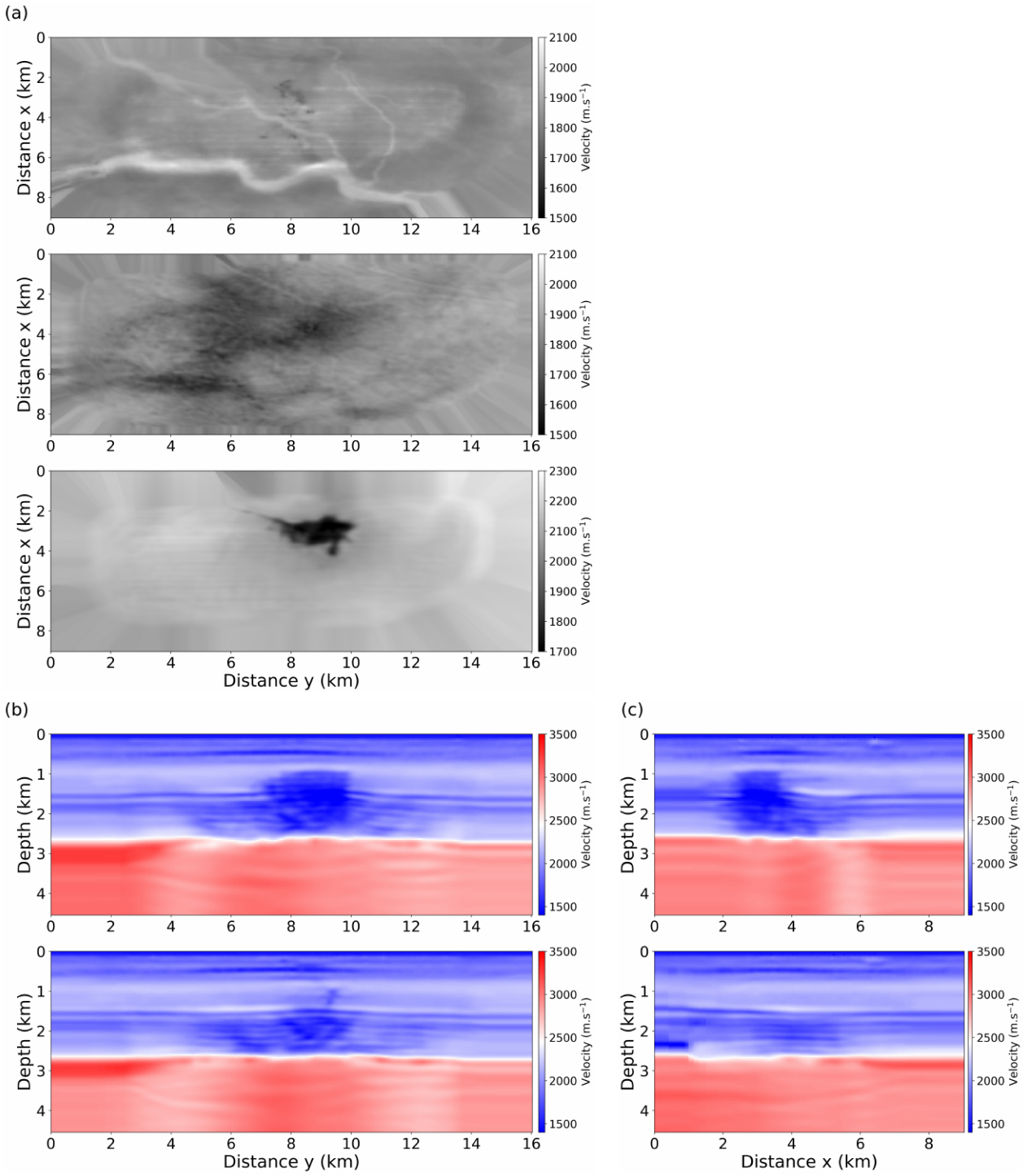

Figure 16: Final v_P model with non-stationary Gaussian smoothing. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

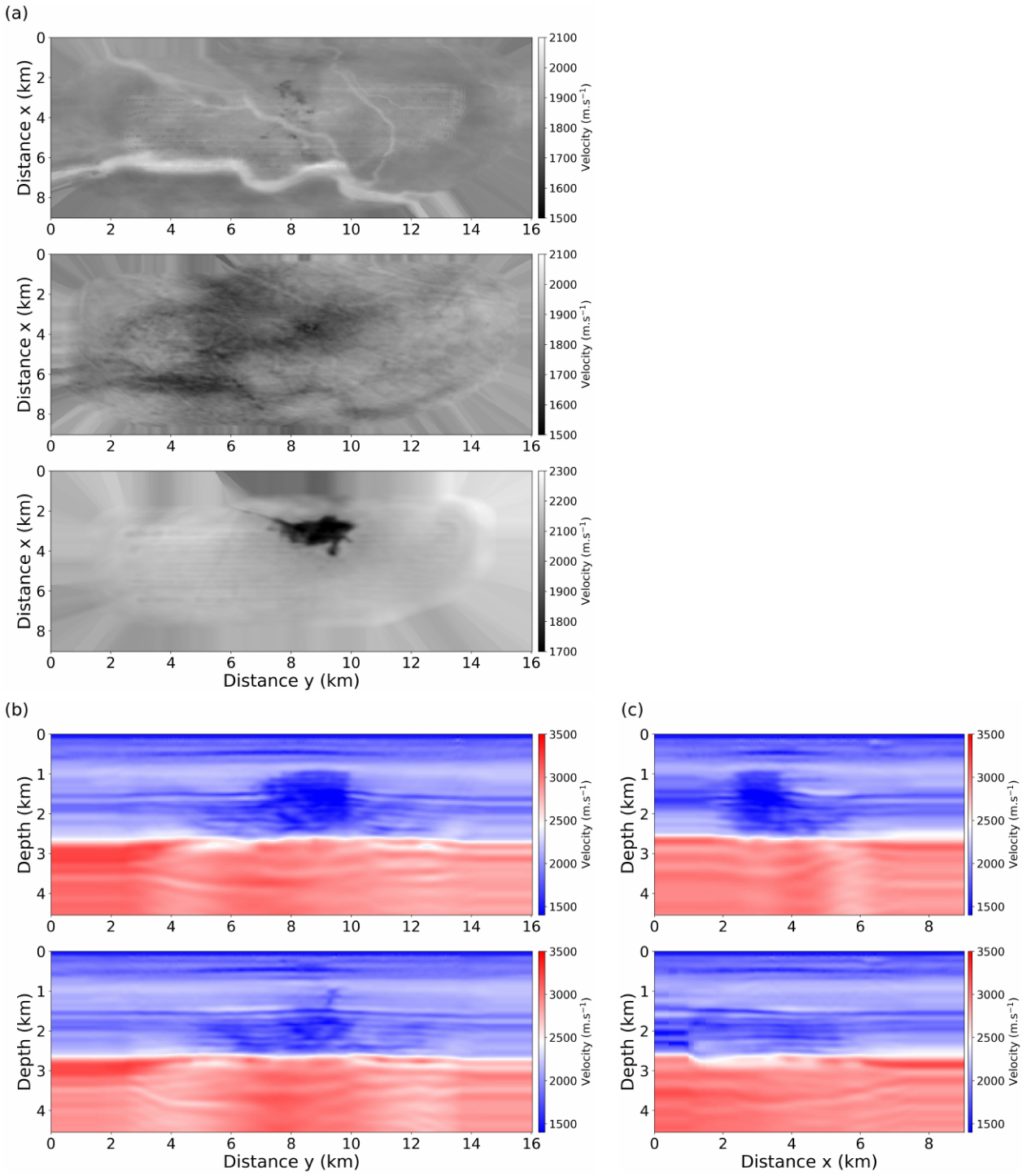

Figure 17: Final v_P model with linear anisotropic diffusion smoothing. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

