

# Fast 3D frequency-domain full-waveform inversion with a parallel block low-rank multifrontal direct solver: Application to OBC data from the North Sea

Patrick Amestoy<sup>1</sup>, Romain Brossier<sup>2</sup>, Alfredo Buttari<sup>3</sup>, Jean-Yves L'Excellent<sup>4</sup>, Theo Mary<sup>5</sup>, Ludovic Métivier<sup>6</sup>, Alain Miniussi<sup>7</sup>, and Stephane Operto<sup>7</sup>

## ABSTRACT

Wide-azimuth long-offset ocean bottom cable (OBC)/ocean bottom node surveys provide a suitable framework to perform computationally efficient frequency-domain full-waveform inversion (FWI) with a few discrete frequencies. Frequency-domain seismic modeling is performed efficiently with moderate computational resources for a large number of sources with a sparse multifrontal direct solver (Gauss-elimination techniques for sparse matrices). Approximate solutions of the time-harmonic wave equation are computed using a block low-rank (BLR) approximation, leading to a significant reduction in the operation count and in the volume of communication during the lower upper (LU) factorization as well as offering great potential for reduction in the memory demand. Moreover, the sparsity of the seismic source vectors is exploited to speed up the forward elimination step during the computation of the monochromatic wavefields. The relevance and the computational efficiency of the frequency-domain

FWI performed in the viscoacoustic vertical transverse isotropic (VTI) approximation was tested with a real 3D OBC case study from the North Sea. The FWI subsurface models indicate a dramatic resolution improvement relative to the initial model built by reflection traveltime tomography. The amplitude errors introduced in the modeled wavefields by the BLR approximation for different low-rank thresholds have a negligible footprint in the FWI results. With respect to a standard multifrontal sparse direct factorization, and without compromise of the accuracy of the imaging, the BLR approximation can bring a reduction of the LU factor size by a factor of up to three. This reduction is not yet exploited to reduce the effective memory usage (ongoing work). The flop reduction can be larger than a factor of 10 and can bring a factor of time reduction of around three. Moreover, this reduction factor tends to increase with frequency, namely with the matrix size. Frequency-domain viscoacoustic VTI FWI can be viewed as an efficient tool to build an initial model for elastic FWI of 4C OBC data.

## INTRODUCTION

Full-waveform inversion (FWI) (Tarantola, 1984) is now routinely used in the oil industry as part of the seismic imaging workflow, at least in soft geologic environments, such as the North Sea that make the acoustic parameterization of the subsurface accept-

able (for a discussion on this latter issue, see Barnes and Charara, 2009; Plessix and Perez Solano, 2015). However, it remains a computational challenge due to the huge number of full-waveform seismic modelings to be performed over the iterations of the FWI optimization. Seismic modeling and FWI can be performed either

Manuscript received by the Editor 25 January 2016; revised manuscript received 10 May 2016; published online 07 September 2016.

<sup>1</sup>Université Toulouse III Paul Sabatier, Institut National Polytechnique de Toulouse (INPT)-Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France. E-mail: amestoy@enseiht.fr.

<sup>2</sup>Université Grenoble Alpes, Institut des Sciences de la Terre (ISTerre), Grenoble, France. E-mail: romain.brossier@univ-grenoble-alpes.fr.

<sup>3</sup>Université Toulouse III Paul Sabatier, Centre National de la Recherche Scientifique (CNRS)-IRIT, Toulouse, France. E-mail: abuttari@enseiht.fr.

<sup>4</sup>Université de Lyon, Institut national de recherche en informatique et en automatique (INRIA)-Laboratoire de l'Informatique du Parallélisme (LIP), LIP Ecole Normale Supérieure (ENS) Lyon, INRIA, Lyon, France. E-mail: jean-yves.l'excellent@ens-lyon.fr.

<sup>5</sup>Université de Toulouse III Paul Sabatier-IRIT, Toulouse, France. E-mail: theo.mary@enseiht.fr.

<sup>6</sup>Université Grenoble Alpes, ISTerre/Laboratoire Jean Kuntzmann (LJK), CNRS, Grenoble, France. E-mail: ludovic.metivier@univ-grenoble-alpes.fr.

<sup>7</sup>Université Nice-Sophia Antipolis, Geoazur-CNRS, CNRS, Institut de recherche pour le développement (IRD), Observatoire de la Côte d'Azur, Valbonne, France. E-mail: alain.miniussi@oca.eu; operto@geoazur.unice.fr.

© 2016 Society of Exploration Geophysicists. All rights reserved.

in the time domain or in the frequency domain (e.g., Plessix, 2007; Vigh and Starr, 2008b; Virieux and Operto, 2009). Hybrid approaches are also possible, where seismic modeling and inversion are performed in the time and frequency domain, respectively (Sirgue et al., 2008). In these hybrid approaches, the monochromatic wavefields required to perform the inversion in the frequency domain are built on the fly in the loop over time steps by discrete Fourier transform during time-domain modeling (Nihei and Li, 2007). Today, most FWI codes are fully implemented in the time domain because the good scalability and the moderate memory demand of the initial-condition evolution problem underlying the forward problem allow one to tackle a wide range of applications in terms of target dimension, survey design, and wave physics. However, the time-domain formulation also requires significant computational resources to be efficient when thousands or tens of thousands of (reciprocal) seismic sources are processed in parallel.

A common parallel strategy consists of distributing these seismic sources over processors. This embarrassing parallel strategy can be combined with shared-memory parallelism and/or domain decomposition of the computational mesh, if enough computational resources are available. Strategies based on source subsampling (Warner et al., 2013) or source blending with random or deterministic (i.e., plane-wave decomposition) encoding (Vigh and Starr, 2008a; Krebs et al., 2009) are commonly used to reduce the number of seismic modelings per FWI iteration. However, these strategies, which reduce the data fold or add random noise in the FWI gradient at each iteration, require increasing the number of iterations to reach a sufficient signal-to-noise ratio in the subsurface models. Considering a computational mesh with  $n^3$  degrees of freedom, an acquisition with  $n^2$  seismic sources and assuming that the number of time steps scales to  $n$ , the time complexity of time-domain modeling scales to  $\mathcal{O}(n^6)$  (Plessix, 2007).

