



HAL
open science

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

► To cite this version:

Ludovic Métivier, Romain Brossier, Alexandre Hoffmann, Jean-Marie Mirebeau, Giuseppe Provenzano, et al.. 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. *Geophysics*, 2024, 89 (1), pp.R33-R58. 10.1190/geo2022-0648.1 . hal-04278886

HAL Id: hal-04278886

<https://hal.science/hal-04278886>

Submitted on 10 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1
2
3
4
5
6
7
8

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

2 Ludovic Métivier^{1,2}, Romain Brossier², Alexandre Hoffmann², Jean-Marie
3 Mirebeau³, Giuseppe Provenzano², Alizia Tarayoun² and Peng Yong²

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

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

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

7 (September 13, 2023)

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

ABSTRACT

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

16 3D and applied so as to filter and enhance the structural coherence of the model updates
17 within a full waveform inversion algorithm. The benefits of this strategy are investigated on
18 a 2D synthetic experiment before considering the multi-parameter inversion of a 3D field
19 dataset from the North sea up to 10 Hz. From this data, the vertical velocity and the density
20 are simultaneously reconstructed. Compared with a conventional nonstationary Gaussian
21 regularization strategy, the models obtained using the coherence enhancing anisotropic dif-
22 fusion strategy show an enhanced resolution, especially for the density model. The high
23 resolution reflectivity image computed from the impedance volume clearly illustrates the
24 benefit this filtering approach can deliver in terms of structural interpretation.

INTRODUCTION

25 Full waveform inversion (FWI) has become a prominent seismic imaging tool during the
26 last decade. It is used at various scales, from global-scale (Bozdağ et al., 2016; Karaoğlu
27 and Romanowicz, 2018; Lei et al., 2020; Thrastarson et al., 2022), lithospheric/continental
28 scales (Yuan et al., 2014; Fichtner and Villaseñor, 2015; Lu et al., 2020) and deep crustal
29 scales (Górszczyk et al., 2017), to exploration scale with dense acquisition and active
30 sources, with depth of investigation reaching few kilometers (Sirgue et al., 2010; Plessix
31 and Perkins, 2010; Warner et al., 2013; Stopin et al., 2014; Vigh et al., 2014; Operto et al.,
32 2015; Raknes et al., 2015; Solano and Plessix, 2019). In the last years, smaller scale ap-
33 plications have also been increasingly investigated: from near surface scales (KoeHN et al.,
34 2018; Irnaka et al., 2022) to medical imaging (Guasch et al., 2020; Marty et al., 2021). A
35 review on FWI and its application can be found for instance in Virieux et al. (2017).

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

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

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

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

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

68 preted as a modification of the scalar product used to express the Taylor's development of
69 the misfit function at the core of local optimization (Zuberi and Pratt, 2017; Adriaens et al.,
70 2021). This provides a nice mathematical framework to include different smoothing oper-
71 ators with different properties. A common choice is to use a Gaussian smoothing operator,
72 potentially non-stationary, with coherent lengths based on an estimation of the expected
73 local resolution. The latter is estimated from the local velocity and a reference frequency
74 (for instance the peak frequency of the inverted data). This strategy has been implemented
75 in early FWI applications to field data (Operto et al., 2006)

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

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

90 (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
91 from the updated reflectivity model to inject it to the smoothing operator applied in the
92 reconstruction of the macro-velocity model. This strategy has been applied in Provenzano
93 et al. (2022) based on a Bessel filtering approach. The same could be performed in the
94 frame of migration velocity analysis (Symes, 2008), which relies on the same scale separation
95 between macro-velocity and reflectivity, and updates these two parameters in an
96 iterative fashion. This has not yet been implemented in this framework to the best of our
97 knowledge.

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

105

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

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

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

134 image. This idea has been exploited for instance in Lewis et al. (2014) and Lee and Pyun
135 (2018) and to a certain extent in Xue et al. (2020). In those studies, the design of the diffu-
136 sion operator is not really detailed, although it relies on the structure tensor associated with
137 the gradient and its spectral decomposition. What we propose here is to generalize to 3D
138 the coherence-enhancing diffusion strategy introduced by Weickert (1998) which is a spe-
139 cific way to design the anisotropic diffusion operator. It uses a local measure of coherence
140 along directions in the plane orthogonal to the direction of fastest variations, and a thresh-
141 old function to control the transition from weak diffusion to strong diffusion depending
142 on the orientation. We shall add that what makes this diffusion-type technique appealing
143 over other aforementioned directional filtering techniques are the mathematical properties
144 inherited from the partial differential equations formalism (preservation of the maximum,
145 minimum and mean value, and the scale-space property) as well as the automation of the
146 structure detection through the design of the anisotropic diffusion operator, without having
147 to compute angle maps and correlation lengths prior the application of the filter, and the
148 relatively limited number of tuning parameters (essentially the diffusion time). We have
149 studied this technique already in the frame of 2D FWI, with the idea that it could be ap-
150 plied in the model space to filter the gradient, and in the data space to filter and denoise
151 the data while enhancing its low frequency content (Métivier and Brossier, 2022). In the
152 present study, after providing some illustrations on a 2D synthetic example, we apply it to
153 the 3D multi-parameter inversion of a North Sea OBC dataset up to 10 Hz. We provide a
154 comparison between applying a conventional non-stationary Gaussian smoothing and our
155 coherence enhancing smoothing. We show that an isotropic Gaussian smoothing has a ten-
156 dency to destroy the structure, especially at depth, which is detrimental to the final model

157 resolution and also penalizes the convergence of the whole FWI algorithm. Conversely,
158 the FWI strategy coupled with the coherence enhancing smoothing makes it possible to
159 recover sharper interfaces in the whole domain. Then this also enhances the reconstruction
160 of a sharper density model. The coherence enhancing strategy is also efficient to remove
161 the footprint of the acquisition in the shallow part of the model, which is all the more im-
162 portant when a source subsampling method is used to decrease the computational cost of
163 FWI as is the case in our 3D OBC application. We believe these features are particularly
164 valuable for high resolution FWI, a current trend in the industry to push FWI towards di-
165 rectly interpretable models inverted from data in frequency bands similar as what is used
166 for migration (Shen et al., 2018; Wang et al., 2019; Huang et al., 2021; Kalinicheva et al.,
167 2020).

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

METHODOLOGY: A NONLINEAR ANISOTROPIC DIFFUSION

APPROACH

172 **Generalities**

173 The structural smoothing technique we propose originates from the work of Weickert
174 (1998) for image processing. Consider an image $I(\mathbf{x})$ with $\mathbf{x} \in \Omega \subset \mathbb{R}^d$, d being the

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

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

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

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

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

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

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

189 with $*$ the convolution in space and

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

190 This shows the equivalence between isotropic diffusion and Gaussian smoothing.

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

$$S(u) = \nabla u \nabla u^T. \quad (5)$$

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

$$u_\sigma = K_\sigma * u. \quad (6)$$

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

$$S_{\ell,\sigma}(u) = K_\ell * (\nabla u_\sigma \nabla u_\sigma^T), \quad (7)$$

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

207 **2D coherence enhancing filter**

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

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

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

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

$$D(u) = P(u)^T \Sigma(u) P(u), \quad (8)$$

221 where $\Sigma(u)$ is defined as

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

222 with g a threshold function from α to 1

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

223 where α and C are user-defined scalar values. A profile of the threshold function $g(x)$
224 depending on C for $\alpha = 10^{-2}$ is given in Figure 1. Playing with the constant C amounts
225 to playing with the steepness of the threshold (how fast the function moves from α to
226 1). Of note, in all the following numerical experiments with the diffusion filter, we use
227 the steepest threshold function presented here, which corresponds to the parameter values
228 $\alpha = 10^{-2}$ and $C = 10^{-8}$. After testing different values it turns out that a stiff transition
229 provides qualitatively better results for our application.

230 [Figure 1 about here.]

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

239 **3D coherence enhancing filter**

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

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

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

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

$$D(u) = P(u)^T \Sigma(u) P(u), \quad (11)$$

253 where $\Sigma(u)$ is defined as

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

254 with g the same threshold function as in equation 10.

255 In the same spirit as in the 2D case, following this definition of $D(u)$, as soon as co-
256 herent structures are detected along v_2 or v_3 by an increase of the quantities $(\lambda_1 - \lambda_2)^2$ or
257 $(\lambda_1 - \lambda_3)^2$, the diffusion weight increases rapidly to 1 while remaining small and equal to
258 α in the direction of the main variation v_1 . If no coherent structure is detected, the diffu-
259 sion remains close to isotropic, with similar small α values in each direction. This extends
260 the coherence-enhancing filter to the 3D case, with again the ability to smooth along de-
261 tected structures and not across them. As for the 2D case, the detection of the structures is
262 performed automatically thanks to the SVD of the averaged structure tensor $S_{\ell,\sigma}(u)$.

263 **Discretization of anisotropic diffusion equations**

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

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

274 non-negative discrete matrices $D(u)$, *i.e.* matrices with non-negative off-diagonal terms.
 275 In Weickert (1998) a 2D scheme is proposed, which guarantees this non-negativity for any
 276 diffusion matrix such that the ratio λ_1/λ_2 between the highest and lowest eigenvalue in
 277 $D(u)$ is bounded by $1 + \sqrt{2} \approx 2.41$. This is the scheme we have used in Métivier and
 278 Brossier (2022). However no extension to 3D is proposed in Weickert (1998).

279 In the current study, we rely on more sophisticated schemes proposed in Fehrenbach and
 280 Mirebeau (2014). These schemes are based on a Lattice-Basis-Reduction technique. They
 281 can be implemented in 2D and 3D, and can guarantee the non-negativity of the discretiza-
 282 tion of $D(u)$ for any value of λ_1/λ_2 or λ_1/λ_3 . The stencils are adapted to finite-difference
 283 Cartesian meshes. The computational complexity to build the stencils is low, which makes
 284 the method very efficient. Describing in detail how these schemes are derived is outside
 285 the scope of this paper, for this we refer the interested reader to Fehrenbach and Mirebeau
 286 (2014) and references therein.

287 **Implementation within a full waveform inversion algorithm**

288 We formulate FWI as

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

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

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

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

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

$$m_{k+1} = m_k + \alpha_k \Delta m_k, \quad (15)$$

302 where α_k is a linesearch parameter satisfying the Wolfe's conditions (Nocedal and Wright,
 303 2006) to ensure convergence towards the nearest local minimum. The descent direction
 304 Δm_k is computed as

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

305 where the operator \mathcal{F} stands for the filtering/smoothing operator we apply to the gradient
 306 and Q_k is a preconditioning matrix. We rely on the l -BFGS algorithm to estimate Q_k as
 307 an approximation of the inverse Hessian operator $(\nabla^2 f(m_k))^{-1}$ from l previously com-
 308 puted filtered gradient values $\mathcal{F}(\nabla f(m_k)), \mathcal{F}(\nabla f(m_{k-1})), \dots, \mathcal{F}(\nabla f(m_{k-l+1}))$ (No-
 309 cedal, 1980). Note that the l -BFGS method makes it possible to combine Q_k with any
 310 other preconditioning operator P approximating $(\nabla^2 f(m_k))^{-1}$ (Métivier and Brossier,

311 2016).

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

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

315 where

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

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

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

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

323 To implement this filter in an efficient way, we exploit the separability of the kernel
324 $\mathcal{G}(\mathbf{x})$ to perform a series of 1D convolutions in space. This makes the computational cost
325 of the application of this nonstationary Gaussian filter negligible compared to the gradient
326 computation as is shown in the numerical experiments presented in the sequel.

SYNTHETIC EXPERIMENTS

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

329 **Data generation**

330 We generate observed data in the 2D acoustic approximation with variable density using
331 our 2D/3D time-domain full waveform modeling and inversion engine TOYxDAC_TIME,
332 developed in the frame of the SEISCOPE project (Yang et al., 2018). This code uses a 4th
333 order finite-differences spatial discretization and a 2nd-order finite-differences time dis-
334 cretization of the velocity/stress equations. The reference data is built from the velocity
335 and density models presented in Figure 2(a,b). These models have been obtained by up-
336 scaling on a 5 m grid the original 1.25 m grid Marmousi II models (Martin et al., 2006). To
337 generate the data, we use a fixed spread acquisition with 128 shot positions each 132.5 m
338 apart, and 169 receivers each 100 m apart. The source wavelet, presented in Figure 2(c) is
339 a Ricker wavelet centered on 5 Hz, low-cut below 2.5 Hz to mimic a realistic exploration
340 scenario in which such low frequencies are not available. The spectrum of the wavelet is
341 presented in Figure 2(d). A Gaussian white noise filtered in the frequency-band from 0 to
342 12.5 Hz is added to the data. Two shot gathers of this reference dataset are presented in
343 Figure 3.

344 [Figure 2 about here.]

345 [Figure 3 about here.]

346 **Initial model and source wavelet estimation**

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

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

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

357 [Figure 4 about here.]

358 **Gradient: comparison between Gaussian and anisotropic diffusion fil-** 359 **ter**

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

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

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

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

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

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

389 [Figure 5 about here.]

390 [Figure 6 about here.]

391 [Figure 7 about here.]

392 **Computational cost: linear vs nonlinear filter, Weickert vs Fehrenbach** 393 **stencil**

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

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

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

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

412 In Figure 8, we present this error map for the gradients presented in Figure 7 and the ones
 413 obtained with a linear anisotropic diffusion filter. The highest error value reaches locally
 414 0.5%. The error is weaker and more localized for a small number of diffusion iterations
 415 (50) and simultaneously increases in amplitude and spreads over the model as the number
 416 of iterations grows (up to 800). This is expected: the nonlinear effect increases with the
 417 number of iterations as, on the one hand, the diffusion matrix $D(u)$ is recomputed at each
 418 diffusion iteration (nonlinear diffusion) while, on the other hand, it is kept the same for all
 419 diffusion iterations (linear diffusion).

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

426 is developed for 2D and 3D filtering.

427 [Figure 8 about here.]

428 [Table 1 about here.]

429 **Inversion: comparison between Gaussian and linear anisotropic diffu-** 430 **sion filter**

431 To conclude this 2D synthetic test, we compare the results obtained after 50 FWI iterations
432 using the non-stationary Gaussian filter and the linear anisotropic diffusion filter, starting
433 from the P-wave velocity and density models presented in Figure 4(a,b). We invert only for
434 the P-wave velocity, the density is not updated (passive parameter). We use the estimated
435 wavelet presented in Figure 4(c,d). The results are presented in Figure 9. While the final
436 models (Fig. 9(a,b)) are similar, especially in the shallowest part above 2 km depth, the
437 resolution of the final model obtained with the anisotropic diffusion filter is higher. The
438 bent layering structure below 3 km depth is better reconstructed especially in the central
439 part between 6 and 12 km in horizontal distance. The central structure, between 1.5 and
440 2.5 km depth and between 6 and 12 km in horizontal distance is also better resolved, with
441 higher contrasts between fast and slow layers. This is confirmed in Figure 9(c,d) where
442 we compare the decrease of the data misfit and the model misfit along the FWI iterations,
443 depending on the choice of the filter. Using the anisotropic diffusion filter, both the data
444 and model errors decrease faster.

APPLICATION TO THE 3D OBC FIELD DATA

446 We apply in this section our coherence enhancing diffusion filter and compare it to the stan-
447 dard nonstationary Gaussian convolution filter in the framework of time-domain FWI of a
448 3D OBC field data. Unlike most of the studies performed on a similar environment (Sirgue
449 et al., 2010; Operto et al., 2015; Kamath et al., 2021; Pladys et al., 2022), we consider here
450 a multi-parameter P-wave velocity (v_P), density (ρ) inversion. The reconstruction of the
451 density together with the vertical velocity from a similar dataset has been tackled only in
452 the frequency-domain in the work of Operto and Miniussi (2018).

453 **Field data presentation and pre-processing**

454 The data we consider has been recorded in a shallow water environment, in the North sea.
455 Several FWI studies have focused on similar data since the seminal paper of Sirgue et al.
456 (2010), which presents the first 3D FWI application to industrial field data. The particular
457 dataset we use is a 4-components ocean bottom cable (OBC) dataset, with 2048 receivers
458 deployed on the seabed and 50824 shots at 5 m depth. The area covered by the shots is
459 around 145 km². We present in Figure 11 the acquisition layout. The field is characterized
460 by a shallow water environment (the water column is approximately constant equal to 70
461 m in the whole domain) and an anticlinal in chalk in the Upper Cretaceous Hod and Tor
462 formations. An apparent low velocity anomaly in the upper part of the model from 1 km

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

466 [Figure 10 about here.]

467 [Figure 11 about here.]

468 We focus on the hydrophone component of the data. We apply source-receiver reci-
469 procity to reduce the computational cost. Therefore, in the following, what is called
470 common shot gather (CSG) refers actually to a common receiver gather (CRG). The pre-
471 processing of the data is minimal: we apply a simple despiking, and apply a minimum-
472 phase band-pass filter to the data to generate three distinct datasets in the 2.5 - 5 Hz, 2.5
473 - 7 Hz and 2.5 - 10 Hz frequency band respectively. For each dataset, we compute the
474 total amplitude of all CSG and remove the ones with anomalous amplitudes. Four of them
475 are excluded in bands 2.5 - 5 Hz and 2.5 - 7 Hz, and one only is excluded in the band
476 2.5 - 10 Hz (the signal over noise ratio increases with the frequency bandwidth). Finally,
477 the visco-acoustic approximation we use makes it impossible to model the Schölte waves
478 propagating at the fluid-solid interface. These waves are thus muted from the observed data,
479 using a simple time-windowing based on a linear velocity of approximately $400 \text{ m}\cdot\text{s}^{-1}$ for
480 these waves. As an illustration, we present in Figure 12 two CSG corresponding to the
481 positions of cables A and B, before and after the mute is applied, filtered in the 2.5 - 10
482 Hz band. We choose these CSG because cable A intersects the low velocity anomaly while
483 cable B is further away from the low velocity anomaly. We reproduce a data anatomy anal-

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

488 [Figure 12 about here.]

489 **FWI setup**

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

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

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

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

$$\rho = 309.6V_p^{0.25}. \quad (22)$$

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

$$\frac{\epsilon - \delta}{1 - 2\delta}, \quad (23)$$

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

527 [Figure 13 about here.]

528 [Figure 14 about here.]

529 [Figure 15 about here.]

530 We use a conventional multi-scale inversion (Bunks et al., 1995), where the model in-
531 verted in band $i - 1$ serves as initial model for the model in band i . The spatial discretization
532 step h is set respectively to 70, 50 and 35 m for frequency bands 2.5 - 5 Hz, 2.5 - 7 Hz, and
533 2.5 - 10 Hz. This satisfies the usual criterion to have at least 4 points per wavelength us-
534 ing our 4-th order finite difference discretization scheme in space to minimize numerical
535 dispersion while ensuring minimal computational cost (Igel, 2017). To decrease the com-
536 putational cost of the inversion further we rely on a source subsampling strategy (Warner
537 et al., 2013; Kamath et al., 2021; Pladys et al., 2022). The initial ensemble of shots is de-
538 composed in 16 groups, the first 15 being composed of 128 shots, and the 16th containing
539 124 sources for the two bands 2.5 - 5 Hz and 2.5 - 7 Hz and 127 sources for the band 2.5
540 - 10 Hz. One inversion over a given frequency band then consists in performing 3 l -BFGS
541 iterations over each group, leading to 48 iterations in total. The 16 groups do not over-
542 lap (one shot belongs to a single group) to ensure each shot is used once per inversion.