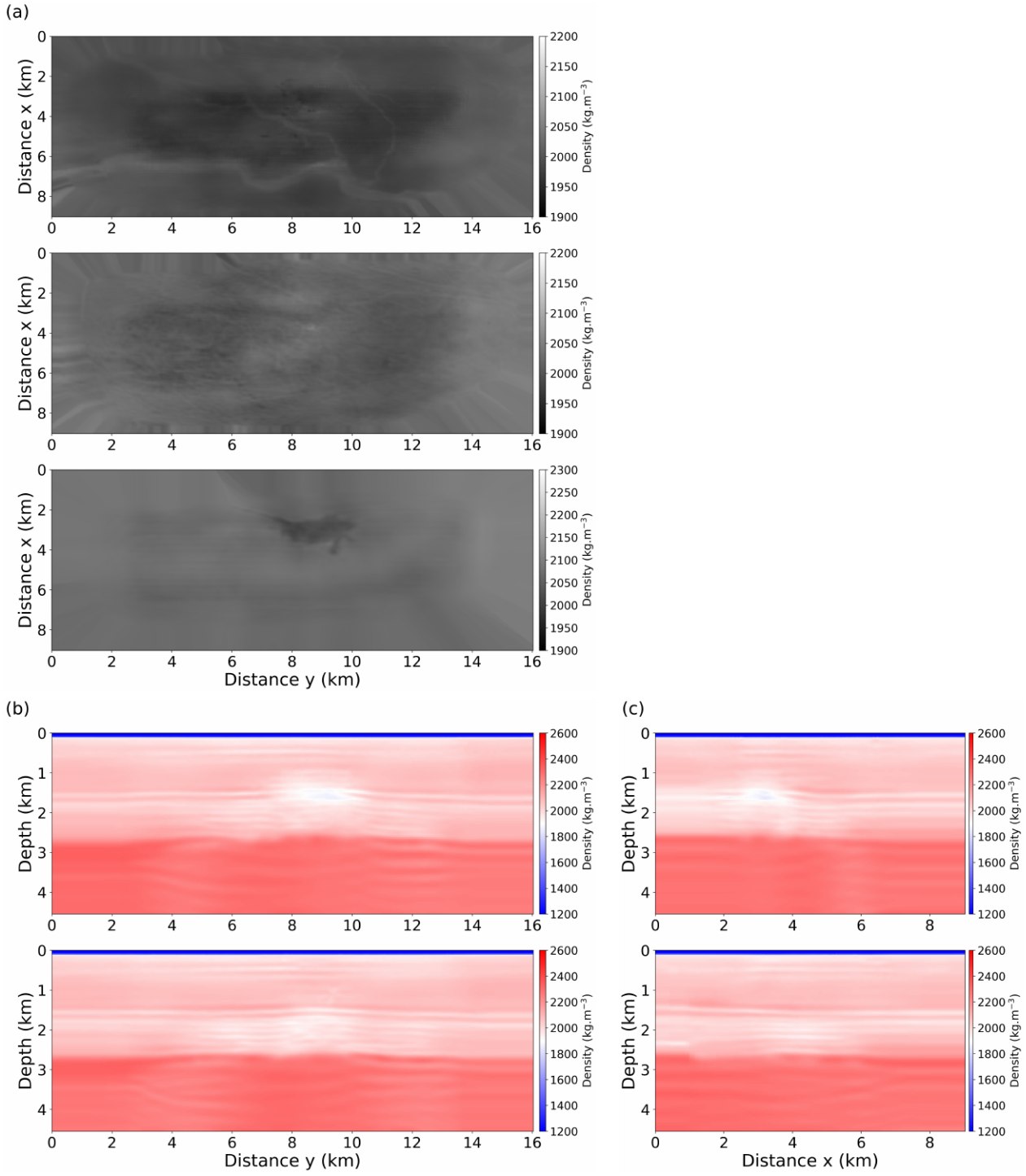

Figure 18: Final ρ model with non-stationary Gaussian smoothing. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