Alternatively, seismic modeling and FWI can be performed in the frequency domain (e.g., Pratt, 1999; Virieux and Operto, 2009). A few applications of 3D frequency-domain FWI on synthetic or real data are presented in Ben Hadj Ali et al. (2008), Plessix (2009), Petrov and Newman (2014), Operto et al. (2015). Solving the time-harmonic wave equation is a stationary boundary-value problem, which requires solving a large and sparse complex-valued system of linear equations with multiple right-hand sides per frequency (e.g., Marfurt, 1984). The sparse right-hand sides of these systems are the seismic sources, the solutions are monochromatic wavefields, and the coefficients of the so-called impedance matrix depend on the frequency and the subsurface properties we want to image. This linear system can be solved either with sparse direct methods, namely, Gauss-elimination techniques (e.g., Operto et al., 2007), iterative solvers (e.g., Riyanti et al., 2006; Plessix, 2007; Petrov and Newman, 2012; Li et al., 2015), or a combination of both in the framework of domain decomposition methods (e.g., Sourbier et al., 2011).

One pitfall of iterative methods is the design of an efficient preconditioner considering that the wave-equation operator is indefinite. More precisely, the iterative approach is competitive with the time-domain approach in terms of operation count as long as the number of iterations can be made independent of the frequency, i.e., the problem size (Plessix, 2007). It seems that this objective has not yet been fully achieved, although using a damped wave equation as a preconditioner or as such for early arrival modeling decreases the iteration count efficiently (Erlangga and Nabben, 2008; Petrov and Newman, 2012). Processing a large number of right sides leads us more naturally toward direct methods because the computation of the solu-

tions by forward/backward substitutions is quite efficient, once a lower upper (LU) factorization of the impedance matrix has been performed. The pitfalls of the direct methods are the memory demand and the limited scalability of the LU factorization that result from the fill-in of the impedance matrix generated during the LU factorization. Fill-reducing matrix orderings based on nested dissections are commonly used to reduce the memory complexity by one order of magnitude, that is  $\mathcal{O}(n^4)$  instead of  $\mathcal{O}(n^5)$  (George and Liu, 1981). The time complexity of the substitution step for  $n^2$  right-hand sides scales to  $\mathcal{O}(n^6)$  accordingly, and it is the same as the complexity of time-domain modeling. The time complexity of one LU factorization (for sparse matrices as those considered in this study) also scales to  $\mathcal{O}(n^6)$ . The conclusions that can be drawn about the relevancy of the direct solver-based approach from this complexity analysis are twofold: the identity between the time complexity of the LU factorization and the solution step for  $n^2$  right-hand sides requires the number of right-hand sides to scale to  $n^2$ . Moreover, the identity between the time complexity of one LU factorization and time-domain modeling for  $n^2$  right-hand sides requires us to limit the inversion to a few discrete frequencies. Both requirements are fulfilled by wide-azimuth long-offset acquisitions implemented with stationary-receiver geometries (ocean bottom cable [OBC] or ocean bottom node [OBN]).

On the one hand, stationary-receiver geometries involve a large number of sources and associated receivers in the computational domain, from which the LU factorization is performed. On the other hand, the wide-azimuth long-offset coverage provided by these geometries generates a strong redundancy in the wavenumber sampling of the subsurface target. This multifold wavenumber coverage results from the redundant contribution of finely sampled (temporal) frequencies and a broad range of finely sampled scattering angles provided by dense point-source acquisitions (Pratt and Worthington, 1990). The strategy consisting of coarsening the frequency sampling in the data space to remove the redundancy of the wavenumber sampling in the model space has been referred to as efficient FWI by Sirgue and Pratt (2004).

The focus of this study is to present an up-to-date status of the computational efficiency of 3D frequency-domain FWI based on sparse direct solvers with a real OBC data case study from the North Sea. We have presented a first application of 3D frequency-domain FWI on the North Sea OBC data set in Operto et al. (2015), which is focused on a detailed quality control of the FWI results based on seismic modeling and source wavelet estimation. Here, we show the benefits resulting from recent advances in the development of a massively parallel sparse direct solver for FWI applications. In this study, we solve the linear system resulting from the discretization of the time-harmonic wave equation with the MUMPS multifrontal massively parallel solver (Amestoy et al., 2001, 2006) that performs the LU factorization with a multifrontal method (Duff and Reid, 1983).

Compared with the results shown in Operto et al. (2015), obtained with the public release 5.0.1 of MUMPS (MUMPS team, 2015), two recent developments in this solver, which will be made available in the next major release, provide further reduction of the computational cost. First, we speedup the forward substitution step by exploiting the sparsity of the right-hand sides, namely the seismic source vectors, by extending the algorithmic work done for computing selected entries of the inverse of a matrix (Amestoy et al., 2015d). Second, we reduce the operation count of the LU factorization by exploiting the existence of blockwise low-rank approximants of the elliptic partial differential operators embedded in the time-harmonic wave equation (Amestoy et al., 2015a). The gov-

erning idea consists of recasting the so-called frontal matrices of the elimination tree in a block representation and compress blocks that have low-rank properties; the effectiveness and accuracy of this compression is controlled by a prescribed threshold. In the remainder of this study, we will refer to this approach as the block low-rank (BLR) format-based approximation. This BLR format is an alternative to other low-rank formats, such as the hierarchically semiseparable (HSS) format that has been proposed for seismic modeling by Wang et al. (2011, 2012a, 2012b). Even though the theoretical complexity of the HSS format is better than that of the BLR format, the simplicity and flexibility of the BLR format makes it easy to use in the context of a general purpose, algebraic multifrontal solver. Recent work has analyzed the theoretical bounds on the complexity of the BLR format (Amestoy et al., 2016b). In this study, we compare these theoretical results with numerical experiments carried out with a 3D FWI application and we illustrate the potential of the proposed format in reducing the complexity of the multifrontal solver.

This paper is organized as follows: The first section reviews the main ingredients of the forward-modeling engine interfaced with the frequency-domain FWI code. We first briefly review the finite-difference stencil with which the time-harmonic wave equation is discretized. This stencil must satisfy some specifications so that the fill-in of the impedance matrix is minimized during the LU factorization. The reader is referred to Operto et al. (2014) for a more detailed description of this finite-difference stencil. Then, we review the basic principles of the multifrontal method, its BLR approximation and the strategy implemented to exploit the sparsity of the source vectors. The reader is referred to Amestoy et al. (2015a) for a more detailed description of the BLR multifrontal method. The second section presents the application on the OBC data from the North Sea. We first show the nature of the errors introduced by the BLR approximation in the monochromatic wavefields and quantify the backward errors for different frequencies. For a given BLR threshold, the ratio between the backward errors obtained with the BLR and the full-rank (FR) solvers decreases with frequency. This suggests that a more aggressive threshold can be used as the frequency increases, hence leading to more efficient compression as the problem size grows. Then, we show that the modeling errors have a negligible impact in the FWI results. The cost of the FWI in terms of memory and time confirms that the computational saving provided by the BLR solver relative to the FR solver increases with frequency (namely, matrix size). We conclude with a strong and weak scalability analysis of the FR and BLR solvers for the subsurface models and frequencies considered in this case study.