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

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

553 The preconditioner we use is similar to the one used in Kamath et al. (2021, Appendix
554 B). It applies both to v_P and ρ gradient, and consists in an illumination compensation
555 through the wavefield values. The Hessian approximation is diagonal where each element
556 is computed as an integration over time and over shots of the source wavefield. For the
557 shallow part, from $z = 0$ to $z = 800$ m, these wavefield based values are replaced with a
558 linear function of depth to avoid a strong imprint of uneven illumination associated with the
559 source subsampling strategy. We also use a unity-based normalization for both parameters
560 to recast the values of the inverted parameters between 0 and 1, based on a given authorized
561 range of variation for both parameters (Yang et al., 2018). The velocity is bounded between
562 1280 m.s^{-1} and 3650 m.s^{-1} . The density is bounded between 1000 kg.m^{-3} and 2800
563 kg.m^{-3} .

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

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

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

$$r_x = 0.4, \quad r_y = 0.4, \quad r_z = 0.3, \quad (24)$$

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

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

$$\sigma_x = 2h, \quad \sigma_y = 2h, \quad \sigma_z = h, \quad (25)$$

592 with h varying from 70, 50 and 35 m depending on the frequency band as already stated. We
593 use the same smoothing both from the noise and coherence scale ($\ell_x = \sigma_x, \ell_y = \sigma_y, \ell_z =$
594 σ_z). In addition, the number of iterations for the anisotropic diffusion is set respectively to
595 80, 40 and 20 depending on the frequency band. These values have been set by trial and
596 error on an initial gradient estimation for each band.

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

599 **Multi-parameter FWI results**

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

605 on the paleo-seafloor at $z = 500$ m, and the refined shape of the low velocity anomaly at
606 $z = 1000$ m. These structures are absent from the initial tomography model in Figure 13.
607 The layering horizontal structure of the low velocity anomaly is revealed by FWI from the
608 smooth blob of the initial model in the traversing and adjacent slices presented in Figures
609 16(b,c) and 17(b,c). A chimney connecting the low velocity anomaly to the top part of the
610 model also appears (Fig. 16(b) and 17(b)). Deeper, the base cretaceous reflector between
611 3.5 and 3.7 km depth is revealed.

612 A comparison between the v_P reconstruction results with the Gaussian (Fig. 16) and
613 anisotropic diffusion (Fig.17) filters shows an enhancement of the resolution using the dif-
614 fusion filter. This enhancement seems to vary with depth. In the shallow part, down to
615 1 km, the results remain very similar, even if some differences can be noted, especially a
616 sharper delineation of the top of the low velocity anomaly using the diffusion filter. Be-
617 tween 1 km and the top of the structure below, the resolution enhancement is more visible.
618 The low velocity anomaly layered structure is made finer and more coherent thanks to the
619 diffusion filter. More details are also visible on the chimney displayed in the second inline
620 section. Finally, the resolution enhancement is striking in the deeper part of the model.
621 The reconstruction of the base cretaceous reflector between 3.5 and 3.7 km depth is much
622 clearer using the diffusion filter compared with the Gaussian filter results.

623 [Figure 16 about here.]

624 [Figure 17 about here.]

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

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

641 [Figure 18 about here.]

642 [Figure 19 about here.]

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

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

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

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

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

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

663 [Figure 20 about here.]

664 [Figure 21 about here.]

665 The overall resolution of the reflectivity images obtained using the diffusion filter ap-
666 pears higher. To further analyze this point, we compare in Figures 22 and 23 the wavenum-
667 ber spectra of the reflectivity images vertical profiles presented in Figures 20(b,c) and
668 21(b,c). The spread of the wavenumber spectrum of the images derived from the FWI
669 results using the diffusion filter is broader than their counterpart obtained using the Gaus-
670 sian filter, which is indicative of a higher resolution.

671 [Figure 22 about here.]

672 [Figure 23 about here.]

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

681 [Figure 24 about here.]

682 [Figure 25 about here.]

683 **Quality control**

684 To assess the quality of the results we start by investigating the decrease of the misfit
685 function over the course of iterations, for the three frequency bands 2.5 - 5 Hz, 2.5 - 7
686 Hz, 2.5 - 10 Hz (Fig. 26). The misfit function is normalized to 1 at the beginning of each
687 band, and 48 iterations are performed on each band. As we rely on a source subsampling
688 strategy, we can identify jumps in the misfit function, each 3 iterations, associated with
689 the inversion of data associated with a new subgroup of sources. This pattern is especially
690 visible in the first and last frequency band, while the convergence appears smoother in
691 the second frequency band. Interestingly, we can observe that FWI using the anisotropic
692 diffusion filter achieves a faster decrease of the misfit function on all of the three frequency
693 bands. This is particularly pronounced on the 2.5 - 7 Hz band and at the beginning of the
694 2.5 - 10 Hz but is also visible in the 2.5 - 5 Hz band. This confirms what has been observed
695 on the 2D Marmousi synthetic experiment: enhancing the coherent features in the gradient
696 can accelerate the convergence of the FWI machinery.

697 [Figure 26 about here.]

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

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

715 [Figure 27 about here.]

716 [Figure 28 about here.]

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

723 [Figure 29 about here.]

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

734 [Figure 30 about here.]

DISCUSSION

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

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

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

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

766 the noise and coherence scale σ and ℓ , and the number of iterations to solve the anisotropic
767 diffusion equation. Regarding α , a value of 10^{-2} ensure a total anisotropy ratio of a factor
768 100 which seems sufficient to capture most of the structure in the gradient. The value
769 of 10^{-8} for C ensures a fast transition from weak to strong diffusion. For the noise and
770 coherence scale σ and ℓ , it appears that selecting the same values for both parameters is
771 efficient. This reduces the number of parameters to set. We have adapted them to the finite-
772 difference grid used to solve the wave equation in our multi-scale approach, using a size of
773 one or two discretization steps h depending on the direction. Given the fact that the model
774 under investigation exhibits more regularity in the horizontal directions x and y we have
775 taken σ and ℓ as respectively $2h$, $2h$ and h in the x, y, z directions. This is a light injection
776 of prior information on the model structure. Finally, regarding the number of diffusion
777 iterations, we set it manually by trial-and-error, but an initial guess can be found benefiting
778 from the analogy between Gaussian smoothing and isotropic diffusion (equations 3 and 4).
779 Using this analogy, one can find a diffusion time T based on a given smoothing length L
780 and devise the corresponding number of diffusion iterations based on the diffusion time-
781 step, controlled by the CFL of the scheme of Fehrenbach and Mirebeau (2014). On top of
782 that, the same filter is applied here to the velocity and density model without the need to
783 design a specific filter for each parameter.

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

788 CPU machine with 64 cores at 2.6 GHz per processors and 2 processors per node. Our code
789 TOYxDAC_TIME implements a double level of parallelization. The outer one is at the shot
790 level. The inner one is an OpenMP acceleration for the computation of the incident and
791 adjoint wavefields for each source. We activate it only for the 2.5 - 7 Hz and 2.5 - 10 Hz
792 frequency bands. In these bands, we use respectively 8 and 64 OpenMP threads per source.
793 As we use batches of 128 shots with our subsampling strategy, FWI thus runs on 128 cores
794 on the 2.5 - 5 Hz band, then 1024 cores on the 2.5 - 7 Hz band, and 8192 cores on the 2.5 -
795 10 Hz. We see in Table 2 that the elapsed time for the incident and adjoint field computation
796 (plus the cost of recomputing the incident field backward in time, see Yang et al. (2016))
797 remains almost constant along the frequency bands thanks to the OpenMP acceleration. We
798 also see that the computational time for the Gaussian filter is always negligible, less than
799 1% of the total time for the gradient computation. In counterpart, the computational time
800 for the anisotropic diffusion filter, despite we use its linear version, is non negligible, and
801 reaches almost 25 % of the computation time for one gradient in the last frequency band.
802 This is due to the fact that the anisotropic diffusion filtering is performed on a single core
803 and does not take advantage of the OpenMP acceleration. One shall keep in mind that the
804 computational complexity of the diffusion process is roughly in $O(N^3)$ with N being the
805 number of discrete points in one direction of space, while the complexity for solving the
806 wave equation is in $O(N^4)$. This favors a better ratio between the time spent in filtering
807 and the time spent in computing the wavefields at higher frequency. This is the reason
808 why despite we lack the OpenMP parallelization for the anisotropic diffusion filter the in-
809 crease in computational cost is maintained at a reasonable fraction of the global gradient
810 computation time.

811

[Table 2 about here.]

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

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

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

840 On a side note regarding computing efficiency, we would like to mention that a porting
841 of the coherence enhancing diffusion filter on GPU architecture has been performed. It is
842 made available as a standalone open-source package¹. This work differs from previous im-
843 plementations of similar PDEs by the use of wide adaptive stencils, which allows to handle
844 arbitrary anisotropy while preserving solution positivity, but could be expected to be less
845 suited to GPUs due to data non-locality. Two points raise specific implementation difficul-
846 ties. First, the computation of the eigenvalues of the structure tensors, for which a custom
847 method was used, since the standard libraries were found to have excessively high memory
848 usage and to be significantly slower. Second, the time step iterations, for which a custom
849 kernel using atomic-addition operations is used, rather than the standard sparse matrix li-
850 braries, for the same reasons. Other routines could be used as is from the standard libraries,
851 such as the convolutions involved in the structure tensor construction, or are embarrassingly
852 parallel hence could be implemented in a straightforward manner, such as a key ingredient
853 of the chosen numerical method known as Selling's matrix decomposition. Filtering the

¹<https://github.com/Mirebeau/AdaptiveGridDiscretizations>

854 gradient for the field data application in the 2.5 - 10 Hz frequency band ($131 \times 259 \times 459$
855 ≈ 15 million points) was run on a laptop equipped with an NVIDIA RTX 4000 GPU. After
856 a preprocessing step of 0.0046 s (construction of the diffusion tensors and of the scheme
857 coefficients), a sequence of 20 time steps takes 0.179 s. This is a speed up by two orders of
858 magnitude over the CPU implementation, which opens promising perspectives, despite its
859 current limitation to a specific subtask of the seismic imaging pipeline.

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

CONCLUSION

865 We present in this study a novel filtering technique for the gradient smoothing step in FWI.
866 This filtering technique is based on the solution of an anisotropic diffusion equation. The
867 diffusion tensor is built automatically from the structure tensor associated with the gradient,
868 using a coherence enhancing technique. Compared to other directional filtering methods,
869 this diffusion based method inherits from the stability properties of a partial-differential-
870 equations based smoothing technique: preservation of the minimum, maximum and mean
871 values, independence to translation and rotations, scale-space property. Compared to pre-
872 vious work on diffusion based smoothing in FWI, what we propose here is a systematic
873 3D extension of the coherence-enhancing technique of Weickert (1998) which provides

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

ACKNOWLEDGMENT

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

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

REFERENCES

- 897 Adriaens, X., L. Métivier, and C. Geuzaine, 2021, A trust-region newton method for
898 frequency-domain full-waveform inversion: First International Meeting for Applied
899 Geoscience & Energy Expanded Abstracts, 757–761.
- 900 Aghamiry, H., A. Gholami, and S. Operto, 2020, Compound regularization of Full-
901 Waveform Inversion for imaging piecewise media: IEEE Transactions on Geoscience
902 and Remote Sensing, **58**, 1192–1204; doi: 10.1109/TGRS.2019.2944464.
- 903 Anagaw, A. Y., and M. D. Sacchi, 2018, Edge-preserving smoothing for simultaneous-
904 source full-waveform inversion model updates in high-contrast velocity models: Geo-
905 physics, **83**, A33–A37.
- 906 Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D.
907 Pugmire, 2016, Global adjoint tomography: first-generation model: Geophysical Journal
908 International, **207**, 1739–1766.
- 909 Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform
910 inversion: Geophysics, **60**, 1457–1473.
- 911 Cance, P., and Y. Capdeville, 2015, Validity of the acoustic approximation for elastic waves
912 in heterogeneous media: Geophysics, **80**, T161–T173; doi: 10.1190/geo2014-0397.1.
- 913 Cao, J., R. Brossier, A. Górszczyk, L. Métivier, and J. Virieux, 2022, 3D multi-parameter
914 full-waveform inversion for ocean-bottom seismic data using an efficient fluid-solid cou-
915 pled spectral-element solver: Geophysical Journal International, **229**, 671–703.
- 916 Capdeville, Y., and J.-J. Marigo, 2013, A non-periodic two scale asymptotic method to take
917 account of rough topographies for 2D elastic wave propagation: Geophysical Journal

918 International, **192**, 163–189.

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

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

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

927 Fehrenbach, J., and J.-M. Mirebeau, 2014, Sparse Non-negative Stencils for Anisotropic
928 Diffusion: Journal of Mathematical Imaging and Vision, **49**, 123–147; doi:
929 10.1007/s10851-013-0446-3.

930 Fichtner, A., and A. Villaseñor, 2015, Crust and upper mantle of the western mediterranean
931 – constraints from full-waveform inversion: Earth and Planetary Science Letters, **428**,
932 52 – 62; doi: <https://doi.org/10.1016/j.epsl.2015.07.038>.

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

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

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

940 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. Métivier, 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 using 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–

987 F25.

988 ———, 2022, On the use of nonlinear anisotropic diffusion filters for seismic imaging using
989 the full waveform: *Inverse Problems*, **38**, 115001; doi: 10.1088/1361-6420/ac8c91.

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

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

993 Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D
994 multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study
995 from the North Sea: *Geophysical Journal International*, **213**, 2037–2059; doi:
996 10.1093/gji/ggy103.

997 Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti,
998 and J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform
999 inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical
1000 transverse isotropic approximation: *Geophysical Journal International*, **202**, 1362–1391.

1001 Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from mul-
1002 tifold ocean bottom seismometers data by frequency-domain full-waveform tomogra-
1003 phy: application to the eastern Nankai trough: *Journal of Geophysical Research*, **111**,
1004 doi:10.1029/2005JB003835.

1005 Peters, B., and F. J. Herrmann, 2017, Constraints versus penalties for edge-preserving full-
1006 waveform inversion: *The Leading Edge*, **36**, 94–100; doi: 10.1190/tle36010094.1.

1007 Pladys, A., R. Brossier, N. Kamath, and L. Métivier, 2022, Robust FWI with graph space
1008 optimal transport: application to 3D OBC Valhall data: *Geophysics*, **87**, 1–76.

1009 Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom

1010 seismometer dataset: *First Break*, **28**, 71–78.

1011 Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I: theory and
1012 verification in a physical scale model: *Geophysics*, **64**, 888–901.

1013 Provenzano, G., R. Brossier, and L. Métivier, 2022, Robust and efficient waveform-based
1014 velocity-model-building by optimal-transport in the pseudotime domain: methodology:
1015 *Geophysics*, **in press**.

1016 Raknes, E. B., B. Arntsen, and W. Weibull, 2015, Three-dimensional elastic full waveform
1017 inversion using seismic data from the sleipner area: *Geophysical Journal International*,
1018 **202**, 1877–1894.

1019 Shen, X., L. Jiang, J. Dellinger, A. Brenders, C. Kumar, M. James, J. Etgen, D. Meaux,
1020 R. Walters, and N. Abdullayev, 2018, High-resolution full-waveform inversion for struc-
1021 tural imaging in exploration: *SEG Technical Program Expanded Abstracts 2018*, 1098–
1022 1102.

1023 Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010,
1024 Full waveform inversion: the next leap forward in imaging at Valhall: *First Break*, **28**,
1025 65–70.

1026 Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging : a strategy for
1027 selecting temporal frequencies: *Geophysics*, **69**, 231–248.

1028 Solano, C. P., and R.-É. Plessix, 2019, Velocity-model building with enhanced shallow
1029 resolution using elastic waveform inversion — an example from onshore oman: *Geo-*
1030 *physics*, **84**, R977–R988; doi: 10.1190/geo2018-0736.1.

1031 Stopin, A., R.-E. Plessix, and S. Al Abri, 2014, Multiparameter waveform inversion of
1032 a large wide-azimuth low-frequency land data set in Oman: *Geophysics*, **79**, WA69–

1033 WA77.

1034 Strong, D., and T. Chan, 2003, Edge-preserving and scale-dependent properties of total
1035 variation regularization: *Inverse Problems*, **19**, S165–S187.

1036 Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical*
1037 *Prospecting*, **56**, 765–790; doi: 10.1111/j.1365-2478.2008.00698.x.

1038 Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation:
1039 *Geophysics*, **49**, 1259–1266.

1040 Thomsen, L. A., 1986, Weak elastic anisotropy: *Geophysics*, **51**, 1954–1966.

1041 Thrastarson, S., D.-P. van Herwaarden, L. Krischer, C. Boehm, M. van Driel, M. Afanasiev,
1042 and A. Fichtner, 2022, Data-adaptive global full-waveform inversion: *Geophysical Jour-*
1043 *nal International*, **230**, 1374–1393; doi: 10.1093/gji/ggac122.

1044 Thurin, J., 2020, Uncertainties estimation in full waveform inversion using ensemble meth-
1045 ods: Theses, Université Grenoble Alpes.

1046 Tikhonov, A., A. Goncharky, V. Stepanov, and A. Yagola, 2013, Numerical methods for
1047 the solution of ill-posed problems: Springer Science & Business Media.

1048 Trinh, P. T., R. Brossier, L. Métivier, J. Virieux, and P. Wellington, 2017, Bessel smoothing
1049 filter for spectral element mesh: *Geophysical Journal International*, **209**, 1489–1512;
1050 doi: 10.1093/gji/ggx103.

1051 Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application
1052 using multicomponent measurements of seismic data collection: *Geophysics*, **79**, R63–
1053 R77.