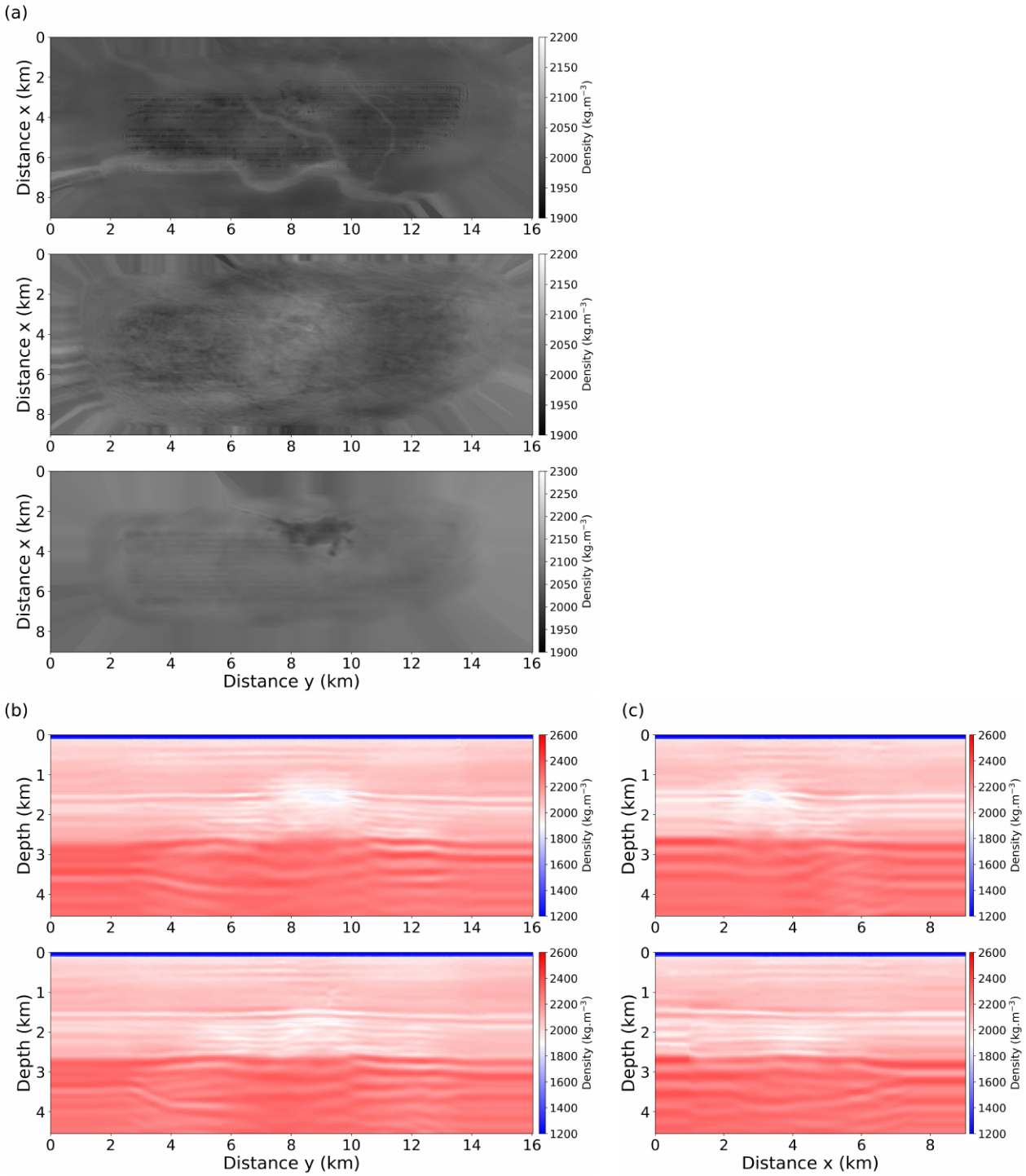

Figure 19: Final ρ model with linear anisotropic diffusion smoothing. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m, in black and white colorscale. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

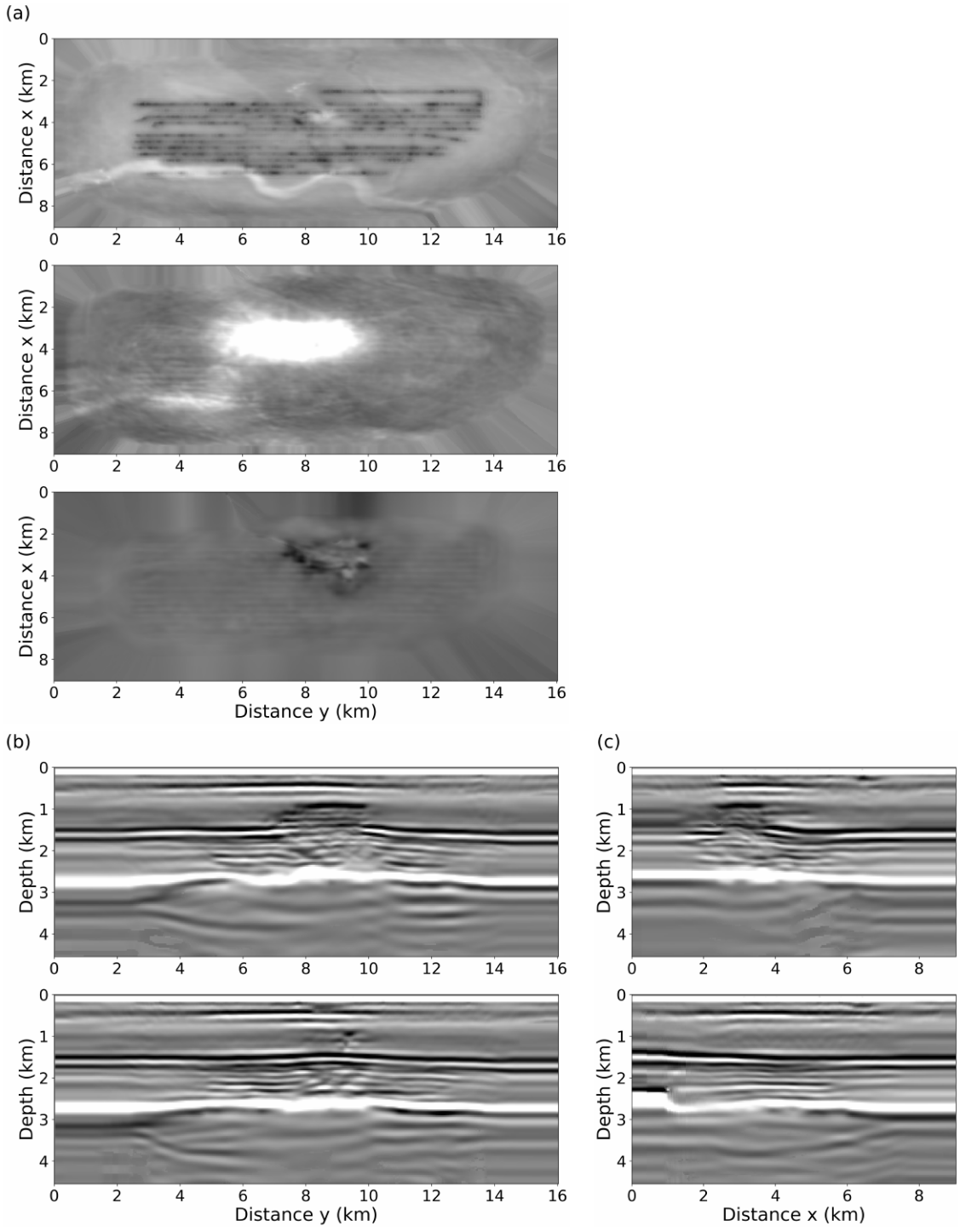

Figure 20: Final "full wavefield images" with Gaussian smoothing. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