The limited computational resources that have been used to perform this case study and the limited computational time required to perform the FWI in the 3.5–10 Hz frequency band using all the sources and receivers of the survey at each FWI iteration highlights the computational efficiency of our frequency-domain approach to process stationary recording surveys in the viscoacoustic vertical transverse isotropic (VTI) approximation.

## METHODS

### Frequency-domain modeling with block low-rank multifrontal solvers

#### *Finite-difference stencils for frequency-domain seismic modeling*

Frequency-domain seismic modeling with direct solvers defines some stringent specifications, whose objective is to minimize the

computational burden of the LU factorization. The first specification aims to minimize the dependencies in the adjacency graph of the matrix, and hence the fill-in of the impedance matrix during the LU factorization. The second one aims to obtain accurate solutions for a coarse discretization that is matched to the spatial resolution of the FWI (i.e., four grid points per wavelengths according to a theoretical resolution of half a wavelength). The first specification can be fulfilled by minimizing the numerical bandwidth and optimizing the sparsity of the impedance matrix using finite-difference stencils with compact spatial support. This precludes using conventional high-order accurate stencils. Instead, accurate stencils are designed by linearly combining different second-order accurate stencils that are built by discretizing the differential operators on different coordinate systems with a spreading of the mass term over the coefficients of the stencil (Jo et al., 1996; Stekl and Pratt, 1998; Min et al., 2000; Hustedt et al., 2004; Gosselin-Cliche and Giroux, 2014). Such stencils are generally designed for the second-order acoustic/elastic wave equations, by opposition to first-order velocity-stress equations, as the second-order formulation involves fewer wavefield components, hence limiting the dimension of the impedance matrix accordingly. For the 3D viscoacoustic wave equation, the resulting finite-difference stencil has 27 nonzero coefficients distributed over two grid intervals and provides accurate solution for arbitrary coarse grids provided that optimal coefficients are estimated by fitting the discrete dispersion equation in homogeneous media (Operto et al., 2007; Brossier et al., 2010). The viscoacoustic 27-point stencil was recently extended to account for VTI anisotropy without generating significant computational overhead (Operto et al., 2014).

The viscoacoustic VTI 27-point stencil results from the discretization of the following equations:

$$A_h \mathbf{p}_h = \mathbf{s}', \quad (1)$$

$$\mathbf{p}_v = A_v \mathbf{p}_h + \mathbf{s}'', \quad (2)$$

$$p = \frac{1}{3}(2\mathbf{p}_h + \mathbf{p}_v), \quad (3)$$

where  $\mathbf{p}_h$ ,  $\mathbf{p}_v$ , and  $\mathbf{p}$  are the so-called horizontal pressure, vertical pressure, and (physical) pressure wavefields, respectively. The operators  $A_h$  and  $A_v$  are given by

$$A_h = \omega^2 \left[ \frac{\omega^2}{\kappa_0} + (1 + 2\epsilon)(\mathcal{X} + \mathcal{Y}) + \sqrt{1 + 2\delta}\mathcal{Z} \frac{1}{\sqrt{1 + 2\delta}} \right] + 2\sqrt{1 + 2\delta}\mathcal{Z} \frac{\kappa_0(\epsilon - \delta)}{\sqrt{1 + 2\delta}}(\mathcal{X} + \mathcal{Y}),$$

$$A_v = \frac{1}{\sqrt{1 + 2\delta}} + \frac{2(\epsilon - \delta)\kappa_0}{\omega^2\sqrt{1 + 2\delta}}(\mathcal{X} + \mathcal{Y}), \quad (4)$$

where  $\kappa_0 = \rho V_0^2$ ,  $\rho$  is the density,  $V_0$  is the vertical wavespeed,  $\omega$  is the angular frequency, and  $\delta$  and  $\epsilon$  are the Thomsen's parameters. Differential operators  $\mathcal{X}$ ,  $\mathcal{Y}$ , and  $\mathcal{Z}$  are given by  $\partial_{\tilde{x}} b \partial_{\tilde{x}}$ ,  $\partial_{\tilde{y}} b \partial_{\tilde{y}}$ , and  $\partial_{\tilde{z}} b \partial_{\tilde{z}}$ , respectively, where  $b = 1/\rho$  is the buoyancy and  $(\tilde{x}, \tilde{y}, \tilde{z})$  define a complex-valued coordinate system in which perfectly matched layers absorbing boundary condition are implemented (Operto et al., 2007). The right side vectors  $\mathbf{s}'$  and  $\mathbf{s}''$  have the following expression:

$$s'(\mathbf{x}, \omega) = \omega^4 s(\omega) \frac{s_h(\mathbf{x})}{\kappa_0(\mathbf{x})} \tilde{\delta}(\mathbf{x} - \mathbf{x}_s) - \omega^2 s(\omega) \sqrt{1 + 2\delta(\mathbf{x})} \mathcal{Z} \times \left( s_v(\mathbf{x}) - \frac{1}{\sqrt{1 + 2\delta(\mathbf{x})}} s_h(\mathbf{x}) \right) \tilde{\delta}(\mathbf{x} - \mathbf{x}_s), \quad (5)$$

$$s''(\mathbf{x}, \omega) = \left( s_v(\mathbf{x}) - \frac{1}{\sqrt{1 + 2\delta(\mathbf{x})}} s_h(\mathbf{x}) \right) \tilde{\delta}(\mathbf{x} - \mathbf{x}_s), \quad (6)$$

where  $\tilde{\delta}$  denotes the Dirac delta function,  $\mathbf{x}_s$  denotes the source position,  $s(\omega)$  is the source excitation term, and  $s_h$  and  $s_v$  are the two quantities that depend on the Thomsen's parameters.

The expression of the operator  $A_h$  shows that the VTI wave equation has been decomposed as a sum of an elliptically anisotropic wave equation (term between brackets) plus an anelliptic correcting term involving the factor  $(\epsilon - \delta)$ . The elliptic part can be easily discretized by plugging the Thomsen's parameters  $\delta$  and  $\epsilon$  in the appropriate coefficients of the isotropic 27-point stencil, whereas the anelliptic correcting term is discretized with conventional second-order accurate stencil to preserve the compactness of the overall stencil.