1054 Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An
1055 introduction to Full Waveform Inversion, *in* *Encyclopedia of Exploration Geophysics*:

- 1056 Society of Exploration Geophysics, R1–1–R1–40.
- 1057 Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration
1058 geophysics: *Geophysics*, **74**, WCC1–WCC26.
- 1059 Wang, P., Z. Zhang, J. Mei, F. Lin, and R. Huang, 2019, Full-waveform inversion for salt:
1060 A coming of age: *The Leading Edge*, **38**, 204–213.
- 1061 Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl,
1062 L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform
1063 inversion: *Geophysics*, **78**, R59–R80.
- 1064 Weickert, J., 1998, *Anisotropic diffusion in image processing*, treubner verlag, stuttgart:
1065 Treubner Verlag.
- 1066 Wu, R. S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography
1067 applied to seismic imaging: *Geophysics*, **52**, 11–25.
- 1068 Xu, S., D. Wang, F. Chen, G. Lambaré, and Y. Zhang, 2012, Inversion on re-
1069 flected seismic wave: *SEG Technical Program Expanded Abstracts 2012*, 1–7; doi:
1070 10.1190/segam2012-1473.1.
- 1071 Xue, Z., Z. Zhang, F. Lin, J. Mei, R. Huang, and P. Wang, 2020, Full-waveform inversion
1072 for sparse obn data: *SEG Technical Program Expanded Abstracts 2020*, 686–690.
- 1073 Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016, Wavefield reconstruction in attenu-
1074 ating media: A checkpointing-assisted reverse-forward simulation method: *Geophysics*,
1075 **81**, R349–R362; doi: 10.1190/geo2016-0082.1.
- 1076 Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018, A Time-Domain
1077 Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full
1078 Waveform Inversion: *SIAM Journal on Scientific Computing*, **40**, B1101–B1130; doi:

1079 10.1137/17M1126126.

1080 Yuan, H., S. French, P. Cupillard, and B. Romanowicz, 2014, Lithospheric ex-
1081 pression of geological units in central and eastern north america from full wave-
1082 form tomography: *Earth and Planetary Science Letters*, **402**, 176–186; doi:
1083 <https://doi.org/10.1016/j.epsl.2013.11.057>. (Special issue on USArray science).

1084 Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving
1085 & reflected waves for velocity model building with impedance inversion based on scale
1086 separation: *Geophysical Journal International*, **202**, 1535–1554.

1087 Zuberi, M. A., and R. G. Pratt, 2017, Mitigating nonlinearity in full waveform inversion
1088 using scaled-sobolev pre-conditioning: *Geophysical Journal International*, **213**, 706–
1089 725; doi: 10.1093/gji/ggx549.

LIST OF FIGURES

1090	1	Threshold function $g(x)$ depending on the constant C for $\alpha = 10^{-2}$	61
1091	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.	62
1092			
1093			
1094	3		(a) Left most and (b) central shot gather computed using the models and the source wavelet presented in Figure 2.
1095			
1096	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.	64
1097			
1098			
1099			
1100			
1101			
1102	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) $r_x = r_z = 0.5$	65
1103			
1104			
1105			
1106			
1107			
1108	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.	66
1109			
1110			
1111			
1112			
1113			
1114	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.	67
1115			
1116			
1117			
1118			
1119			
1120	8	Difference in percentage (following equation 21) between a gradient obtained after <i>nonlinear</i> anisotropic diffusion and a gradient after <i>linear</i> 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 %.	68
1121			
1122			
1123			
1124			
1125	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.	69
1126			
1127			
1128			

1129	10	Location of the Valhall field on the North Sea (from Thurin (2020)).	70
1130	11	Layout of the acquisition overlapped on an horizontal P-wave velocity	
1131		slice at 1 km depth obtained by FWI. Location of sources (gray dots)	
1132		and receivers (blue diamonds). Two receivers positions (A and B) are lo-	
1133		cated with black stars. Cables A ($x = 2950$ m), B ($x = 5530$ m) and	
1134		C ($x = 3080$ m) are identified. Black dots denote the position of three	
1135		available P-wave velocity sonic-logs.	71
1136	12	(a) CSG aligned with the positions of cable A before muting the Schölte.	
1137		(b) CSG aligned with the positions of cable B before muting the Schölte	
1138		waves. (c) CSG aligned with the positions of cable A after muting the	
1139		Schölte waves. (d) CSG aligned with the positions of cable B after muting	
1140		the Schölte waves. We reproduce the data anatomy analysis presented in	
1141		Operto et al. (2015). The red, white, black arrows point on the reflection	
1142		from a shallow reflector, the top of the low velocity anomaly and the top	
1143		of the reservoir, respectively. The solid arrow points on the pre-critical	
1144		reflections, while the dashed ones points on the post-critical reflections. . . .	72
1145	13	Initial v_P model. (a) Constant depth slices at, from top to bottom: $z = 175$	
1146		m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x	
1147		slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity	
1148		anomaly), and $x = 3.95$ km (nearby its periphery). (c) Constant y slices at,	
1149		from top to bottom: $y = 9$ km and $y = 6$ km.	73
1150	14	Initial density model. (a) Constant depth slices at, from top to bottom: $z =$	
1151		175 m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant	
1152		x slices at, from top to bottom: $x = 2.95$ km (passing through the low	
1153		velocity anomaly), and $x = 3.95$ km (nearby its periphery). (c) Constant y	
1154		slices at, from top to bottom: $y = 9$ km and $y = 6$ km.	74
1155	15	Initial anisotropy model η . (a) Constant depth slices at, from top to bottom:	
1156		$z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b)	
1157		Constant x slices at, from top to bottom: $x = 2.95$ km (passing through	
1158		the low velocity anomaly), and $x = 3.95$ km (nearby its periphery). (c)	
1159		Constant y slices at, from top to bottom: $y = 9$ km and $y = 6$ km.	75
1160	16	Final v_P model with non-stationary Gaussian smoothing. (a) Constant	
1161		depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in	
1162		black and white colorscale. (b) Constant x slices at, from top to bottom: x	
1163		$= 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$ km	
1164		(nearby its periphery). (c) Constant y slices at, from top to bottom: $y = 9$	
1165		km and $y = 6$ km.	76

1166	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.	77
1167			
1168			
1169			
1170			
1171			
1172	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.	78
1173			
1174			
1175			
1176			
1177			
1178	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.	79
1179			
1180			
1181			
1182			
1183			
1184	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.	80
1185			
1186			
1187			
1188			
1189	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.	81
1190			
1191			
1192			
1193			
1194	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.	82
1195			
1196			
1197			
1198	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.	83
1199			
1200			
1201			
1202	24	3D view of the velocity model (top), density model (middle), reflectivity image (bottom) obtained by FWI with the nonstationary Gaussian filter. . .	84
1203			
1204	25	3D view of the velocity model (top), density model (middle), reflectivity image (bottom) obtained by FWI with the anisotropic diffusion filter.	85
1205			

1206	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.	86
1207			
1208	27	Data fit in mirror display for cable A. (a) Data fit in the initial model. (b)	
1209		Data fit in the final model obtained using a Gaussian filter. (c) Data fit in	
1210		the final model obtained using an anisotropic diffusion filter.	87
1211	28	Data fit in mirror display for cable B. (a) Data fit in the initial model. (b)	
1212		Data fit in the final model obtained using a Gaussian filter. (c) Data fit in	
1213		the final model obtained using an anisotropic diffusion filter.	88
1214	29	Comparison of the data fit by superposition of the field data in black/white	
1215		colourscale and the predicted data in a red/blue colourscale. A good match is	
1216		indicated by the absence of white and red color. (a,b) Data fit in the initial	
1217		model for (a) cable A and (b) cable B. (c,d) Data fit in the FWI model	
1218		obtained using a Gaussian filter for (c) cable A and (d) cable B. (e,f) Data	
1219		fit in the FWI model obtained using an anisotropic diffusion filter for (e)	
1220		cable A and (f) cable B. The red, white, black arrows point on the reflection	
1221		from a shallow reflector, the top of the low velocity anomaly and the top of	
1222		the structure below, respectively. The solid arrow points on the pre-critical	
1223		reflections, while the dashed ones points on the post-critical reflections. . . .	89
1224	30	Comparison between sonic logs and velocity model values at the location of	
1225		the wells where the sonic logs have been extracted. The sonic logs appear	
1226		in solid black line. The initial model is in dotted-red line. The FWI model	
1227		using a Gaussian filter is in solid orange line. The FWI model using an	
1228		anisotropic diffusion filter is in solid purple line.	90

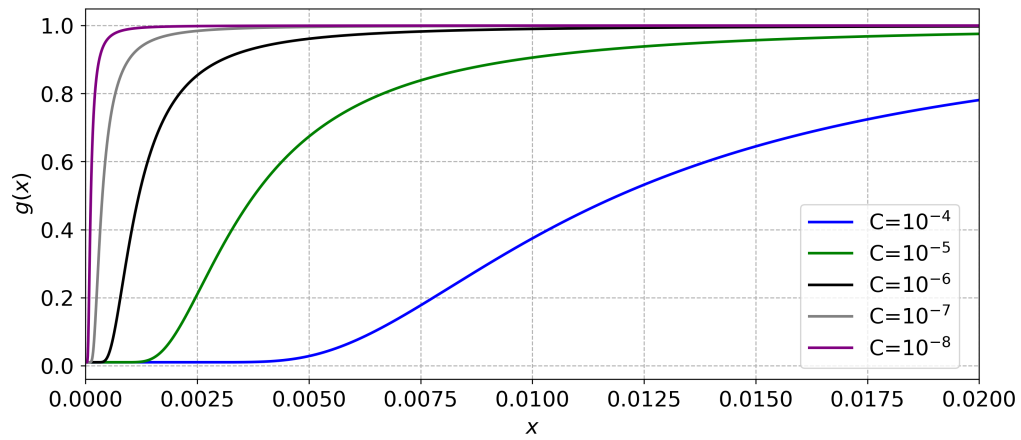


Figure 1: Threshold function $g(x)$ depending on the constant C for $\alpha = 10^{-2}$.

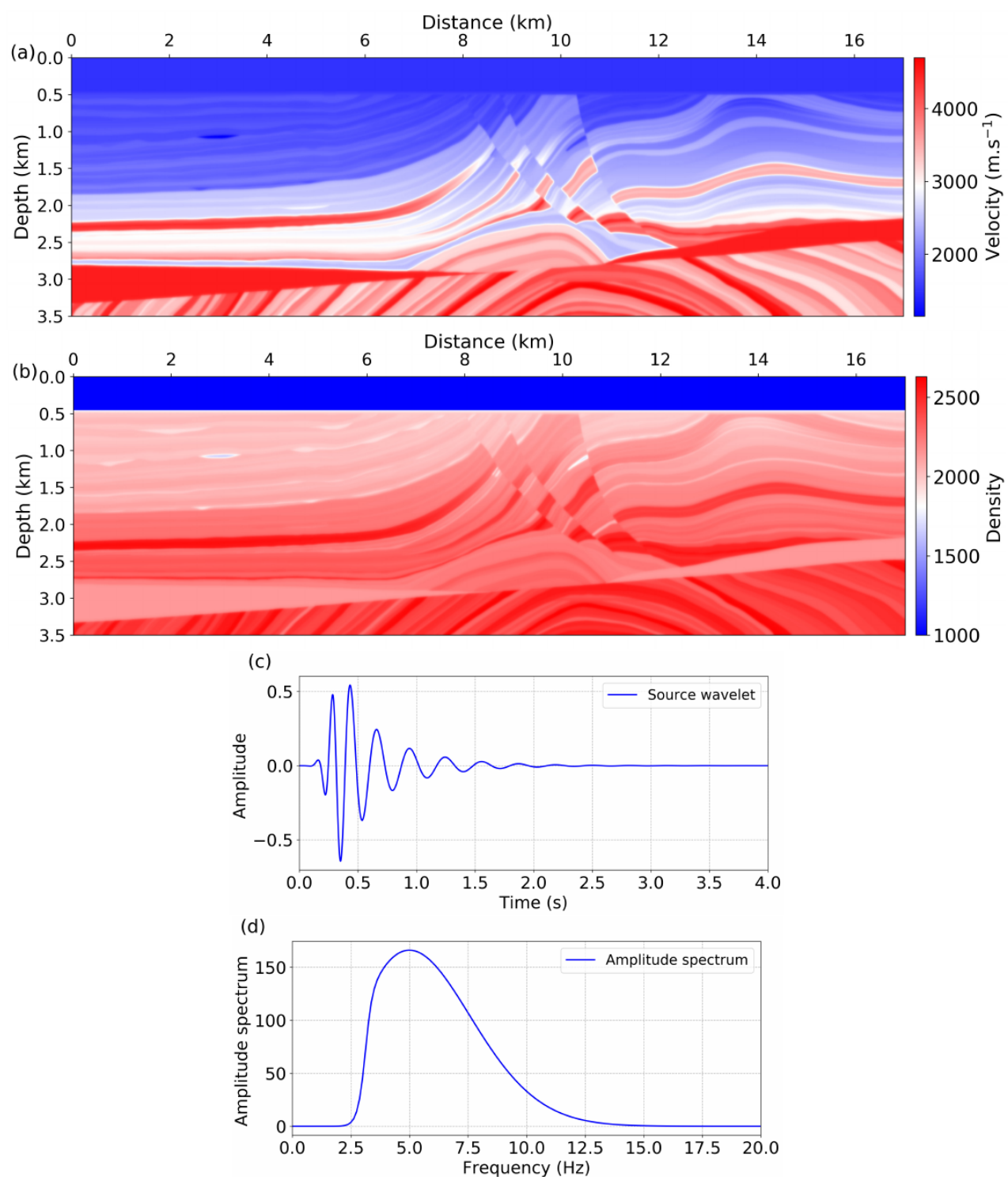


Figure 2: (a) Velocity and (b) density 5 m grid models used as reference to build the data for the Marmousi II synthetic experiment. (c) Source wavelet used to generate the data using these models and (d) its spectrum.

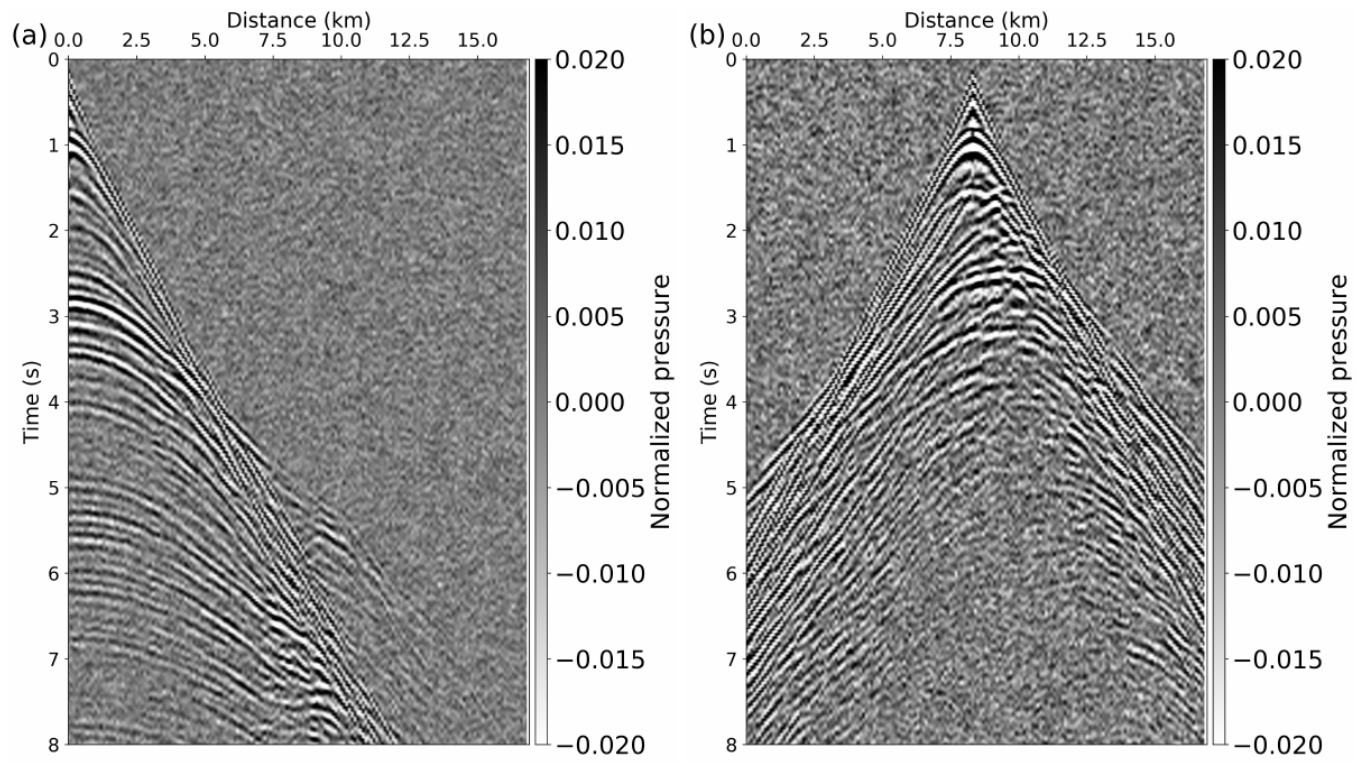


Figure 3: (a) Left most and (b) central shot gather computed using the models and the source wavelet presented in Figure 2.