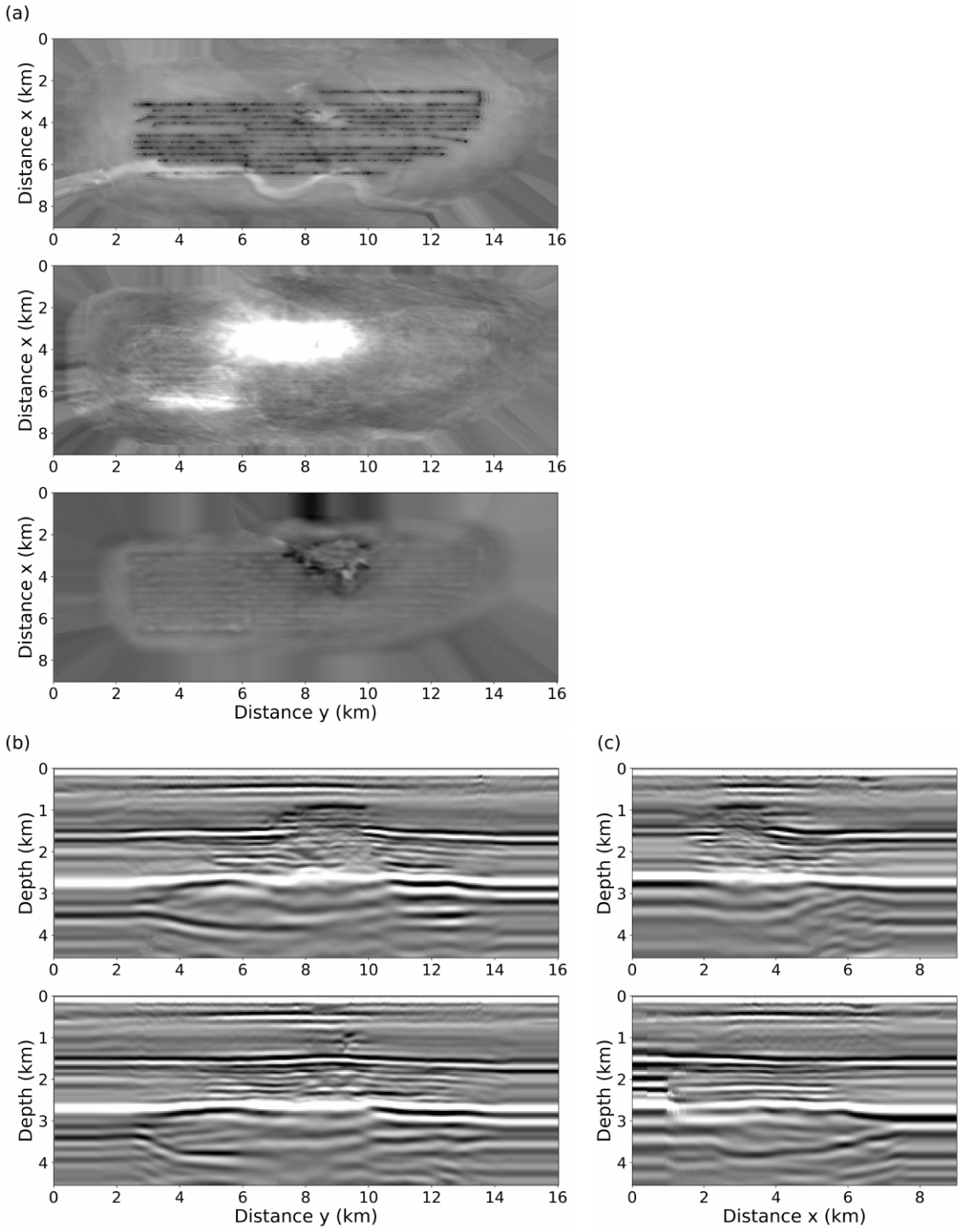

Figure 21: Final "full wavefield images" with linear anisotropic diffusion smoothing. (a) Constant depth slices at, from top to bottom: z = 175 m, z = 500 m, z = 1000 m. (b) Constant x slices at, from top to bottom: x = 2.95 km (passing through the low velocity anomaly), and x = 3.95 km (nearby its periphery). (c) Constant y slices at, from top to bottom: y = 9 km and y = 6 km.

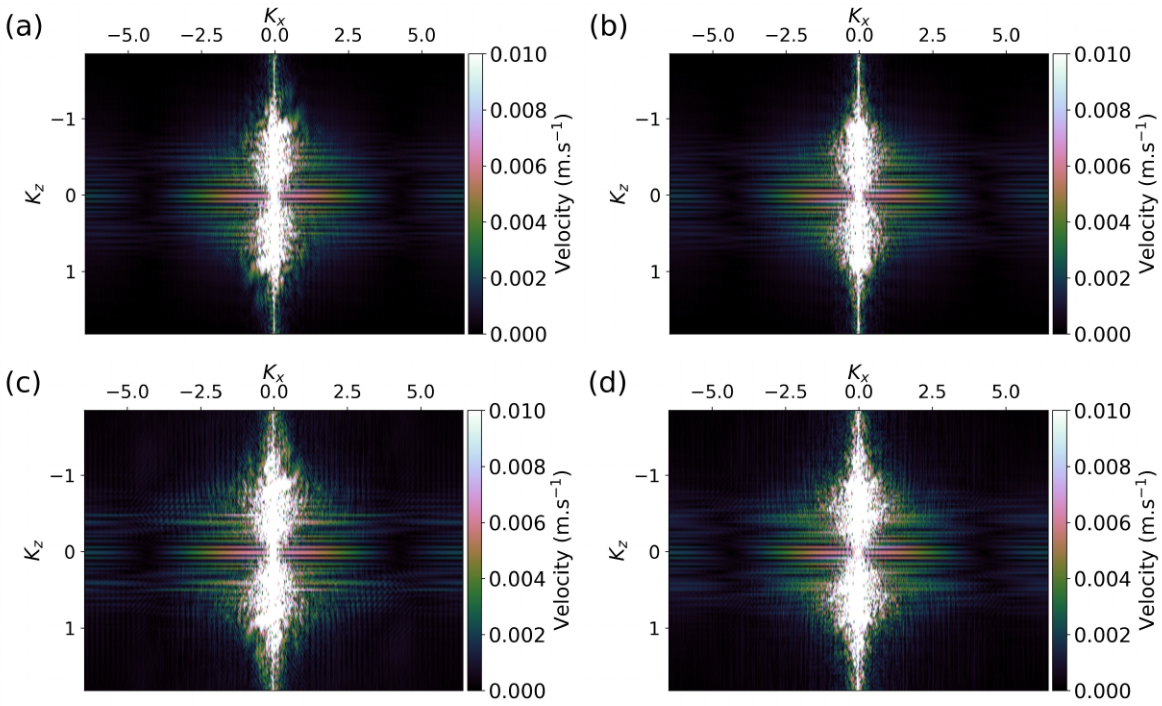

Figure 22: (a,b) Wavenumber spectrum of the reflectivity images depth sections in Figure 20b obtained with a Gaussian filter. (c,d) Wavenumber spectrum of the reflectivity images depth sections in Figure 21b obtained with the coherence enhancing diffusion filter.

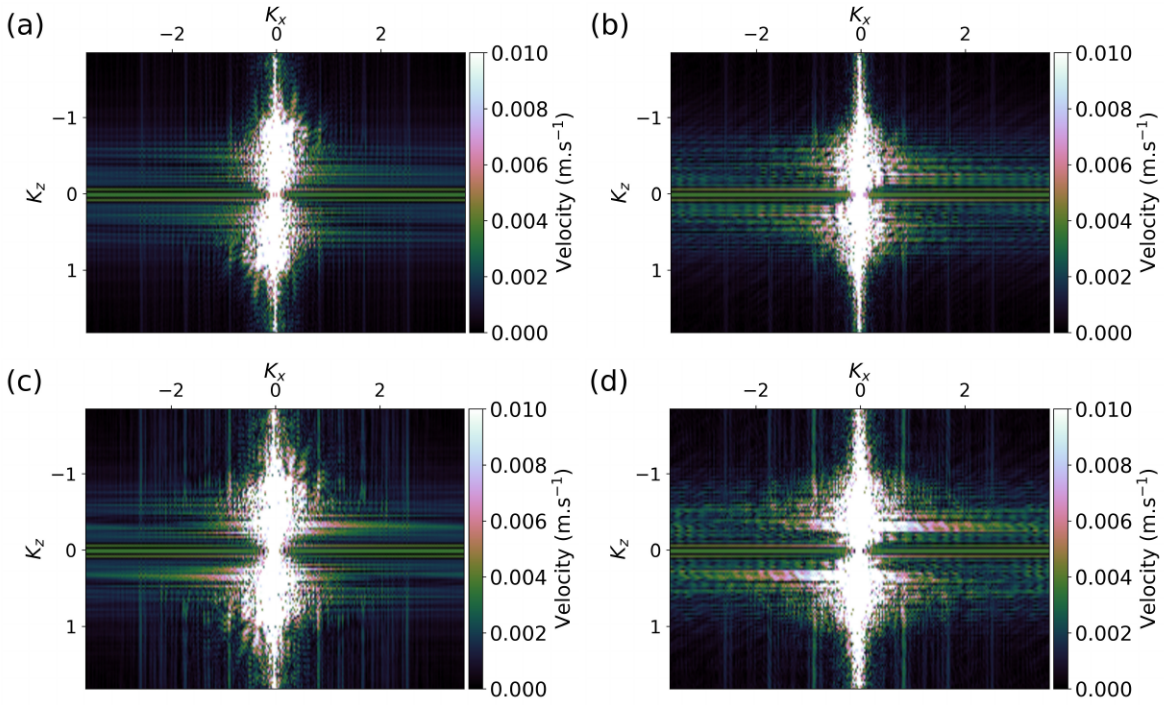

Figure 23: (a,b) Wavenumber spectrum of the reflectivity images depth sections in Figure 20c obtained with a Gaussian filter. (c,d) Wavenumber spectrum of the reflectivity images depth sections in Figure 21c obtained with the coherence enhancing diffusion filter.