The direct solver is used to solve the linear system involving the matrix  $A_h$  (equation 1). In a second step, the vertical pressure  $\mathbf{p}_v$  is explicitly inferred from the expression of  $\mathbf{p}_h$ , equation 2, before forming the physical pressure wavefield  $\mathbf{p}$  by linear combination of  $\mathbf{p}_h$  and  $\mathbf{p}_v$  (equation 3).

The accuracy of the 27-point stencil for the viscoacoustic isotropic and VTI equations is assessed in details in [Operto et al. \(2007\)](#), [Brossier et al. \(2010\)](#), and [Operto et al. \(2014\)](#) and allows for accurate modeling with a discretization rule of four grid points per minimum wavelength.

*LU factorization and solution with the multifrontal method*

Solving the time-harmonic wave equation requires solving a large and sparse complex-valued system of linear equations

$$\mathbf{A}\mathbf{X} = \mathbf{B}, \quad (7)$$

where  $\mathbf{B}$  is also a large sparse matrix of so-called right-hand sides, which correspond to the seismic sources. To solve the system, we first factor the matrix  $\mathbf{A}$  ( $\mathbf{A} = \mathbf{L}\mathbf{U}$ ) with the multifrontal method (first introduced by [Duff and Reid, 1983](#)), and then perform a forward step  $\mathbf{L}\mathbf{Y} = \mathbf{B}$  followed by a backward step  $\mathbf{U}\mathbf{X} = \mathbf{Y}$ .

The multifrontal method performs a sequence of partial factorizations of dense matrices, called frontal matrices or, simply, fronts. This sequence is established by a tree-shaped dependency graph called the elimination tree ([Schreiber, 1982](#)), which has a front associated with each of its nodes (Figure 1). This graph is traversed from the bottom (the leaves) to the top (the root) and, each time a node is visited, first the front is formed through assembly operations, and then it is partially factorized through a truncated dense LU reduction. Two sets of variables are associated with each front: The fully summed (FS) variables, whose corre-

sponding rows and columns of  $\mathbf{L}$  and  $\mathbf{U}$  are computed within the current front, and the nonfully summed (NFS) variables, which receive updates resulting from the elimination of the FS variables. The structure of a front before and after the partial factorization is shown in Figure 2. The results of the front partial factorization are the partial factors  $[L_{11} L_{21}]$  and  $[U_{11} U_{12}]$ , which are stored apart and a Schur complement, referred to as a contribution block (CB), which will be later used in the assembly operations of the parent front before being discarded. All the CBs are temporarily stored in a memory area called CB stack, whose size varies throughout the tree traversal.

The computational and memory requirements for the complete factorization strongly depend on the sparsity structure of the input matrix. Matrix permutations are commonly applied prior to the actual factorization to reduce the fill-in (i.e., zero coefficients, which are turned into nonzero by the factorization), which is mostly responsible for this cost, in terms of floating-point operations and memory consumption. Among the most commonly used techniques, those based on nested dissection commonly give the best results on large scale problems ([George and Liu, 1981](#)). The shape of the elimination tree and the size of the fronts therein, result from the applied matrix permutation. Although any topological-order traversal of this tree gives equivalent properties in terms of factor matrix size and computational cost, the size of the CB stack area can greatly vary depending on the chosen traversal order; methods exist to minimize the peak size of the CB stack in a sequential execution ([Guermouche et al., 2003](#)) and to control its size in a parallel setting where multiple branches of the tree are traversed in parallel ([Agullo et al., 2016](#)).

In a parallel environment, the sparse matrix is distributed onto the processors according to a mapping of the processors on the nodes of the elimination tree. Two kinds of parallelism, referred to as tree parallelism and node parallelism, are exploited (Figure 3). In MUMPS, both levels of parallelism are exploited with message passing interface (MPI). In tree parallelism, fronts in different subtrees are processed concurrently by different processes, whereas in node parallelism, large enough fronts are mapped on several proc-

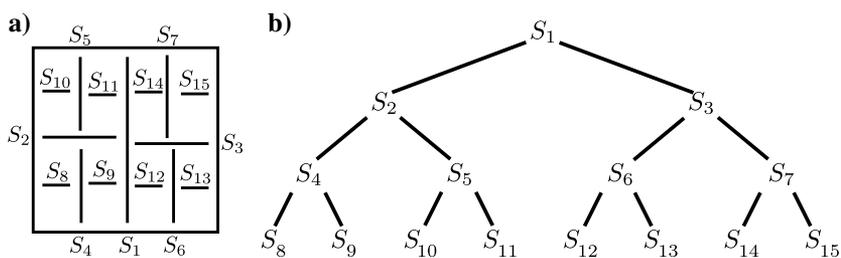


Figure 1. An example of (b) elimination tree, with the ordering based on (a) a nested dissection.

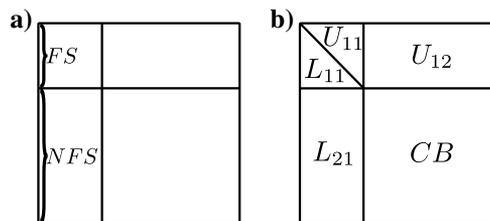


Figure 2. A front (a) before and (b) after partial factorization.

esses: the master process is assigned to process the FS rows and is in charge of organizing computations; the NFS rows are distributed following a 1D row-wise partitioning, so that each slave holds a range of rows. Within the block row of each MPI process, node parallelism is exploited at a finer level with multithreading.

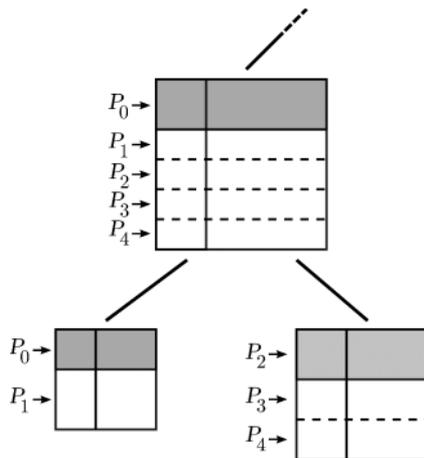


Figure 3. Illustration of tree and node parallelism. The shaded part of each front represents its FS rows. The fronts are row-wise partitioned in our implementation, but column-wise partitioning is also possible.