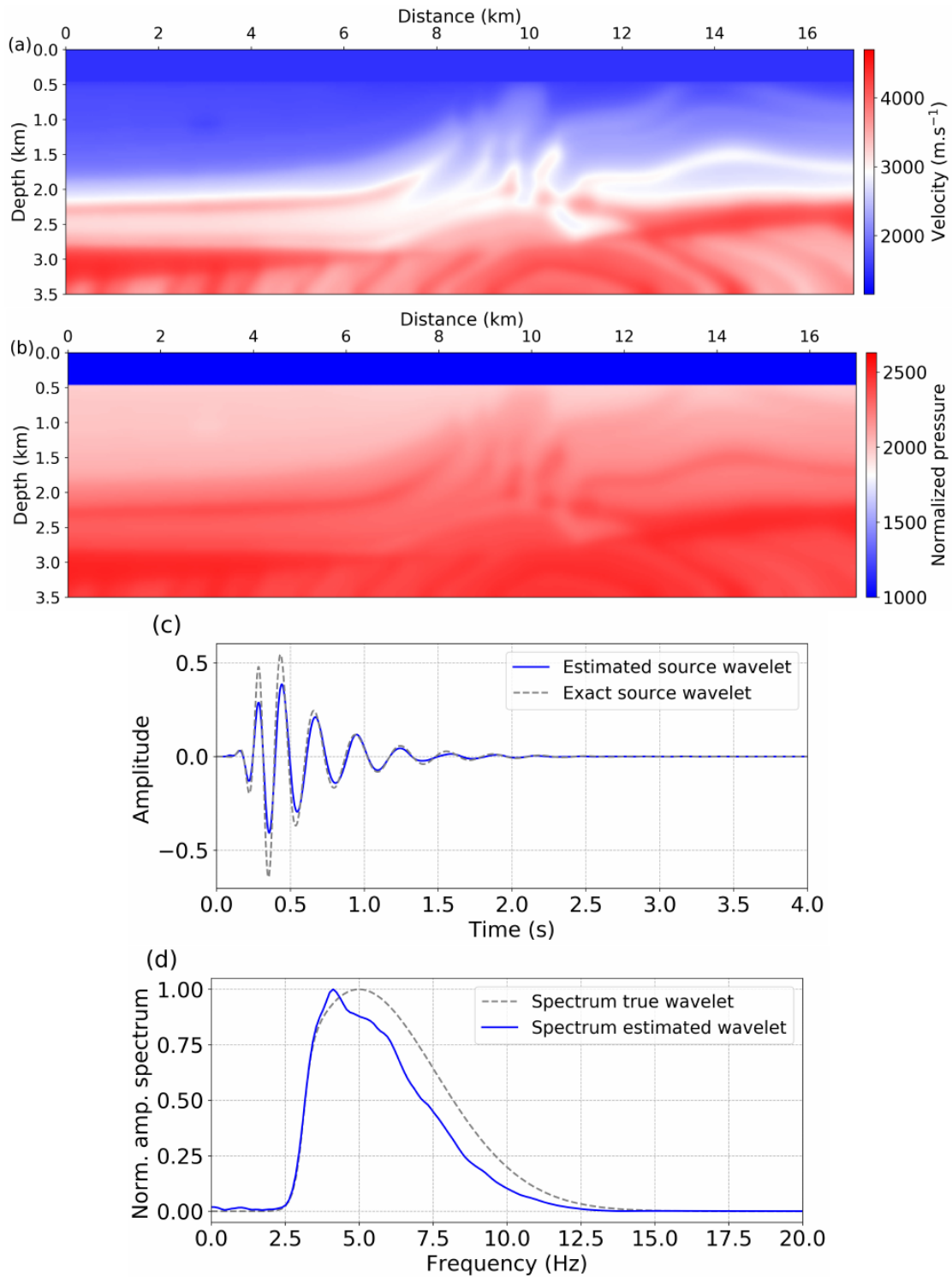


Figure 4: (a) Initial velocity and (b) initial density. (c) Estimated source wavelet compared to the true source wavelet. (d) Comparison of the corresponding two amplitude spectra. Due to the inaccuracy of the initial velocity and density models and the noise on the data, the estimated wavelet exhibits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

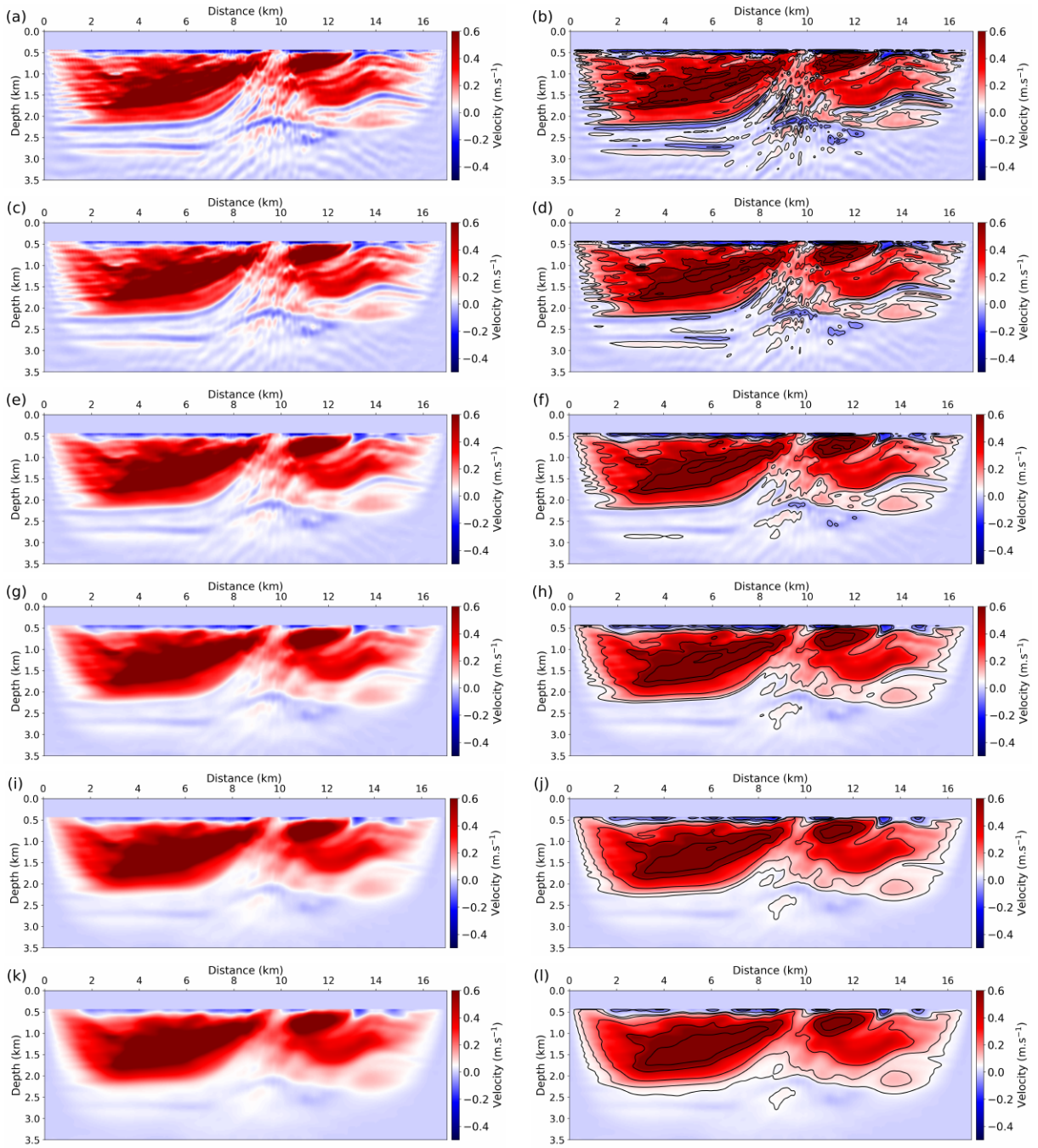


Figure 5: Evolution of the gradient for different Gaussian smoothing with $r_x = r_z$ from 0.1 to 0.5. On the left column the gradient alone is presented, while on the right column, the level set of the gradient are superimposed to better delineate its main structures. Gradient with (a,b) no smoothing, (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) $r_x = r_z = 0.5$.

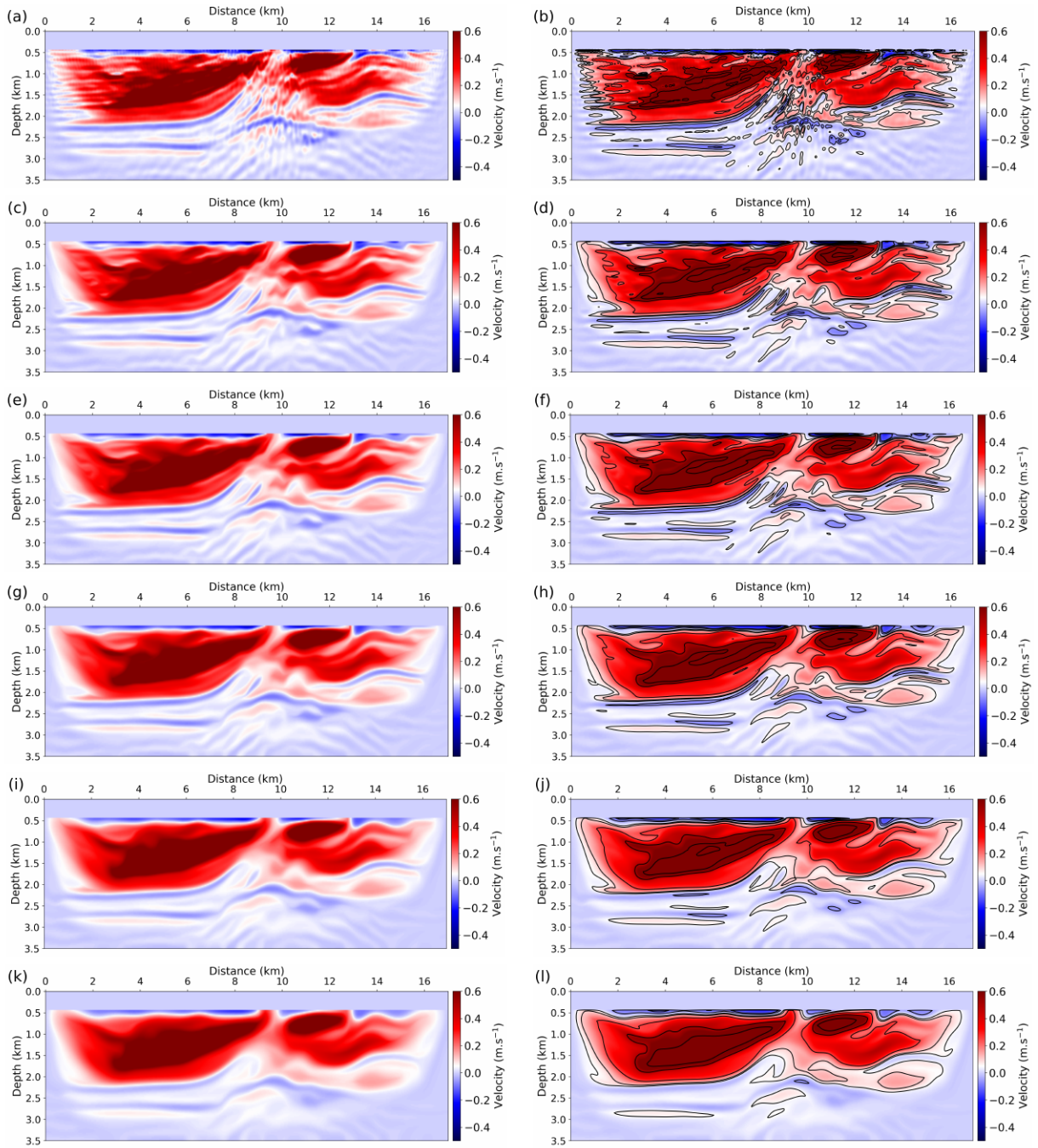


Figure 6: Evolution of the gradient along the nonlinear anisotropic diffusion iterations using the stencil from Fehrenbach and Mirebeau (2014). On the left column the gradient alone is presented, while on the right column, the level set of the gradient are superimposed to better delineate its main structures. 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.

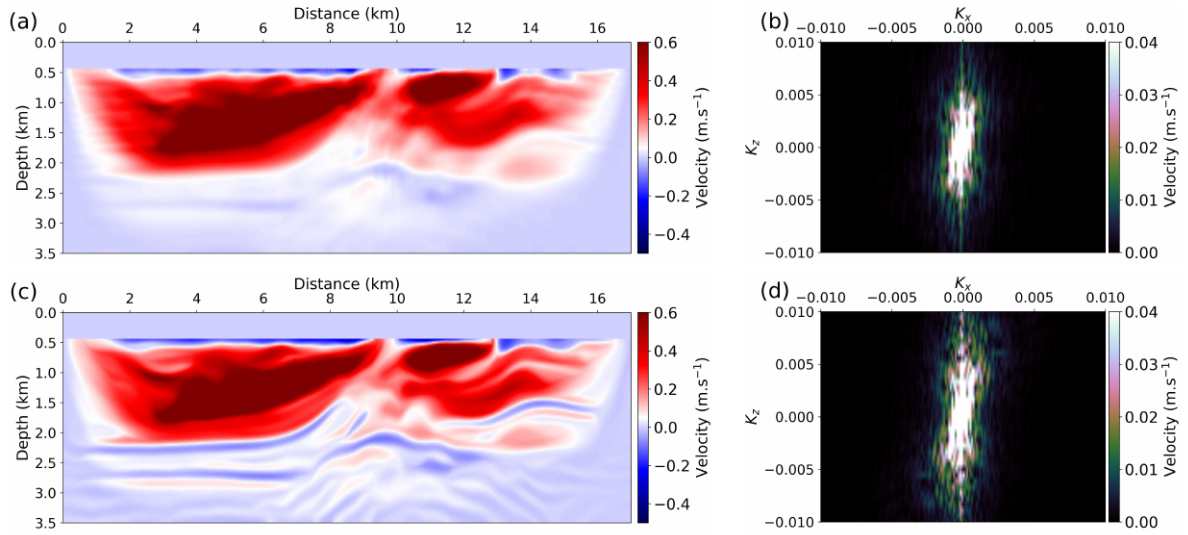


Figure 7: Comparison between the initial gradient obtained after (a,b) a Gaussian smoothing (c,d) 200 iterations of nonlinear anisotropic diffusion using the stencil from Fehrenbach and Mirebeau (2014). In (a) and (c) the gradients are presented, while in (b) and (d) their corresponding k_x, k_z spectrum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

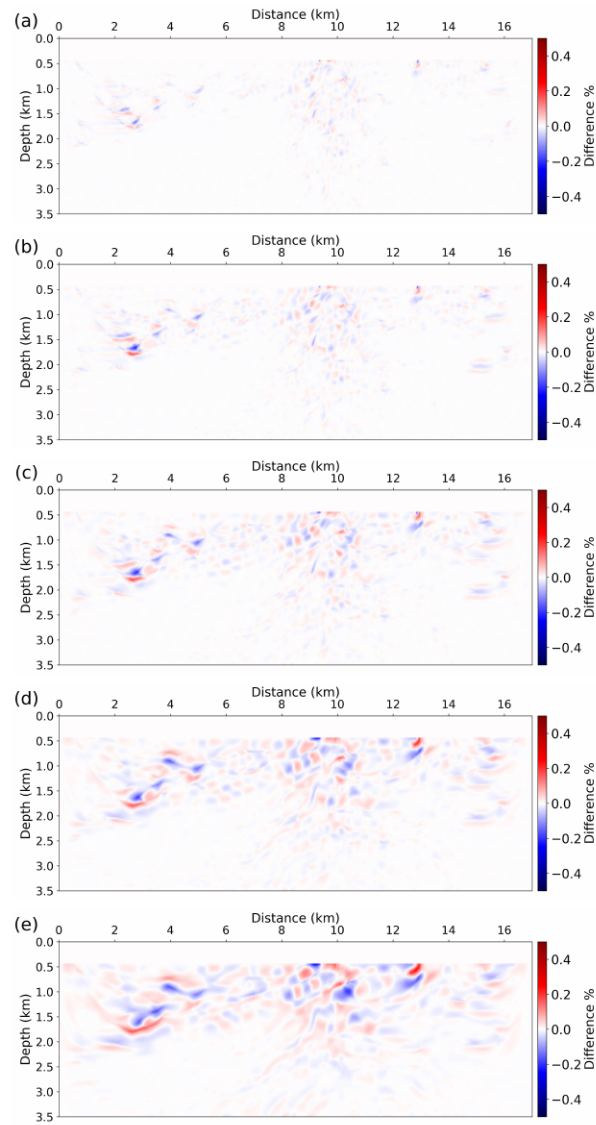


Figure 8: Difference in percentage (following equation 21) between a gradient obtained after *nonlinear* anisotropic diffusion and a gradient after *linear* anisotropic diffusion. Comparison for (a) 50 diffusion iterations, (b) 100 diffusion iterations, (c) 200 diffusion iterations, (d) 400 diffusion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

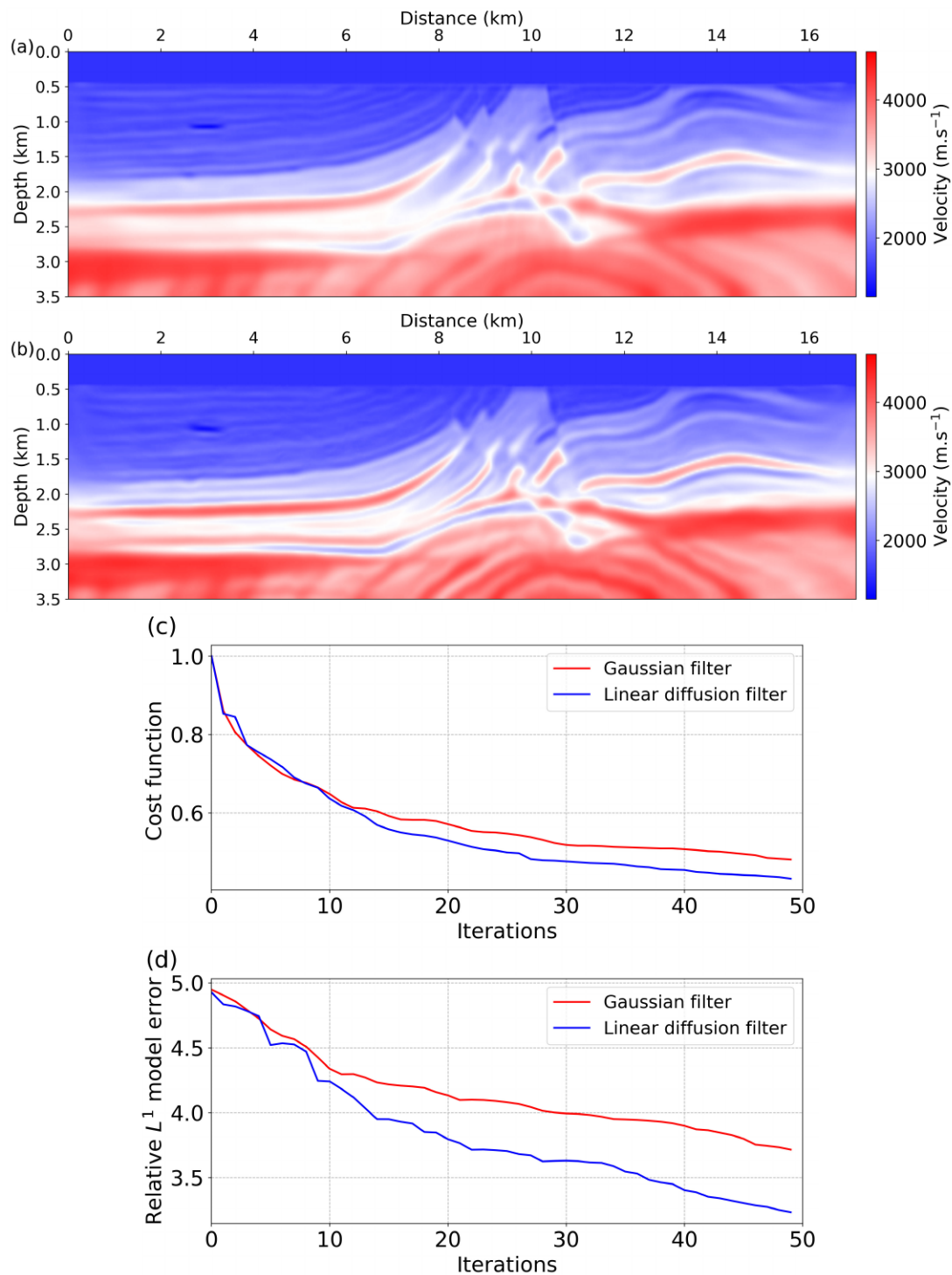


Figure 9: Final P-wave velocity model after 50 FWI iterations using (a) Gaussian smoothing (b) linear anisotropic diffusion smoothing. (c) Model error depending on the iteration number. The use of the linear anisotropic diffusion filter accelerates the convergence towards the exact model.