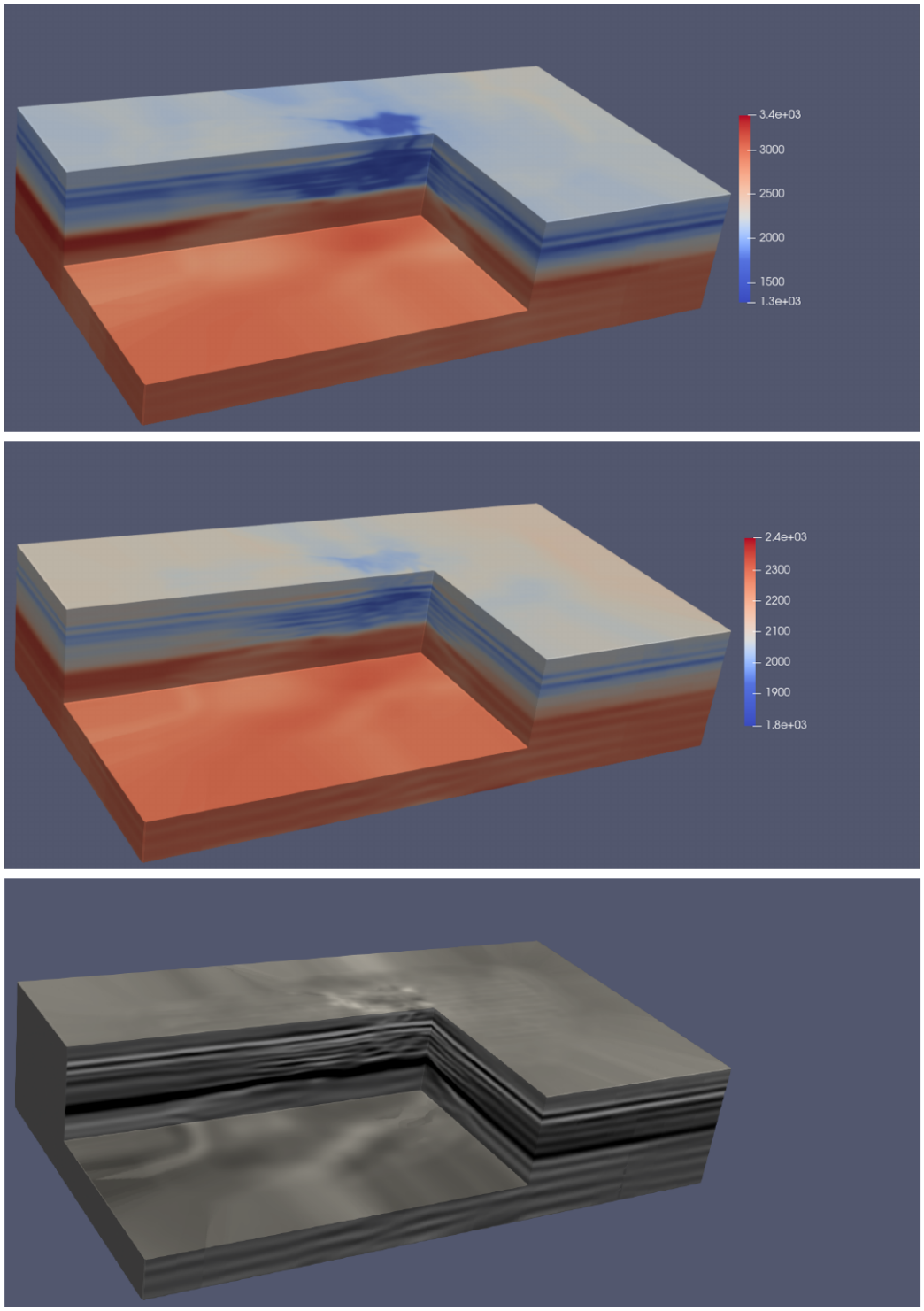

Figure 24: 3D view of the velocity model (top), density model (middle), reflectivity image (bottom) obtained by FWI with the nonstationary Gaussian filter.