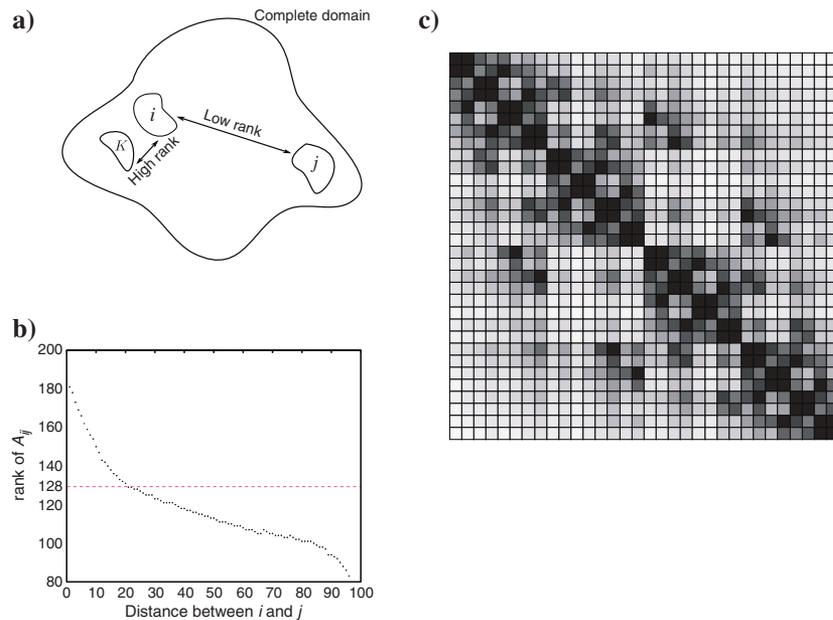


Figure 4. Illustrations of the admissibility condition for elliptic PDEs. (a) Strong and weak interactions in the geometric domain as a function of the distance between degrees of freedom. (b) Correlation between graph distance and rank ( $\epsilon = 10^{-8}$ ). The experiment has been performed on the dense Schur complement associated with the top-level separator of a 3D  $128^3$  Helmholtz problem (from Amestoy et al., 2015a). (c) Illustration of a BLR matrix of the top-level separator of a 3D  $128^3$  Poisson’s problem. The darkness of a block is proportional to its storage requirement: the lighter a block is, the smaller is its rank. Off-diagonal blocks are those having low-rank properties (from Amestoy et al., 2015a).

Block low-rank multifrontal solver

A matrix  $\mathbf{A}$  of size  $m \times n$  is said to be low rank if it can be approximated by a low-rank product  $\tilde{\mathbf{A}} = \mathbf{X}\mathbf{Y}^T$  of rank  $k_\epsilon$ , such that  $k_\epsilon(m + n) \leq mn$  and  $\|\mathbf{A} - \tilde{\mathbf{A}}\| \leq \epsilon$ . The first condition states that the low-rank form of the matrix requires less storage than the standard form, whereas the second condition simply states that the approximation is of good enough accuracy. Using the low-rank form also allows for a reduction of the number of floating-point operations performed in many kernels (e.g., matrix-matrix multiplication).

It has been shown that matrices resulting from the discretization of elliptic partial differential equations (PDEs), such as the generalization of the Helmholtz equation considered in this study, have low-rank properties (Bebendorf, 2004). In the context of the multifrontal method, frontal matrices are not low rank themselves but exhibit many low-rank subblocks. To achieve a satisfactory reduction in the computational complexity and the memory footprint, sub-blocks have to be chosen to be as low rank as possible (e.g., with exponentially decaying singular values). This can be achieved by clustering the unknowns in such a way that an admissibility condition is satisfied. This condition states that a subblock interconnecting different variables will have a low rank, if these associated variables are far away in the domain, because they are likely to have a weak interaction. This intuition is illustrated in Figure 4a and 4b. Diagonal blocks that represent self-interactions are always FR, whereas the off-diagonal blocks are potentially low rank (Figure 4c). In the framework of an algebraic solver, a graph partitioning tool is used to partition the subgraphs induced by the FS and NFS variables (Amestoy et al., 2015a). The blocks are compressed with a truncated QR factorization with column pivoting.

Unlike other formats, such as  $\mathcal{H}$ -matrices (Hackbusch, 1999), HSS (Xia et al., 2009; Xia, 2013), and HODLR (Aminfar and Darve, 2016) matrices, we use a flat, nonhierarchical blocking of the fronts (Amestoy et al., 2015a). Thanks to the low-rank compression, the theoretical complexity of the factorization is reduced from  $\mathcal{O}(n^6)$  to  $\mathcal{O}(n^{5.5})$  and can be further reduced to  $\mathcal{O}(n^5 \log n)$  with the best variant of the BLR format (Amestoy et al., 2016b). Although compression rates may not be as good as those achieved with hierarchical formats (hierarchical formats can achieve a complexity in  $\mathcal{O}(n^5)$  [Xia et al., 2009], and even  $\mathcal{O}(n^4)$  in a fully structured context), BLR offers a good flexibility thanks to its simple, flat structure. This makes BLR easier to adapt to any multifrontal solver without a deep redesign of the code. The comparison between hierarchical and BLR formats is an ongoing work, in particular in Amestoy et al. (2016a), where the HSS-based STRUMPACK (Ghysels et al., 2015; Rouet et al., 2015) and the BLR-based MUMPS (Amestoy et al., 2015a) solvers are compared. In the following, the low-rank threshold used to compress the blocks of the frontal matrices will be denoted by  $\epsilon$ .

In parallel environments, the row-wise partitioning imposed by the distribution of the front onto several processes constrains the clustering of the unknowns. However, in practice, we manage to maintain nearly the same flop compression rates when the number of processes grows as discussed in the following scalability analysis. The LU compression during the factorization of the fronts contributes to reducing the volume of communication by a substantial factor and to maintaining a good parallel efficiency of the solver. Exploiting the compression of the CB matrices would further reduce the volume of communication and the memory footprint (but not the operation count), but is not available in our current implementation. The MPI parallelism is hybridized with thread parallelism to fully exploit multicore architectures. The FR tasks (i.e., not involving low-rank blocks) are efficiently parallelized through multithreaded basic linear algebra subroutines (BLAS; Coleman and vanLoan, 1988) because of their large granularity. In contrast, the low-rank tasks have a finer granularity that makes multithreaded BLAS less efficient and the conversion of flops compression into time reduction more challenging. To overcome this obstacle, we exploit OpenMP-based multithreading to execute multiple low-rank tasks concurrently, which allows for a larger granularity of computations per thread (Amestoy et al., 2015b).

#### *Exploiting the sparsity of the right-hand sides*

Even though the BLR approximation has not yet been implemented in the solution phase, we reduce the number of operations performed and the volume of data exchanged during this phase by exploiting the sparsity of the right-hand sides. Furthermore, we also improved the multithreaded parallelism of this phase.