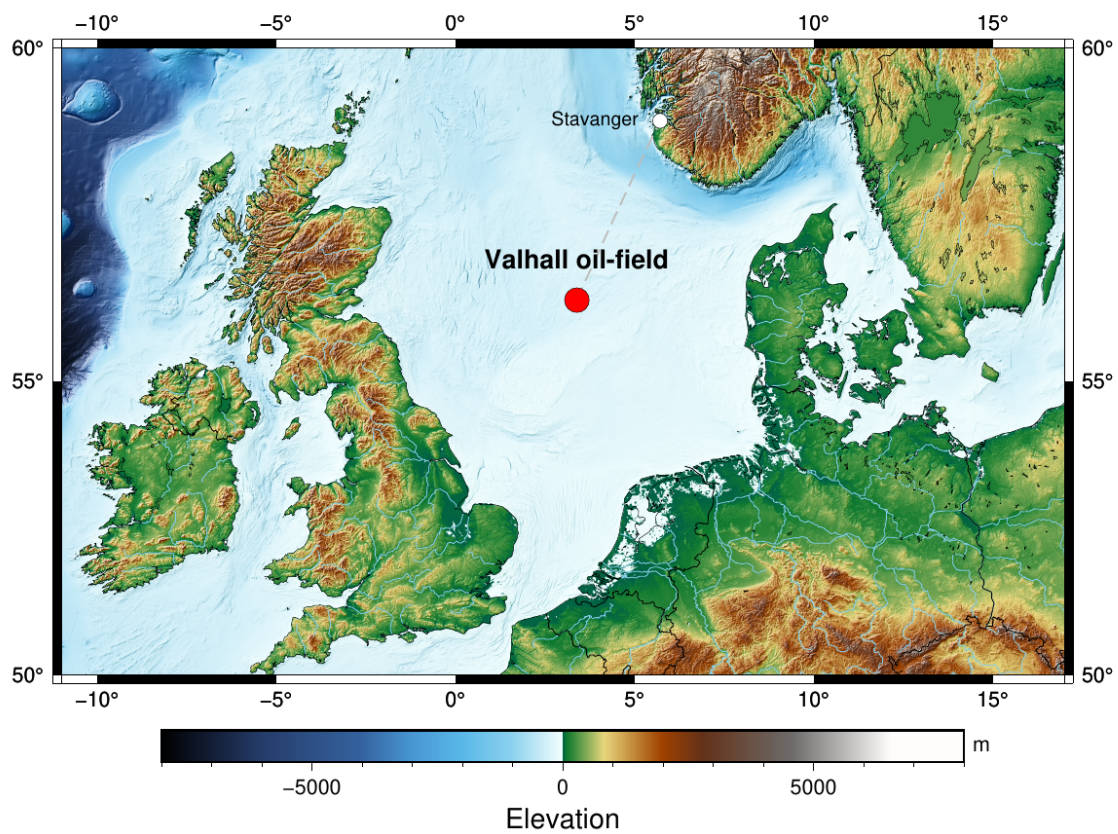


Figure 10: Location of the Valhall field on the North Sea (from Thurin (2020)).

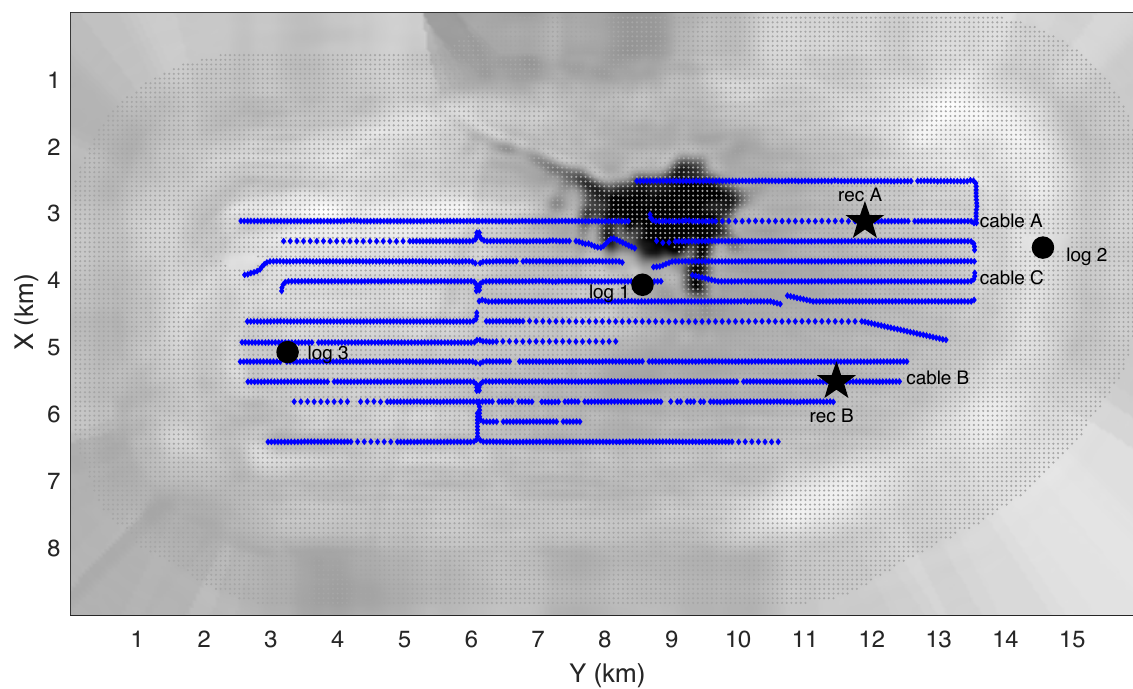


Figure 11: Layout of the acquisition overlapped on an horizontal P-wave velocity slice at 1 km depth obtained by FWI. Location of sources (gray dots) and receivers (blue diamonds). Two receivers positions (A and B) are located with black stars. Cables A ($x = 2950$ m), B ($x = 5530$ m) and C ($x = 3080$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

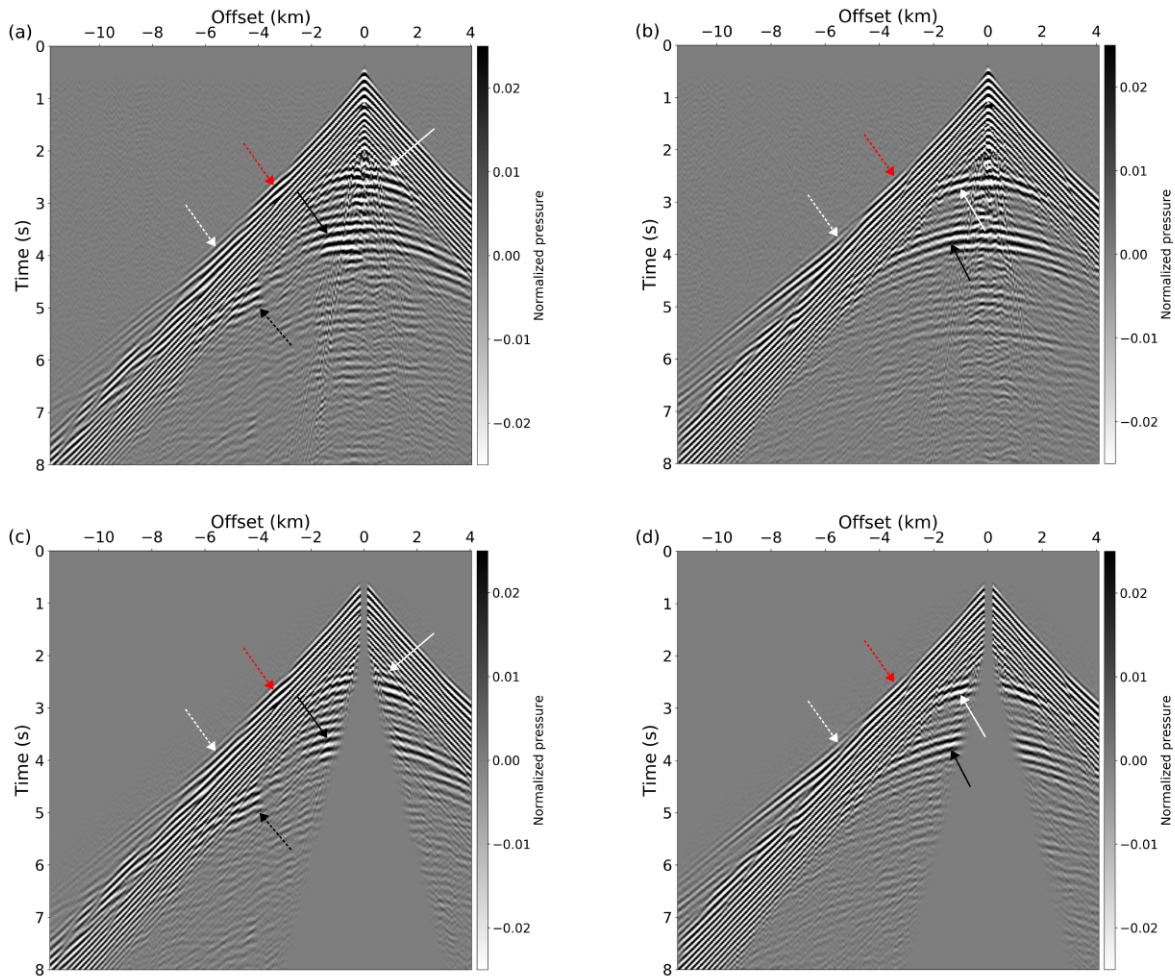


Figure 12: (a) CSG aligned with the positions of cable A before muting the Schölte. (b) CSG aligned with the positions of cable B before muting the Schölte waves. (c) CSG aligned with the positions of cable A after muting the Schölte waves. (d) CSG aligned with the positions of cable B after muting the 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.

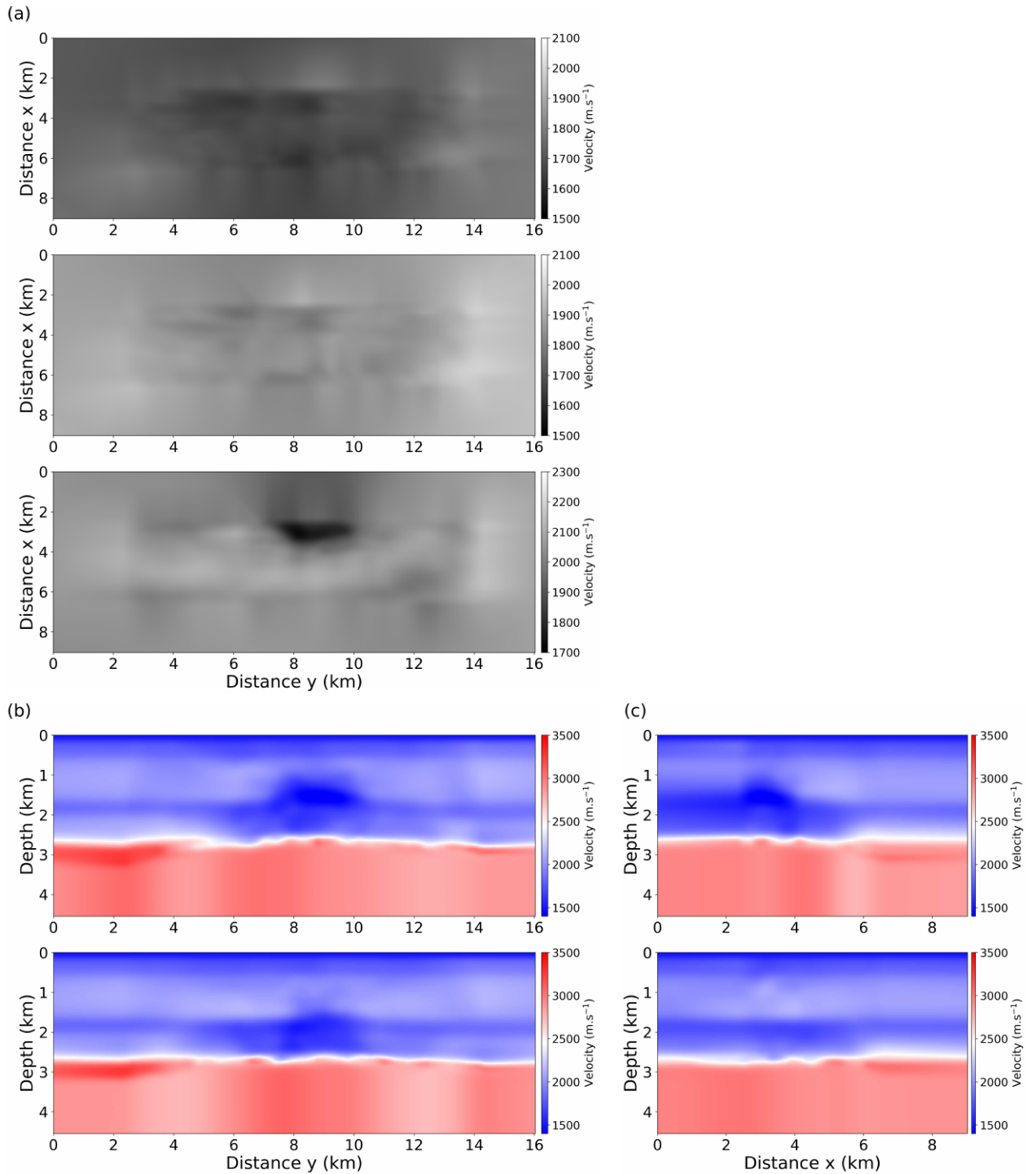


Figure 13: Initial v_P model. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$ km (nearby its periphery). (c)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

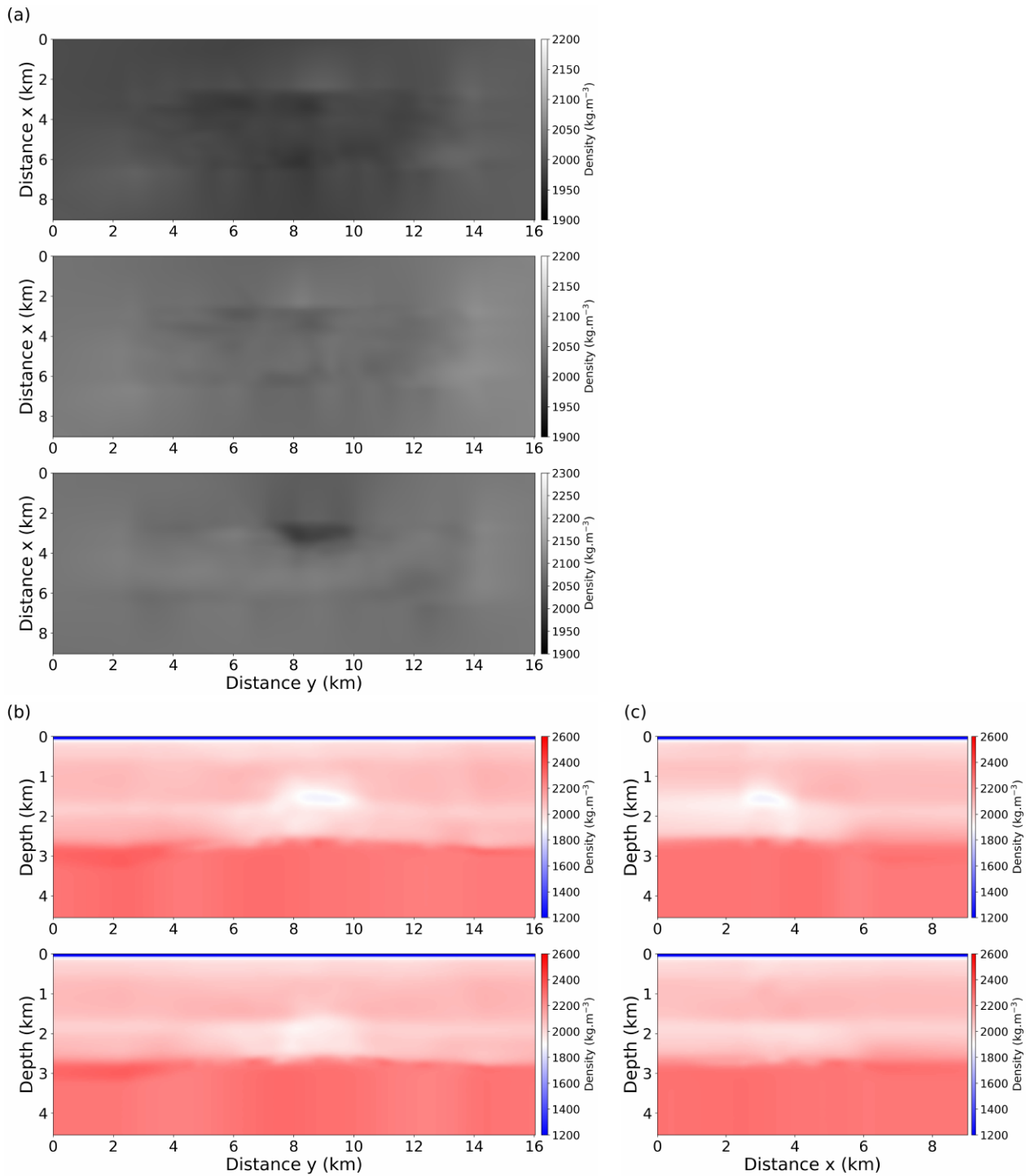


Figure 14: Initial density model. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$ km (nearby its periphery).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