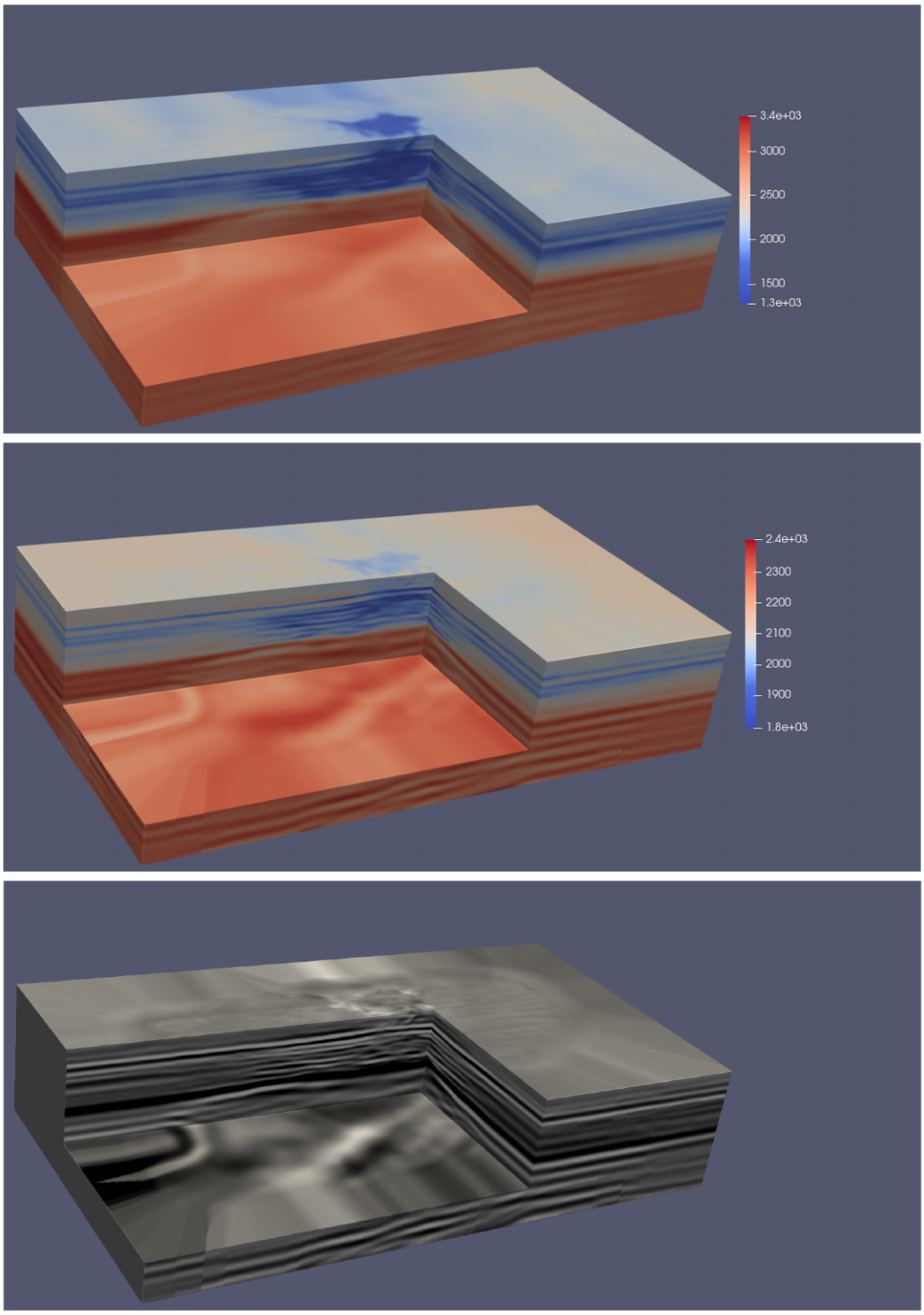

Figure 25: 3D view of the velocity model (top), density model (middle), reflectivity image (bottom) obtained by FWI with the anisotropic diffusion filter.

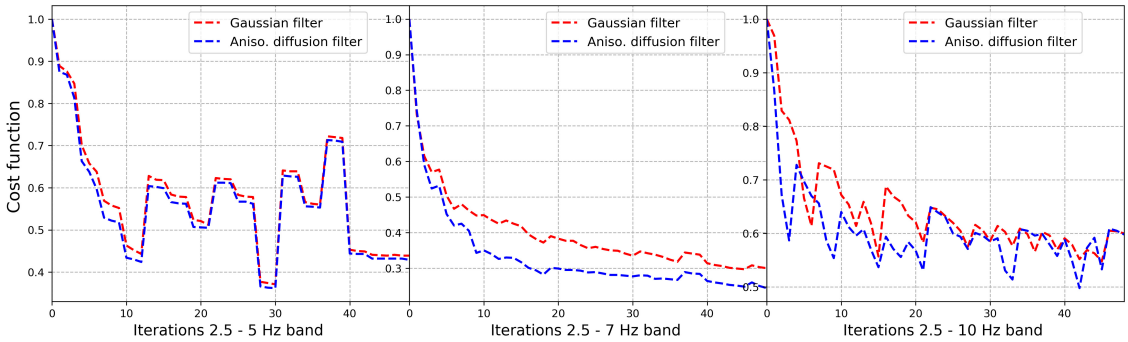

Figure 26: Misfit function decrease along the course of iterations over the three frequency bands 2.5 - 5 Hz, 2.5 - 7 Hz, 2.5 - 10 Hz.

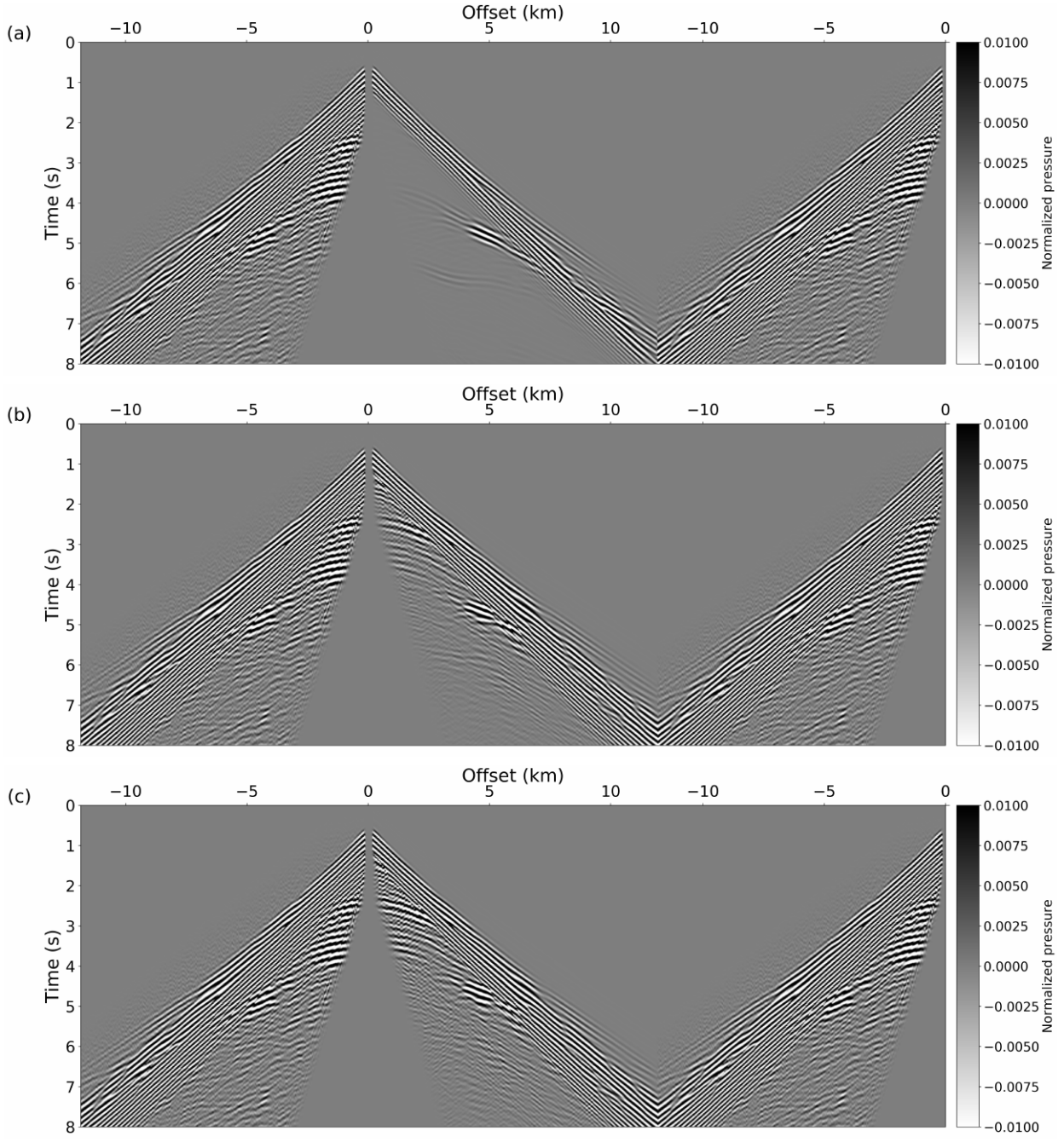

Figure 27: Data fit in mirror display for cable A. (a) Data fit in the initial model. (b) Data fit in the final model obtained using a Gaussian filter. (c) Data fit in the final model obtained using an anisotropic diffusion filter.