During the forward phase and thanks to the sparsity of the right-hand side matrix  $\mathbf{B}$ , equation 7, a property shown in the work of Gilbert and Liu (1993) can be used to limit the number of operations for computing the solution  $\mathbf{Y}$  of  $\mathbf{LY} = \mathbf{B}$ : for each column  $b_j$  of  $\mathbf{B}$ , one has only to follow a union of paths in the elimination tree, each path being defined by a nonzero entry in  $b_j$ . The flop reduction results from this property and from the fact that the nonzero entries in each source  $b_j$  have a good locality in the elimination tree, so that only a few branches of the elimination tree per source need to be traversed.

This is sometimes not enough to provide large gains in a parallel context because processing independent branches of the elimination tree is necessary to provide significant parallelism. The question is then how to permute the columns of matrix  $\mathbf{B}$  to reduce the number of operations while providing parallelism. This combinatorial problem has been studied for computing selected entries of the inverse of a matrix by Amestoy et al. (2015d), who explain how to permute and process columns of the right-hand side matrix in a blockwise fashion. This work has been extended to the forward step  $\mathbf{LY} = \mathbf{B}$  with a sparse  $\mathbf{B}$ . In the framework of seismic imaging application, we exploit the fact that the nonzero entries of one block of columns of the right-hand side matrix  $\mathbf{B}$  are well-clustered in the elimination tree, when the corresponding sources are contiguous in the computational mesh (i.e., follow the natural ordering of the acquisition) and when nested dissection is used for reordering the unknowns of the problem. We exploit this geometric property of the source distribution to choose those relevant subsets of sparse right-hand sides that are simultaneously processed in parallel during the forward elimination phase.

## APPLICATION TO OBC DATA FROM THE NORTH SEA

### Acquisition geometry, geologic context, and FWI experimental setup

The subsurface target and the data set are the same as in Operto et al. (2015). The FWI experimental setup is also similar except that we perform FWI with a smaller number of discrete frequencies (6 instead of 11), and we fix the maximum number of FWI iterations per frequency as stopping criterion of iterations. A brief review of the target, data set, and experimental setup is provided here. The reader is referred to Operto et al. (2015) for a more thorough description.

#### *Geologic target*

The subsurface target is the Valhall oil field located in the North Sea in a shallow water environment (70 m water depth; Barkved et al., 2010; Sircue et al., 2010). The reservoir is located at approximately 2.5 km depth. The overburden is characterized by soft sediments in the shallow part. A gas cloud, whose main zone of influence is delineated in Figure 5a, makes seismic imaging at the reservoir depths challenging. The parallel geometry of the wide-azimuth OBC acquisition consists of 2302 hydrophones, which record 49,954 explosive sources located 5 m below the sea surface (Figure 5a). The seismic sources cover an area of 145 km<sup>2</sup> leading to a maximum offset of 14.5 km. The maximum depth of investigation in the FWI model is 4.5 km. The seismograms recorded by a receiver for a shot profile intersecting the gas cloud and the receiver position are shown in Figure 5b. The wavefield is dominated by the diving waves, which mainly propagate above the gas zone, the reflection from the top of the gas cloud (Figure 5b, white arrow), and the reflection from the reservoir (Figure 5b, black arrow) (Prioux et al., 2011, 2013a, 2013b; Operto et al., 2015). In Figure 5b, the solid black arrow points to the precritical reflection from the reservoir, whereas the dashed black arrow points to the critical and postcritical reflection. The discontinuous pattern in the time-offset domain of this wide-angle reflection highlights the complex interaction of the wavefield with the gas cloud.

#### *Initial models*

The vertical-velocity  $V_0$  and the Thomsen's parameter models  $\delta$  and  $\epsilon$ , which are used as initial models for FWI, were built by reflection traveltime tomography (courtesy of BP) (Figures 6a–6d and 7a and 7d). The  $V_0$  model describes the long wavelengths of the subsurface except at the reservoir level, which are delineated by a sharp positive velocity contrast at approximately 2.5 km depth (Figure 7a and 7d). We do not smooth this velocity model before FWI for reasons explained in Operto et al. (2015, their Figure 7). The velocity model allows us to match the first-arrival traveltimes as well as those of the critical reflection from the reservoir Operto et al. (2015, their Figure 8), hence providing a suitable framework to prevent cycle skipping during FWI. A density model was built from the initial vertical velocity model using a polynomial form of the Gardner law given by  $\rho = -0.0261V_0^2 + 0.373V_0 + 1.458$  (Castagna et al., 1993) and was kept fixed over iterations. A homogeneous model of the quality factor was used below the sea bottom with a value of  $Q = 200$ .

### FWI experimental setup

We review here the experimental setup that was used to apply frequency-domain FWI on the OBC data set. The reader is referred to Appendix A for some algorithmic aspects of 3D frequency-domain FWI because they can have a significant impact together with the seismic modeling engine on the computational efficiency of the inversion. The discrete frequencies, the computational resources, and the finite-difference grid dimensions that are used to perform FWI are reviewed in Table 1. We perform FWI for six discrete frequencies in the 3.5–10 Hz frequency band. Only one frequency is processed at a time, and the inversion proceeds sequentially from the lowest frequency to the highest one following a frequency-driven multiscale approach (Pratt, 1999). The grid interval in the subsurface models is periodically matched to the frequency  $f$  to minimize the computational time and regularize the inversion by avoiding over parameterization. We use a grid interval of 70 m for  $3.5 \text{ Hz} \leq f \leq 5 \text{ Hz}$ , 50 m for  $f = 7 \text{ Hz}$ , and 35 m for  $f = 10 \text{ Hz}$ . These intervals approximately lead to four grid points per wavelength for a minimum wavespeed of 1400 m/s. The number of degrees of freedom in the 70, 50, and 35 m finite-difference grids are 2.9, 7.2, and 17.4 million, respectively, after adding perfectly matched absorbing layers (Bérenger, 1996; Operto et al., 2007). We perform FWI on 12, 16, and 34 computer nodes for the 70, 50, and 35 m grids, respectively (Table 1). The computer nodes are equipped with two 2.5 GHz Intel Xeon IvyBridge E5-2670v2 processors with 10 cores per processor. The shared memory per node is 64 GB. The connecting network is InfiniBand fourth data rate (FDR) at 56 Gb/s. The operations are performed in single-precision complex arithmetic, for which the peak performance of the machine is 10 Gflops/s/core (which corresponds to a double precision peak of 20 Gflops/s/core).