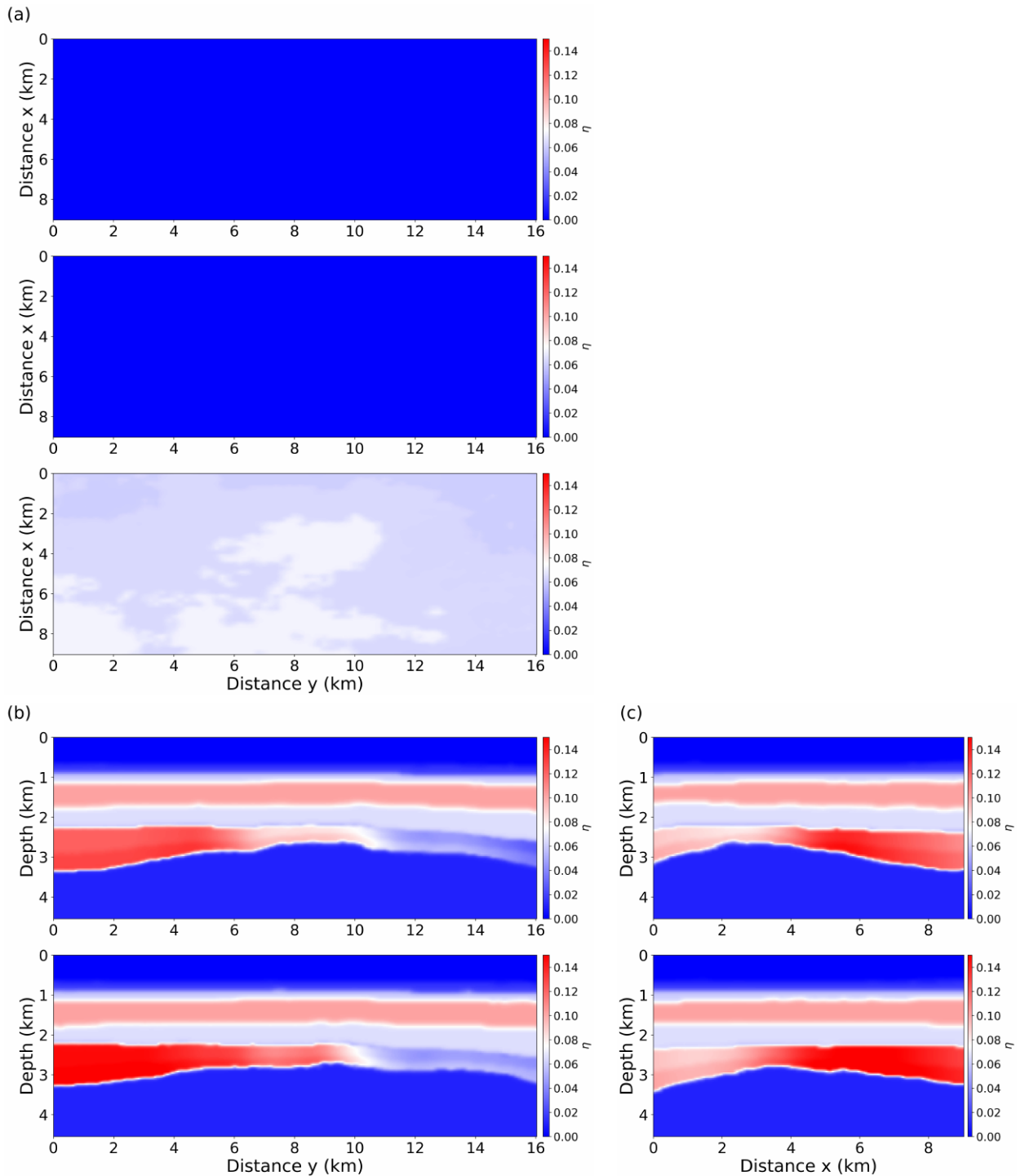


Figure 15: Initial anisotropy model η . (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$ km (nearby its periphery).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

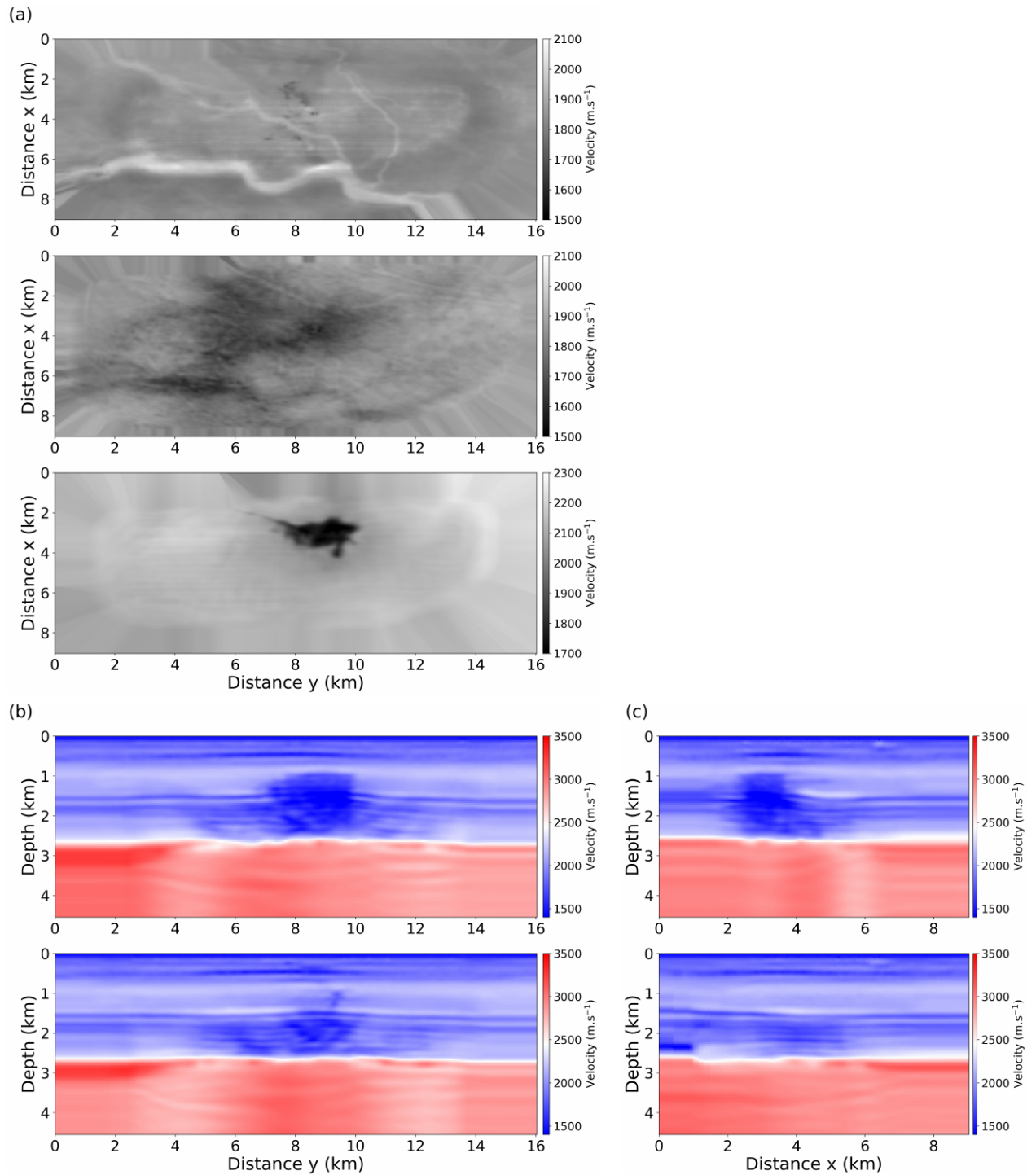


Figure 16: Final v_P model with non-stationary Gaussian smoothing. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x =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

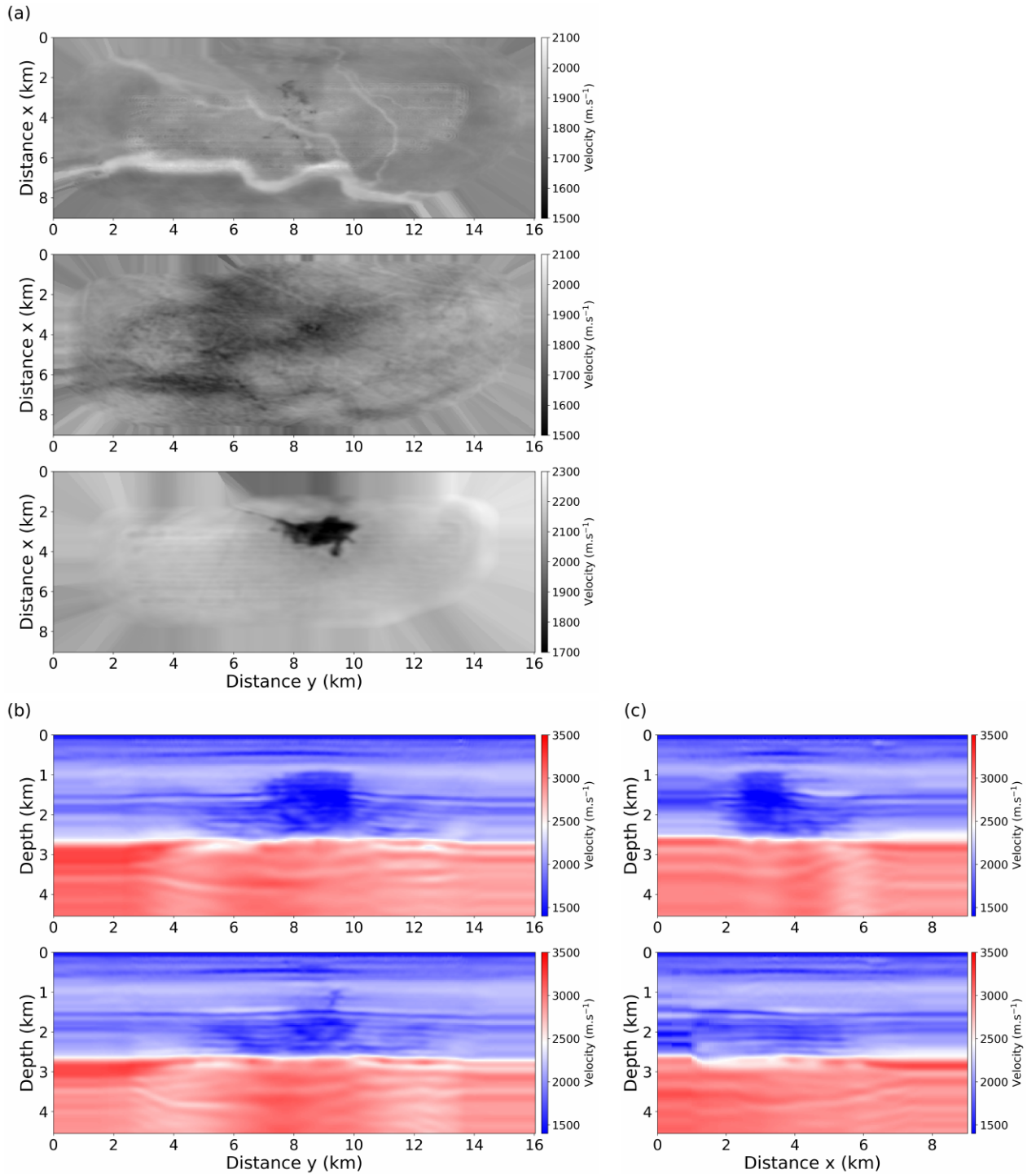


Figure 17: Final v_P model with linear anisotropic diffusion smoothing. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and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

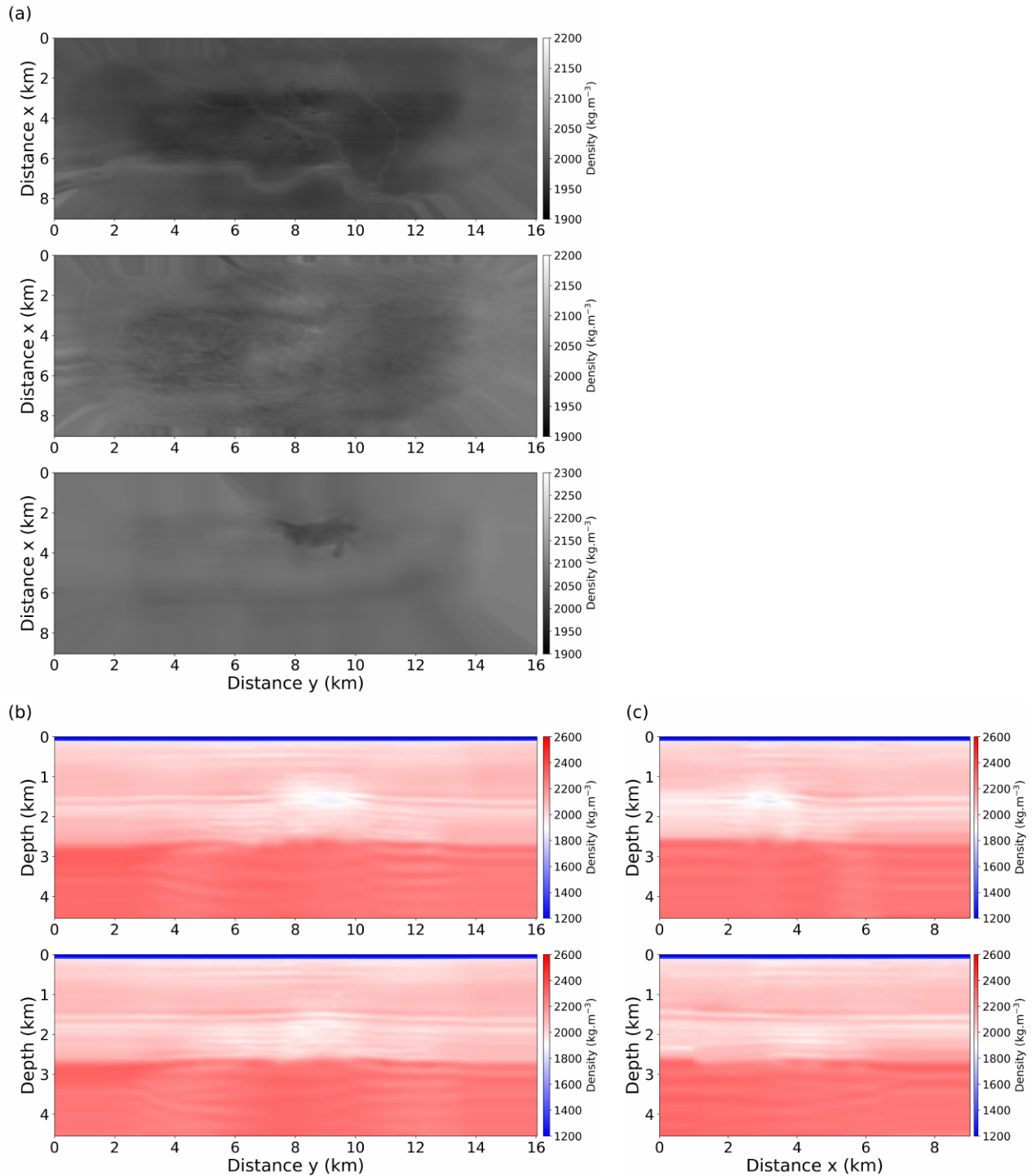


Figure 18: Final ρ model with non-stationary Gaussian smoothing. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$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

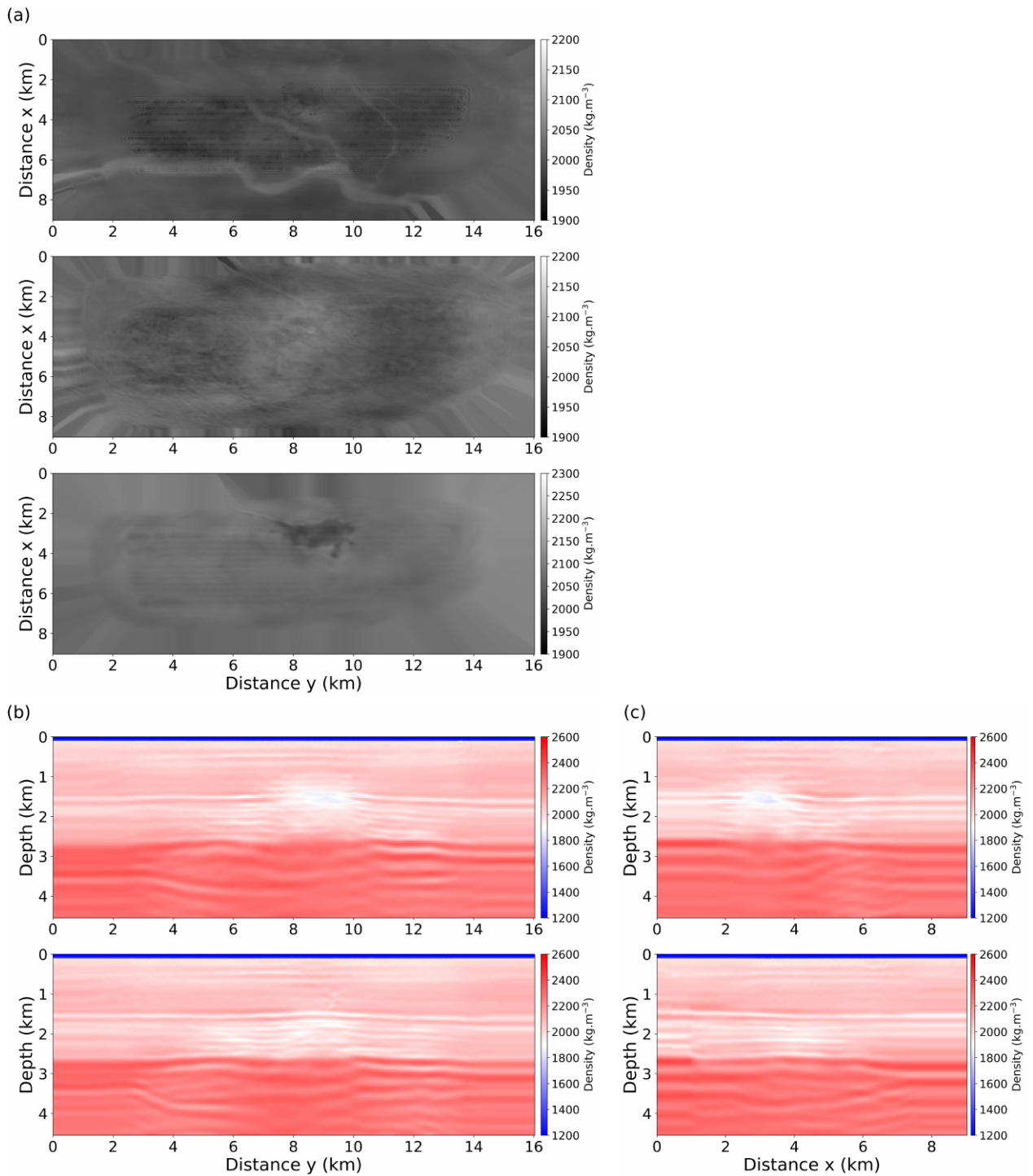


Figure 19: Final ρ model with linear anisotropic diffusion smoothing. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m, in black and white colorscale. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x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