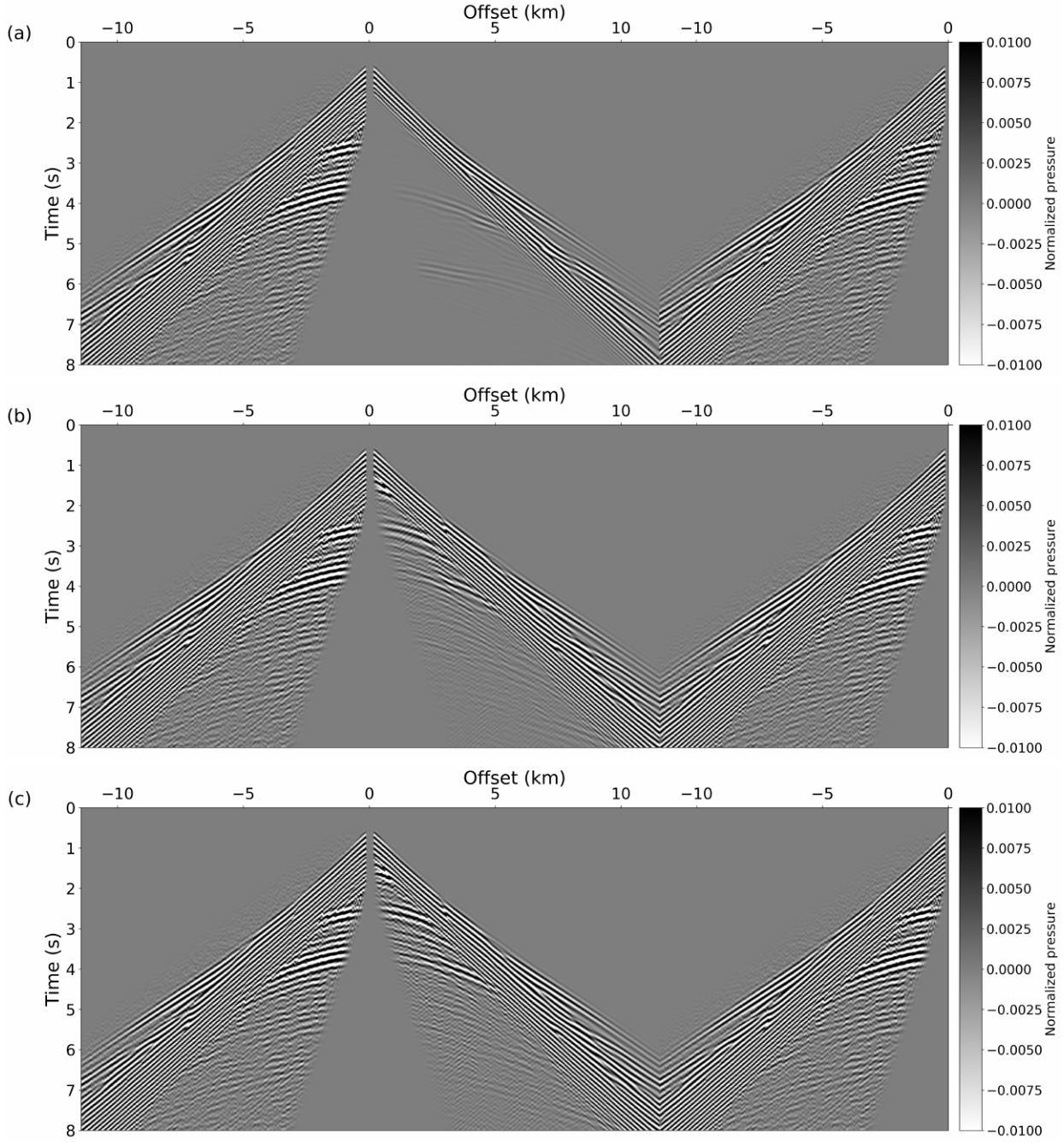

Figure 28: Data fit in mirror display for cable B. (a) Data fit in the initial model. (b) Data fit in the final model obtained using a Gaussian filter. (c) Data fit in the final model obtained using an anisotropic diffusion filter.