We launch two MPI processes per node (i.e., one MPI process per processor) and use 10 threads per MPI process such that the number of MPI processes times the number of threads is equal to the number of cores on the node (see Appendix A for a more thorough discussion). The number of nodes for the three grids (12, 16, and 34) were chosen pragmatically to find the best compromise between the computational efficiency, the memory use and the fast access to the computational resources according to the operating rule of the available cluster. In the following, we present a strong and weak scalability of the FR and BLR solvers in the framework of this case study to provide insights on the computational performance that could have been achieved with more resources.

We use all the (reciprocal) shots and receivers at each FWI iteration, whatever the grid interval and the frequency. This implies that 4604 wavefield solutions need to be computed for each FWI gradient computation. Seismic modeling is performed with a free surface on top of the finite-

difference grid during inversion. Therefore, free-surface multiples and ghosts were left in the data and accounted for during FWI. We only update  $V_0$  during inversion, whereas  $\rho$ ,  $Q$ ,  $\delta$ , and  $\epsilon$  are kept to their initial values. We use the preconditioned steepest-descent method implemented in the SEISCOPE toolbox (Métivier and Brossier, 2015) to perform FWI, where the preconditioner is provided by the diagonal terms of the so-called pseudo-Hessian (Shin et al., 2001). No regularization and no data weighting were used. We simultaneously process subsets of 240 (first-frequency band), 256 (second-frequency band), and 68 (third-frequency band) right sides during the substitution step performed by the direct solver (see Appendix A). These numbers (denoted by  $N_s$  in Appendix A) are chosen to be a multiple of the number of MPI processes

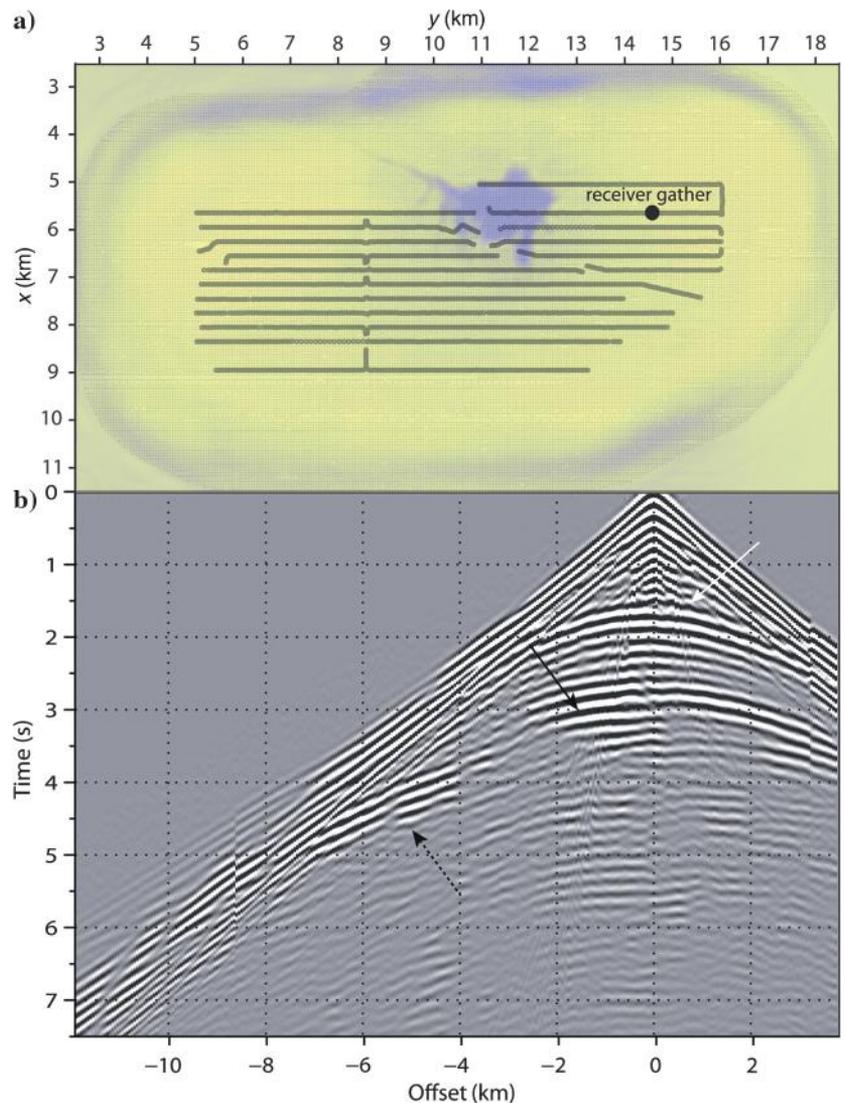


Figure 5. North Sea case study. (a) Acquisition layout. The green lines are cables. The dot pattern shows the area covered by the 50,000 explosive sources. The black circle shows the position of the receiver, whose records are shown in panel (b). A depth slice at 1 km depth across the gas cloud is superimposed in transparency to show the zone of influence of this gas cloud. (b) Common receiver gather for a shot profile crosscutting the receiver position (circle in panel [a]) and the gas cloud. The white arrow points to the reflection from the top of the gas. The black arrows point to precritical (solid) and post-critical (dash) reflections from the reservoir.

(24, 32, and 68 on the 70, 50, and 35 m grids, respectively; Table 1) because the building of the right sides of the state and adjoint problems is distributed over MPI processes following an embarrassing parallelism. A high number of right-sides is favorable to minimize the disk traffic during data reading and optimize the multiple-right-side substitution step. However, we choose to keep these numbers relatively low because one source wavelet is averaged over the number of simultaneously processed sources in our current implementation.

The stopping criterion of iterations consists of fixing the maximum iteration to 15 for the 3.5 and 4 Hz frequencies, 20 for the 4.5, 5, and 7 Hz frequencies, and 10 for the 10 Hz frequency. We use a limited number of iterations at 3.5 and 4 Hz because of the poor signal-to-noise ratio. Although this stopping criterion of iteration might seem

quite crude, we show that a similar convergence rate was achieved by FWI performed with the FR and BLR solvers at the 7 and 10 Hz frequencies, leading to very similar final FWI results. Therefore, the FWI computational costs achieved with the FR and BLR solvers in this study are provided for FWI results of similar quality.