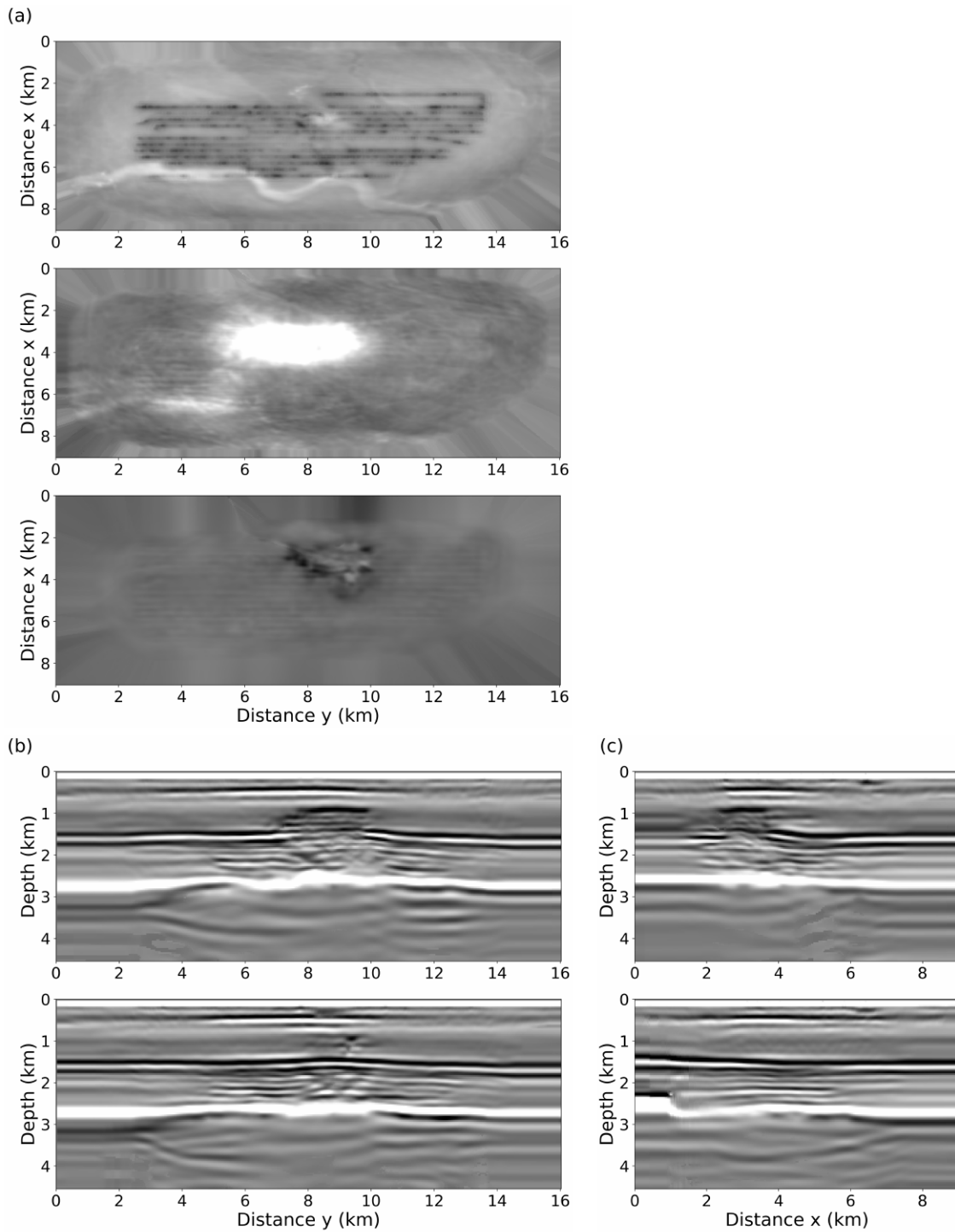


Figure 20: Final “full wavefield images” with Gaussian smoothing. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$ km (nearby its periphery).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

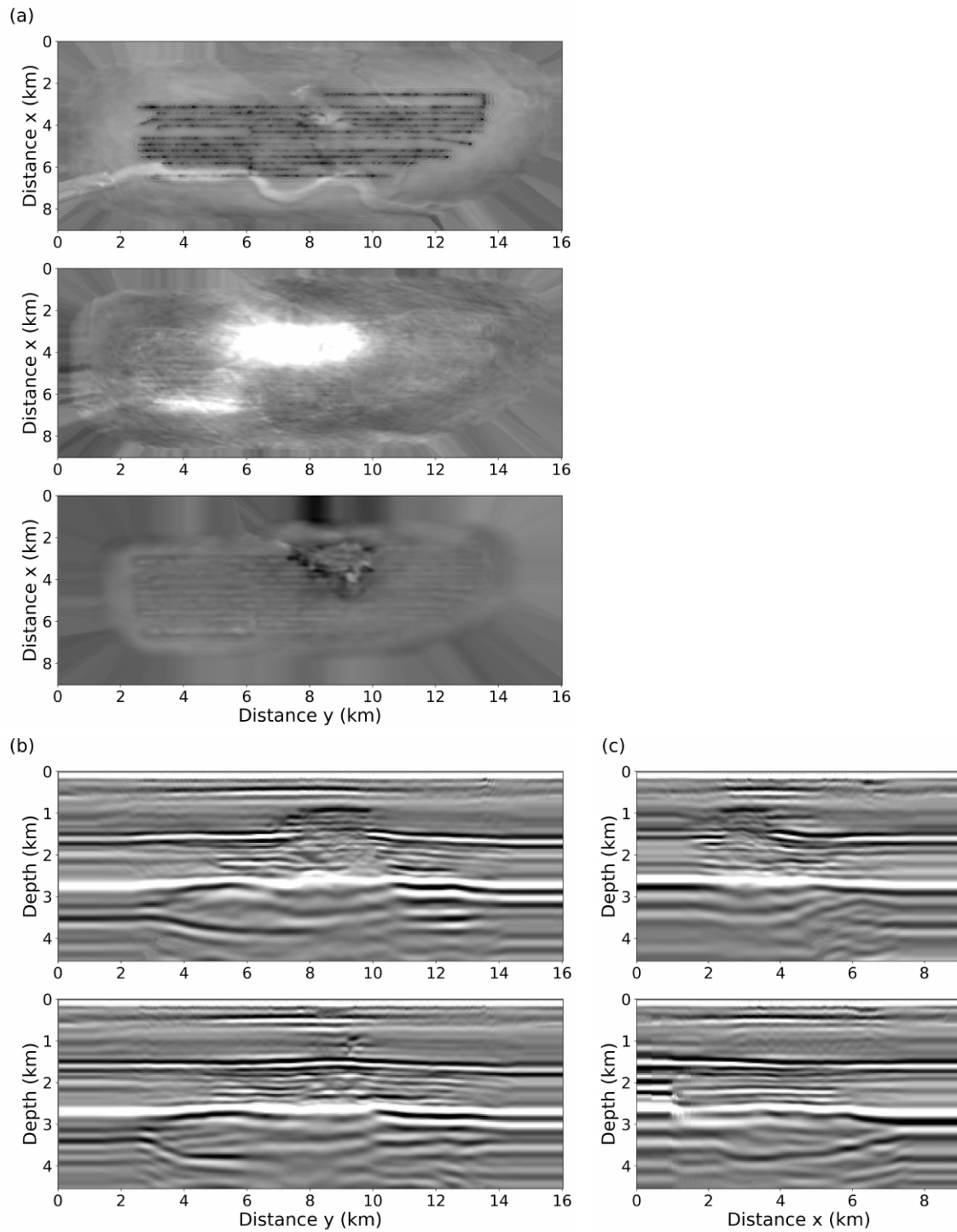


Figure 21: Final “full wavefield images” with linear anisotropic diffusion smoothing. (a) Constant depth slices at, from top to bottom: $z = 175$ m, $z = 500$ m, $z = 1000$ m. (b) Constant x slices at, from top to bottom: $x = 2.95$ km (passing through the low velocity anomaly), and $x = 3.95$ km (nearby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

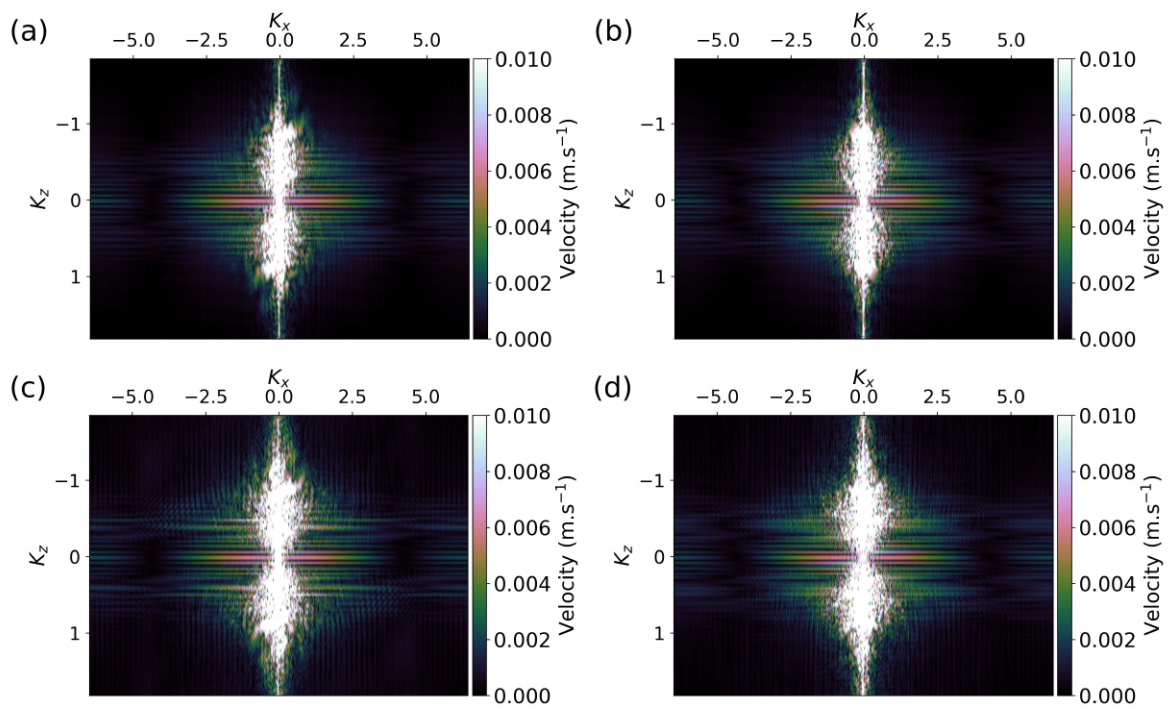


Figure 22: (a,b) Wavenumber spectrum of the reflectivity images depth sections in Figure 20b obtained with a Gaussian filter. (c,d) Wavenumber spectrum of the reflectivity images depth sections in Figure 21b obtained with the coherence enhancing diffusion filter.

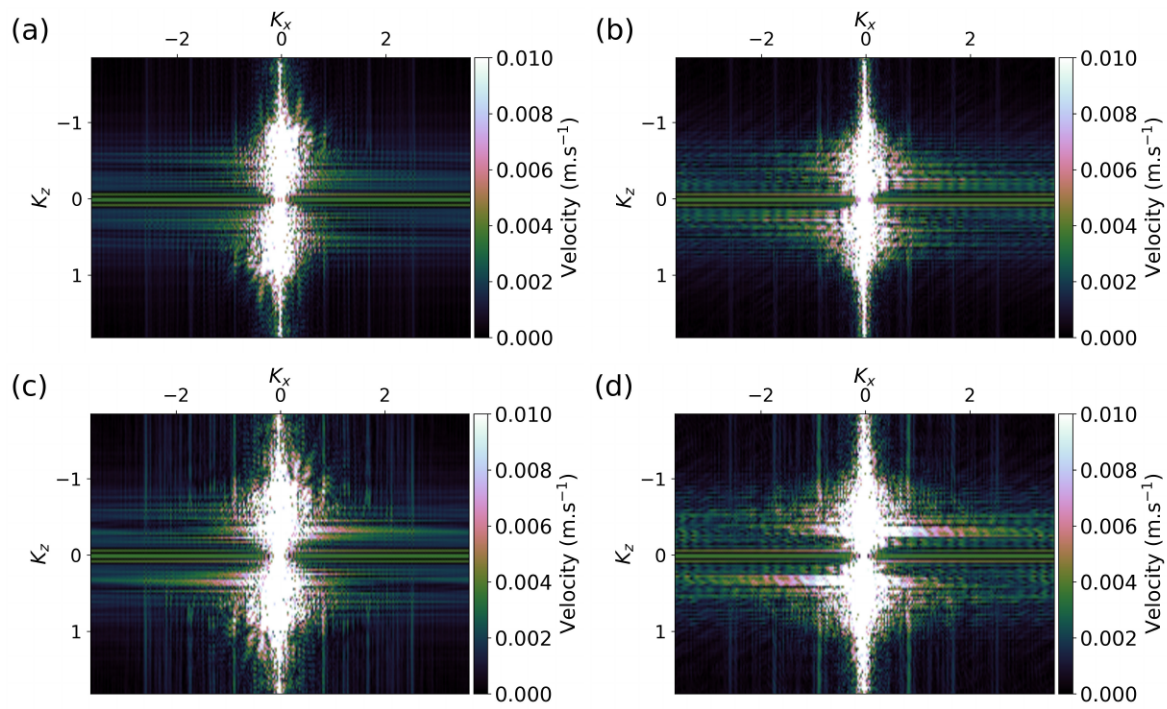


Figure 23: (a,b) Wavenumber spectrum of the reflectivity images depth sections in Figure 20c obtained with a Gaussian filter. (c,d) Wavenumber spectrum of the reflectivity images depth sections in Figure 21c obtained with the coherence enhancing diffusion filter.

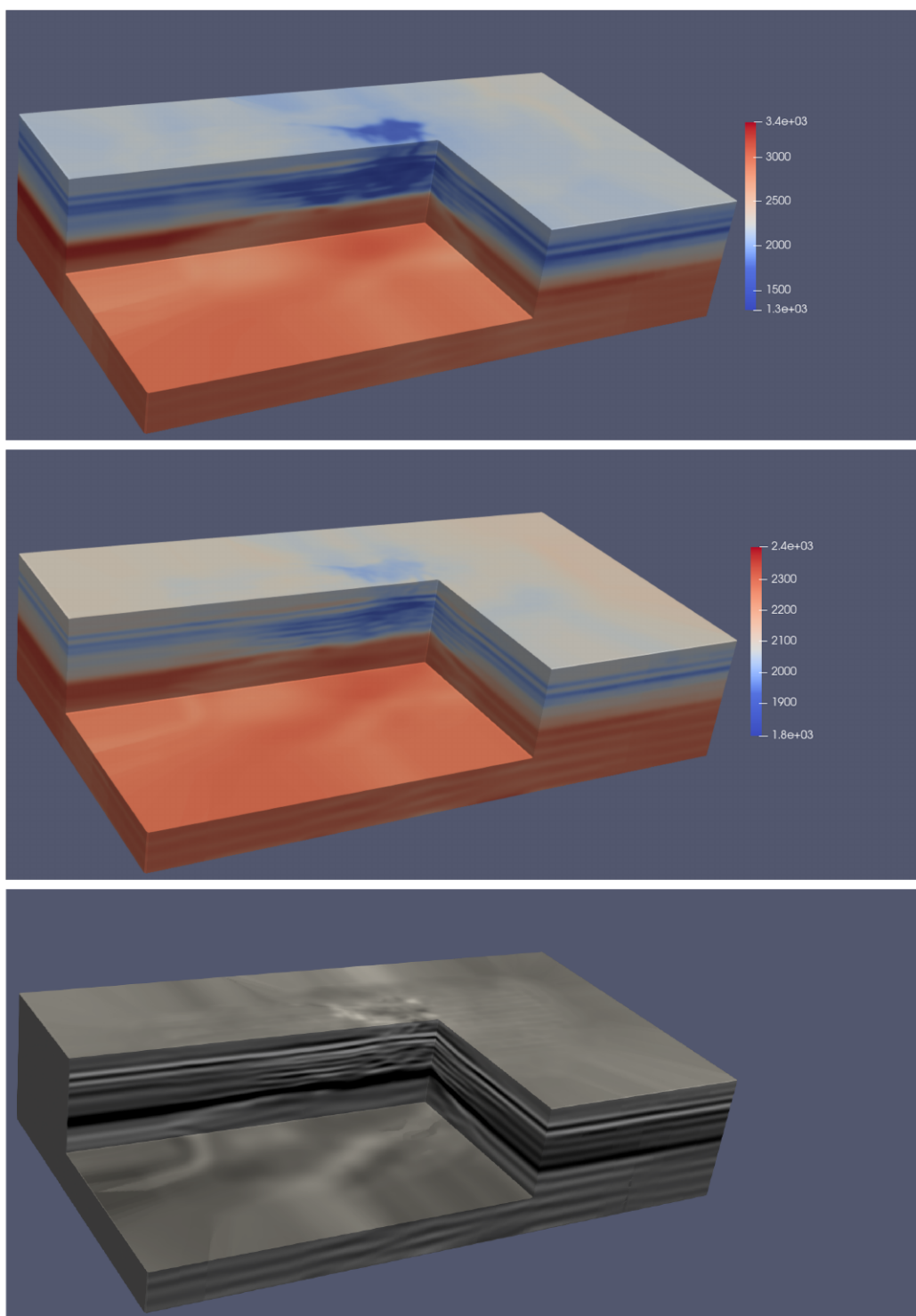


Figure 24: 3D view of the velocity model (top), density model (middle), reflectivity image (bottom) obtained by FWI with the nonstationary Gaussian filter.

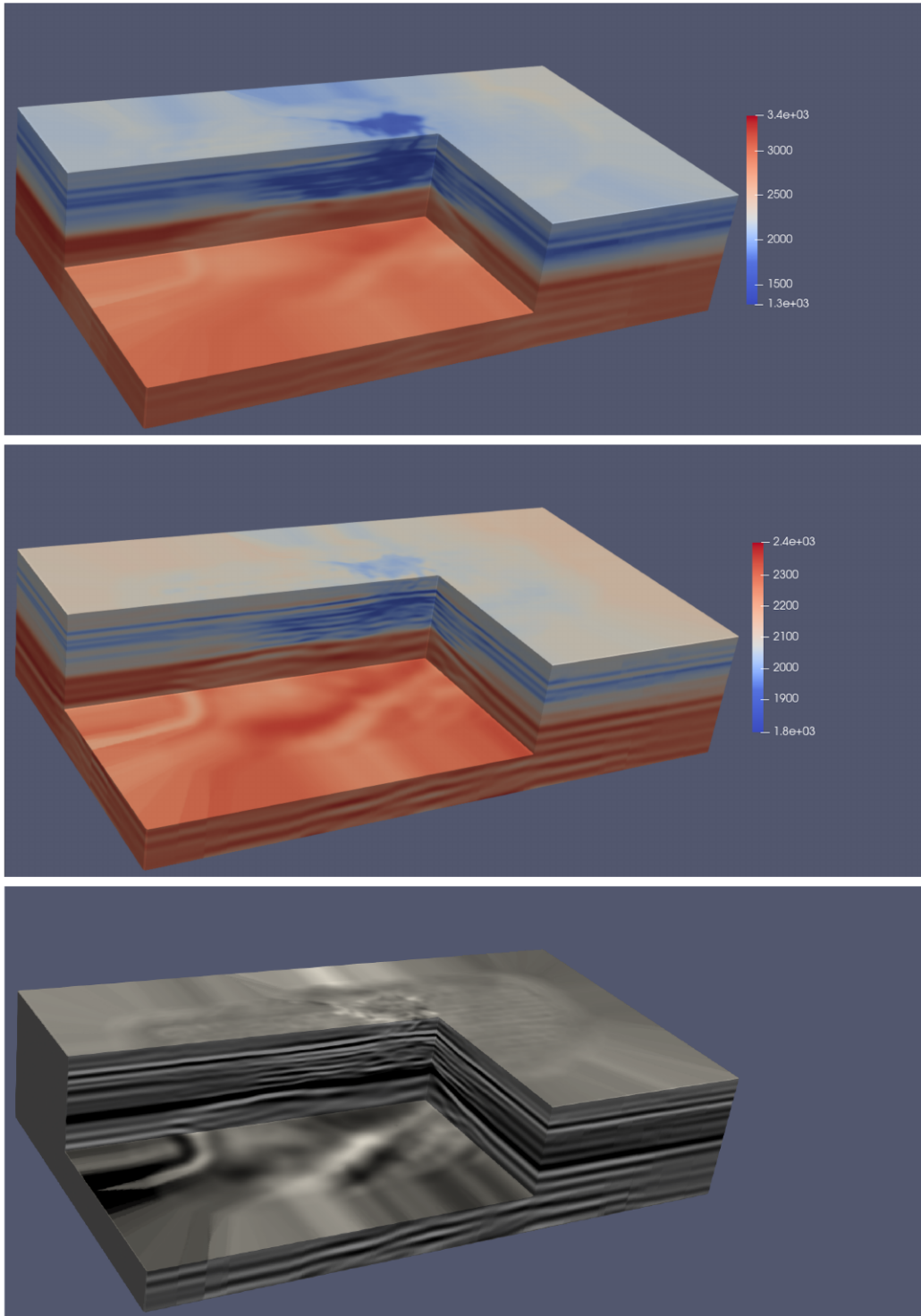


Figure 25: 3D view of the velocity model (top), density model (middle), reflectivity image (bottom) obtained by FWI with the anisotropic diffusion filter.