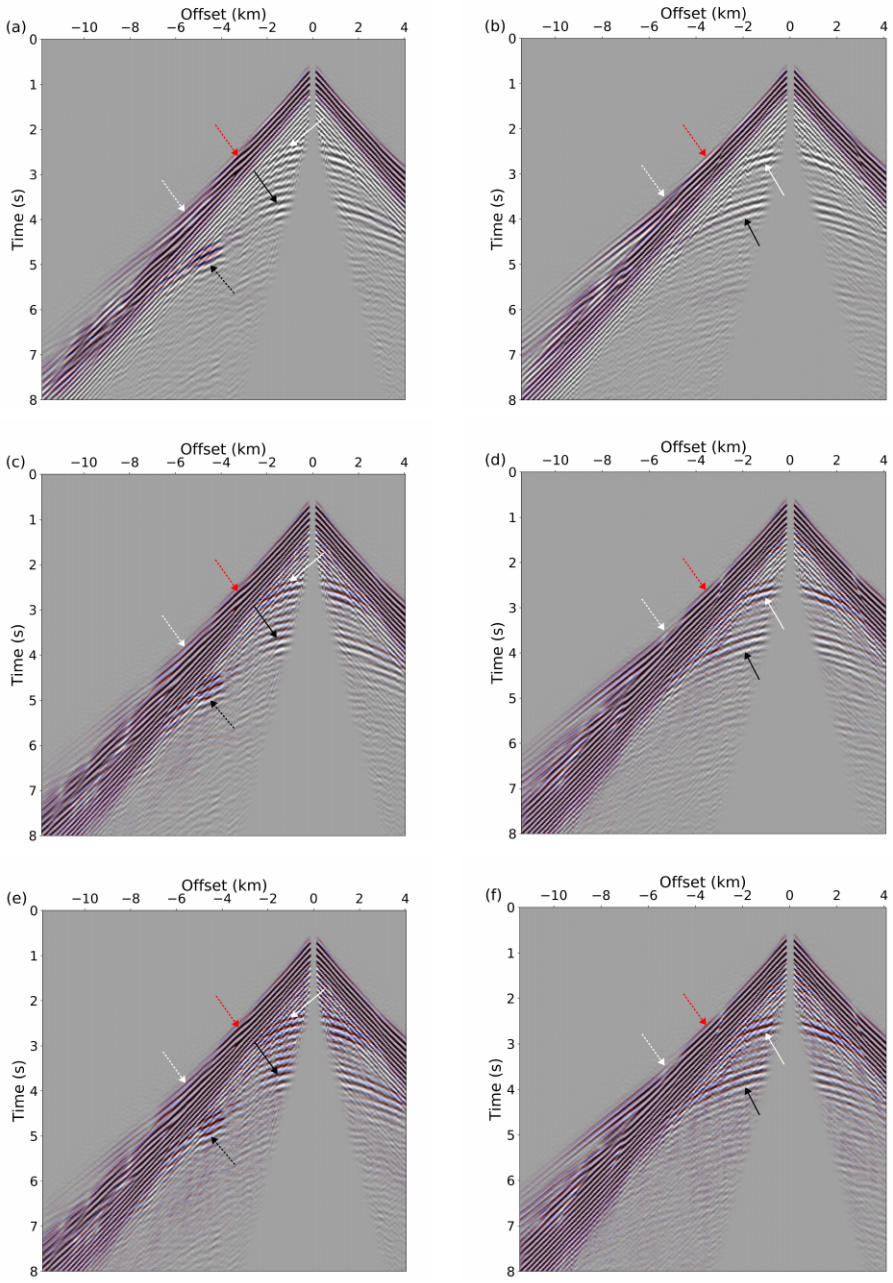

Figure 29: Comparison of the data fit by superposition of the field data in black/white colorscale and the predicted data in a red/blue colorscale. A good match is indicated by the absence of white and red color. (a,b) Data fit in the initial model for (a) cable A and (b) cable B. (c,d) Data fit in the FWI model obtained using a Gaussian filter for (c) cable A and (d) cable B. (e,f) Data fit in the FWI model obtained using an anisotropic diffusion filter for (e) cable A and (f) cable B. The red, white, black arrows point on the reflection from a shallow reflector, the top of the low velocity anomaly and the top of the structure below, respectively. The solid arrow points on the pre-critical reflections, while the dashed ones points on the post-critical reflections.

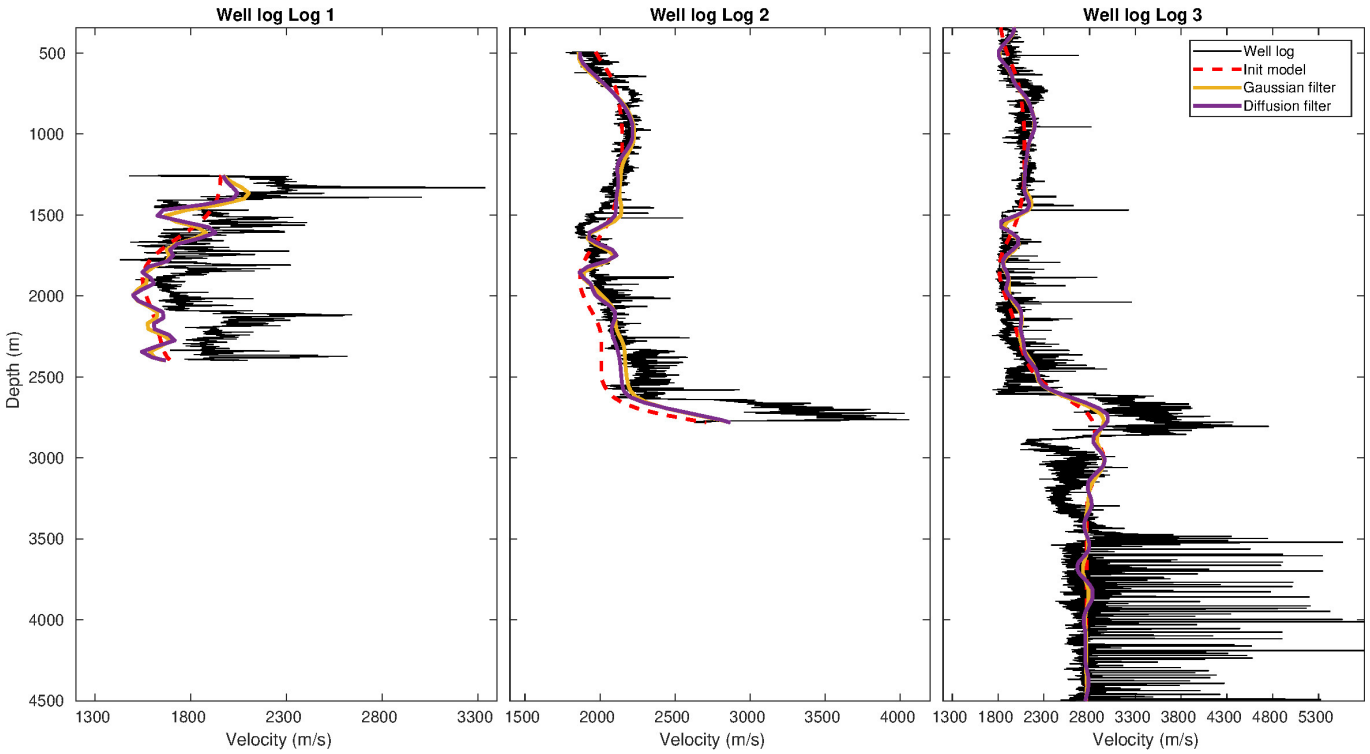

Figure 30: Comparison between sonic logs and velocity model values at the location of the wells where the sonic logs have been extracted. The sonic logs appear in solid black line. The initial model is in dotted-red line. The FWI model using a Gaussian filter is in solid orange line. The FWI model using an anisotropic diffusion filter is in solid purple line.

LIST OF TABLES

Type	Inc. field	Adj. field	Smoothing	Total	% smoothing
Gaussian $r_x = r_z = 0.4$	9.5 s	29.5 s	≤ 0.1 s	39 s	$\simeq 0\%$
Nonlin. Fehrenbach	9.5 s	29.5 s	13.6 s	52.6 s	25.8 $\%$
Nonlin. Weickert	9.5 s	29.5 s	12.1 s	51.1 s	23.7 $\%$
Lin. Fehrenbach	9.5 s	29.5 s	1.5 s	40.5 s	3.7 $\%$

Table 1: Comparison of different computation times depending on the type of smoothing which is applied. The Gaussian smoothing over truncated windows benefits from the possibility to tensorize the operation which makes it negligible in terms of computational cost. In comparison, the diffusion filter requires to solve a parabolic partial differential equation. The nonlinear version, independently of the stencil which is used, generates a non-negligible computational cost increase. The smoothing operation becomes 25% of the total time to build the gradient. The linear version, however, offers computational time closer from those associated with the use of Gaussian smoothing, for the 2D Marmousi experiment.

	2.5 - 5 Hz *(1 OMP)*	2.5 - 7 Hz *(8 OMP)*	2.5 - 10 Hz *(64 OMP)*
inc. field	163 s	150 s	154 s
adj. + inc. fields	649 s	632 s	609 s
Gaussian filt.	0.6 s ($< 1\%$ total time)	1.3 s ($< 1\%$ total time)	3.8 s ($< 1\%$ total time)
Linear aniso. filt.	142 s ($\simeq 14$ % total time)	149 s ($\simeq 18$ % total time)	285 s ($\simeq 26$ % total time)

Table 2: Computational cost associated with the filtering approach on the three different frequency bands 2.5 - 5 Hz, 2.5 - 7 Hz and 2.5 - 10 Hz. The total time refers to the total time for computing a single gradient. Depending on the frequency bands, a different number of OpenMP process are used, 1 for the 2.5 - 5 Hz band, 8 for the 2.5 - 7 Hz band, and 64 for the 2.5 - 10 Hz band. The Gaussian and anisotropic diffusion filter implementations do not enjoy OpenMP acceleration, which explains why the computational cost associated to the filtering part increases while the average elapsed time for building the incident and adjoint field remains similar.