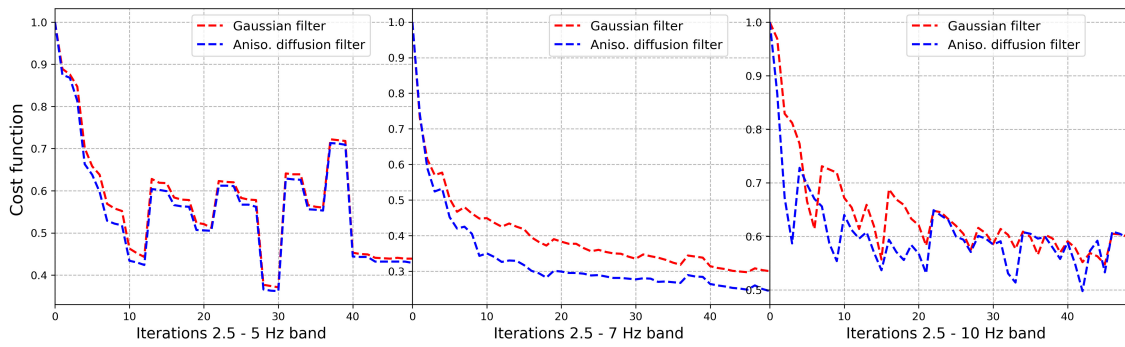


Figure 26: Misfit function decrease along the course of iterations over the three frequency bands 2.5 - 5 Hz, 2.5 - 7 Hz, 2.5 - 10 Hz.

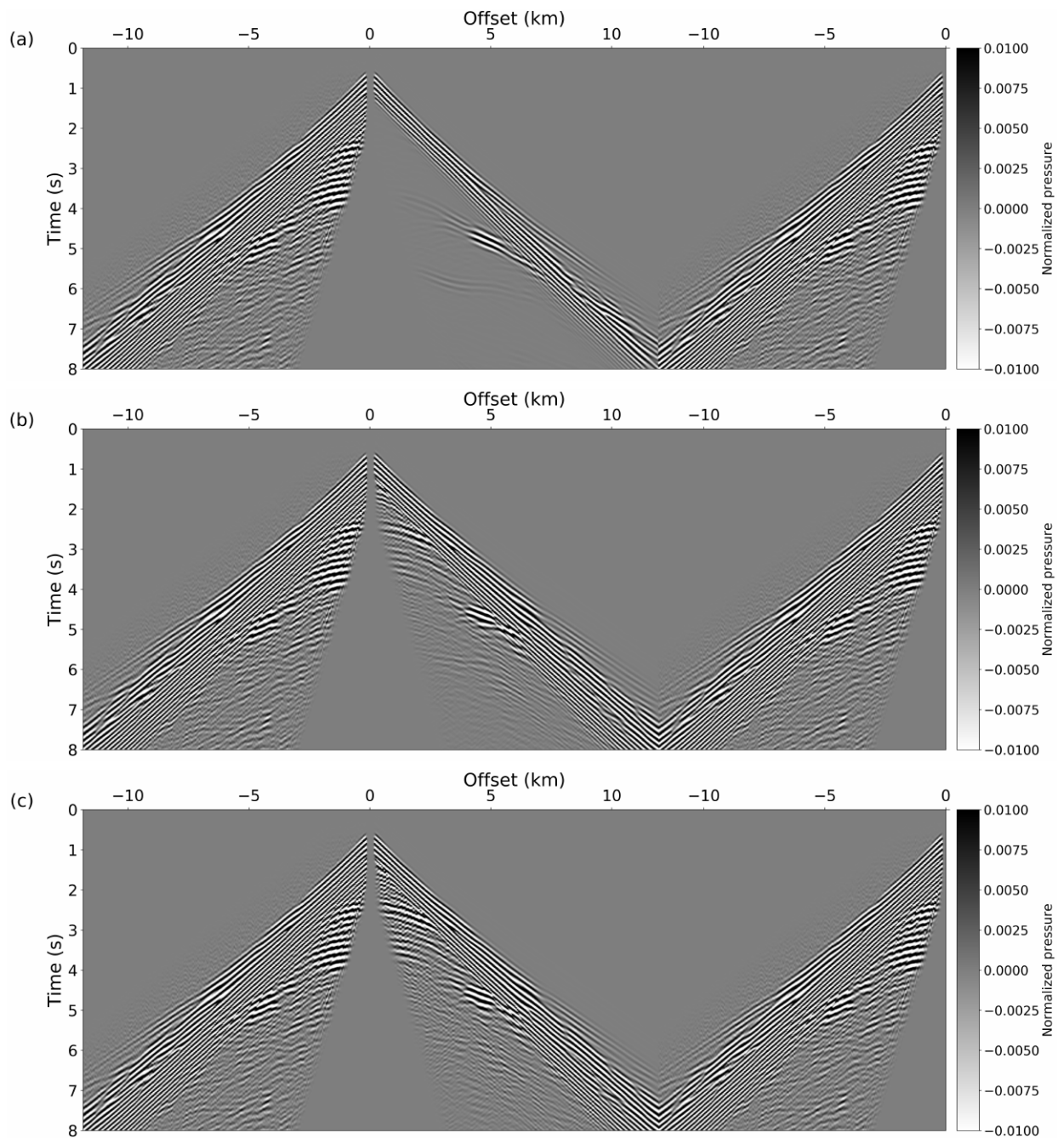


Figure 27: Data fit in mirror display for cable A. (a) Data fit in the initial model. (b) Data fit in the final model obtained using a Gaussian filter. (c) Data fit in the final model obtained using an anisotropic diffusion filter.

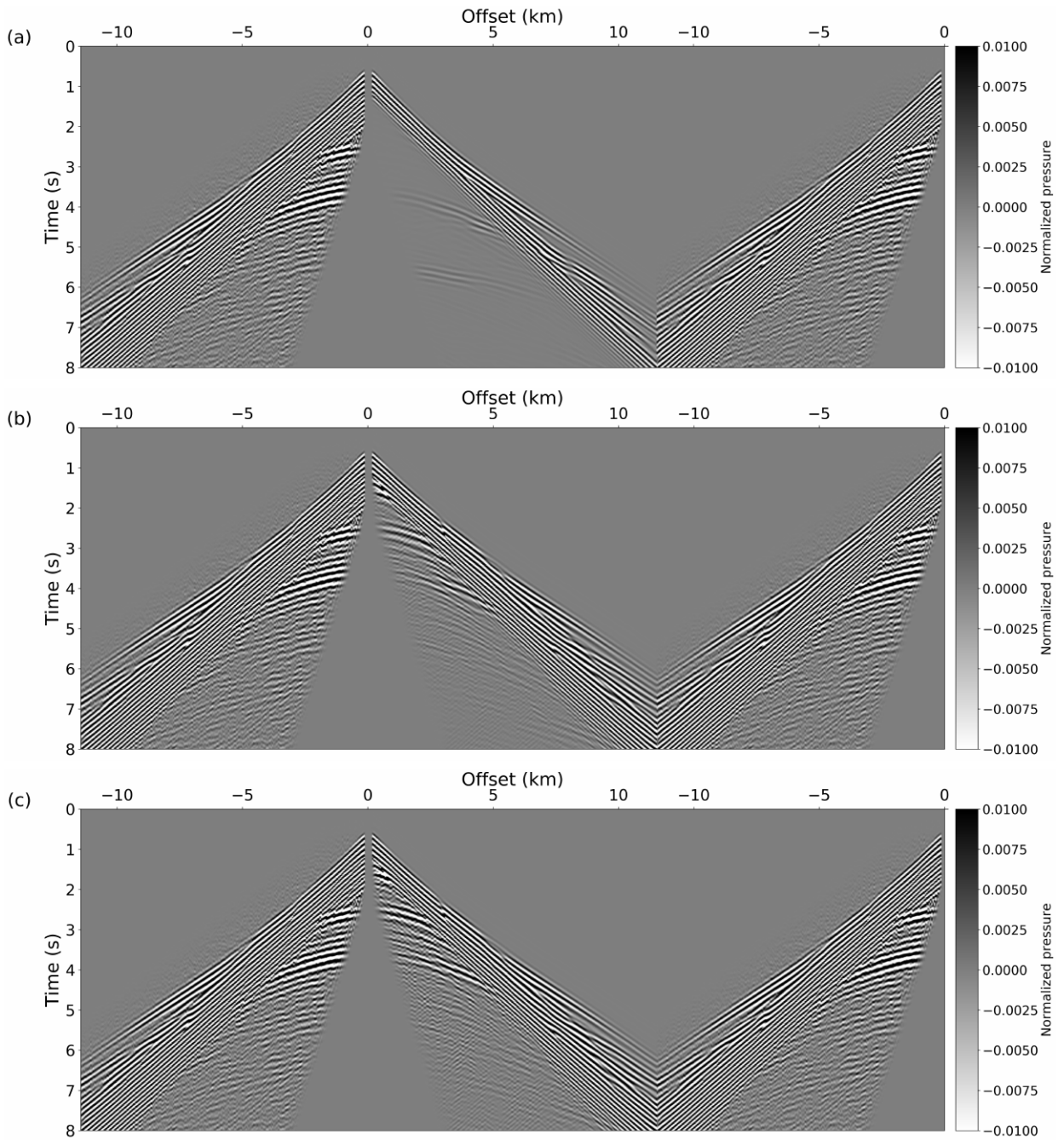


Figure 28: Data fit in mirror display for cable B. (a) Data fit in the initial model. (b) Data fit in the final model obtained using a Gaussian filter. (c) Data fit in the final model obtained using an anisotropic diffusion filter.

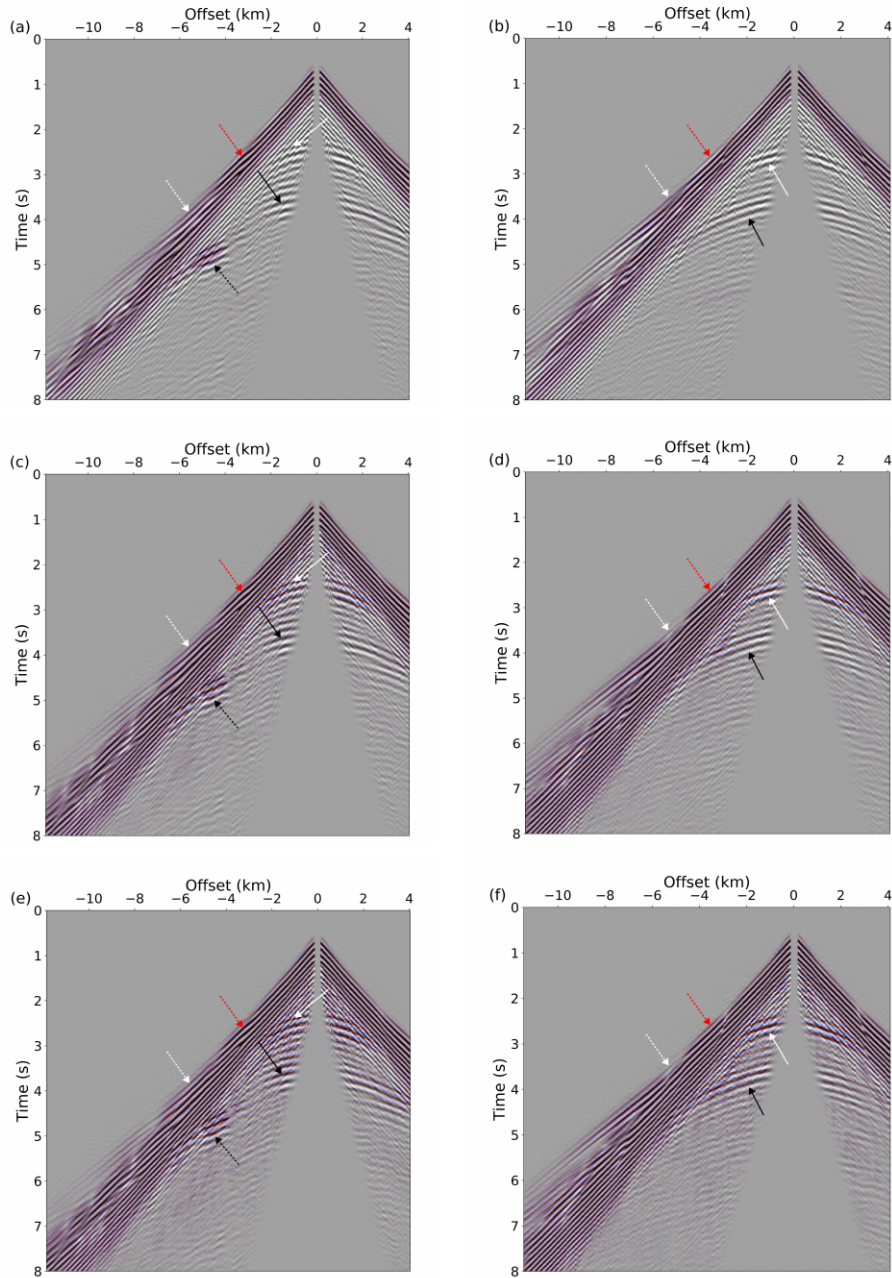


Figure 29: Comparison of the data fit by superposition of the field data in black/white colorscale and the predicted data in a red/blue colorscale. A good match is indicated by the absence of white and red color. (a,b) Data fit in the initial model for (a) cable A and (b) cable B. (c,d) Data fit in the 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.

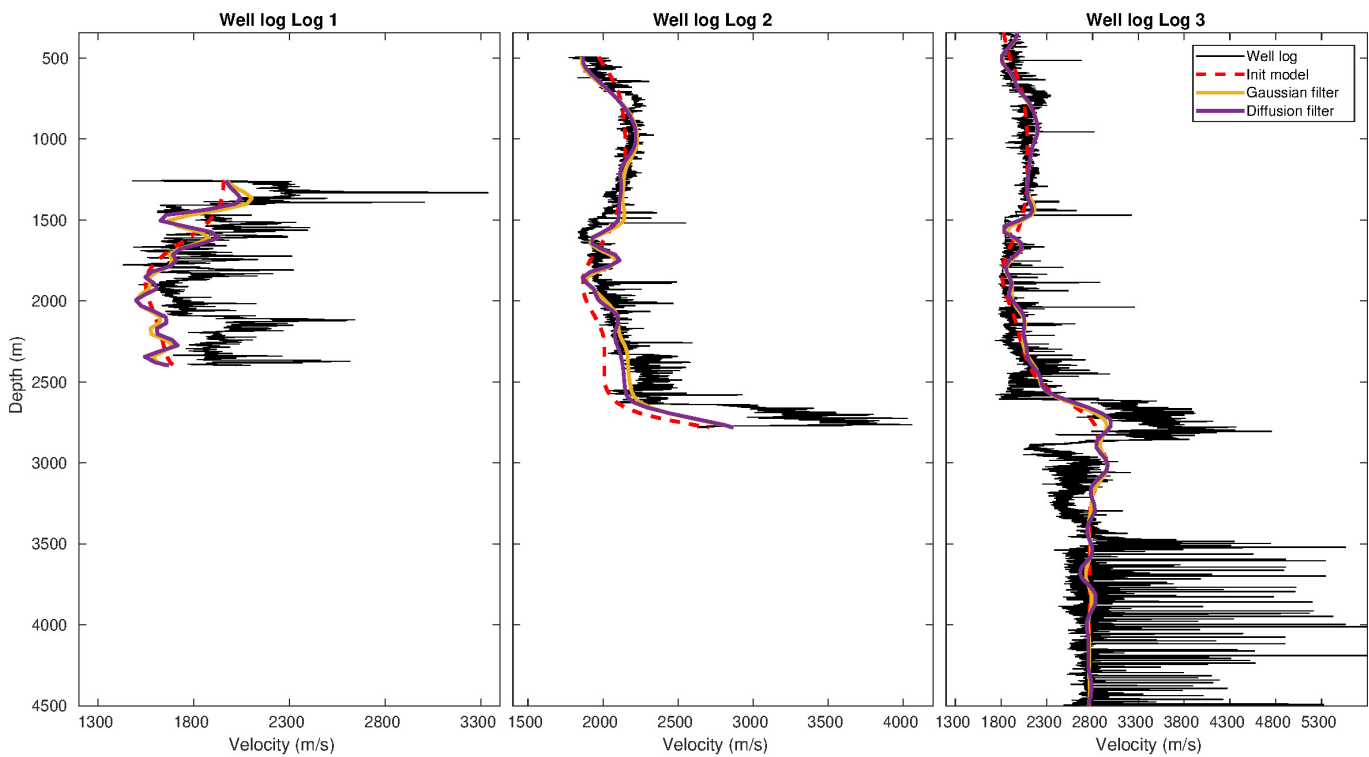


Figure 30: Comparison between sonic logs and velocity model values at the location of the wells where the sonic logs have been extracted. The sonic logs appear in solid black line. The initial model is in dotted-red line. The FWI model using a Gaussian filter is in solid orange line. The FWI model using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

LIST OF TABLES

1229	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.	92
1230			
1231			
1232			
1233			
1234			
1235			
1236			
1237			
1238			
1239	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.	93
1240			
1241			
1242			
1243			
1244			
1245			
1246			
1247			

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 (<i>1 OMP</i>)	2.5 - 7 Hz (<i>8 OMP</i>)	2.5 - 10 Hz (<i>64 OMP</i>)
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.