#### *Nature of the modeling errors introduced by the BLR solver*

We show now the nature of the errors introduced in the wavefield solutions by the BLR approximation. Figures 8a, 9a, and 10a show a 5, 7, and 10 Hz monochromatic common-receiver gather computed with the FR solver in the FWI models obtained after the 5, 7, and 10 Hz inversions (the FWI results are shown in the next

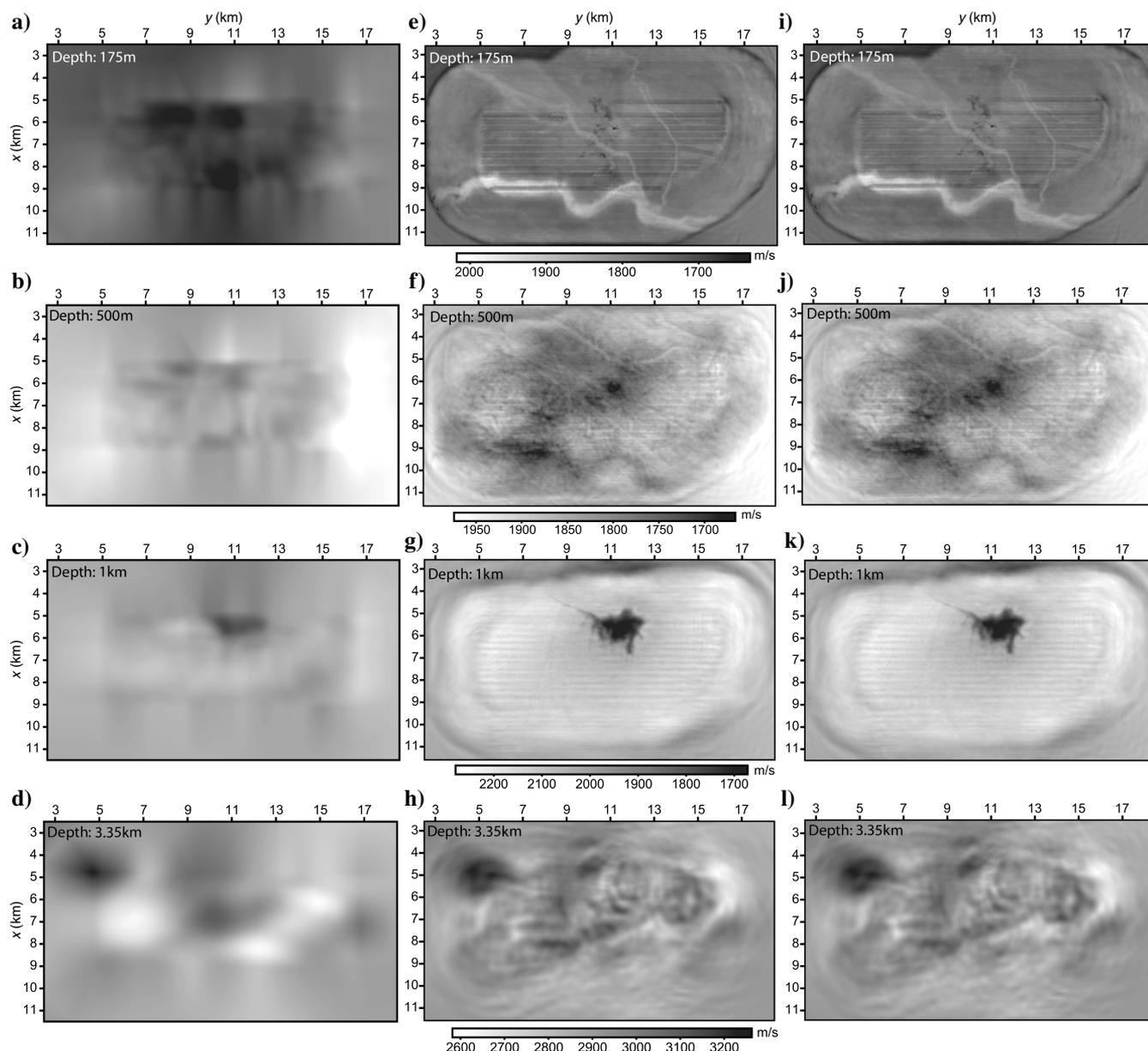


Figure 6. North Sea case study. (a-d) Depth slice extracted from the initial model at (a) 175 m, (b) 500 m, (c) 1 km, and (d) 3.35 km depths. (e-h) Same as panels (a-d) for depth slices extracted from the FWI model obtained with the FR solver. (i-l) Same as panels (a-d) for depth slices extracted from the FWI model obtained with the BLR solver.

subsection). Figures 8b–8d, 9b–9d, and 10b–10d show the differences between the common-receiver gathers computed with the FR solver and those computed with the BLR solver using  $\epsilon = 10^{-5}$ ,  $10^{-4}$ , and  $10^{-3}$  (the same subsurface model is used to perform the FR and the BLR simulations). These differences are shown after multiplication by a factor 10. A direct comparison between the FR and the BLR solutions along a shot profile intersecting the receiver position is also shown. Three conclusions can be drawn for this case study: For these values of  $\epsilon$ , the magnitude of the errors generated by the BLR approximation relative to the reference FR solutions is small. Second, these relative errors mainly concern the amplitude of the wavefields, not the phase. Third, for a given value of  $\epsilon$ , the magnitude of the errors decreases with the frequency. This last statement can be more quantitatively measured by the ratio between the scaled residual obtained with the BLR and the FR solver, where the scaled residual is given by  $\delta = \|A_h \tilde{p}_h - b\|_\infty / \|A_h\|_\infty \|\tilde{p}_h\|_\infty$  and  $\tilde{p}_h$  denotes the computed solution. We show that, for a given value of  $\epsilon$ ,  $\delta_{\text{BLR}}/\delta_{\text{FR}}$  decreases with frequency (Table 2).

### FWI results

The FWI results obtained with the FR solver are shown in Figures 6e–6h, 7b, and 7e. Comparison between the initial model and the FWI model highlights the resolution improvement achieved by FWI. The structures reviewed below are also described by [Sirgue et al. \(2010\)](#), [Barkved et al. \(2010\)](#), and [Operto et al. \(2015\)](#) for comparison. The depth slice at 175 m depth shows high-velocity

glacial sand channel deposits as well as small-scale low-velocity anomalies (Figure 6e). The depth slice at 500 m depth shows linear structures interpreted as scrapes left by drifting icebergs on the paleo seafloor as well as a wide low-velocity zone (Figure 6f) represented by a horizontal reflector in the vertical sections (Figure 7b and 7e, black arrow). The depth slice at 1 km depth (Figure 6g) and the inline vertical section at  $x = 5575$  m (Figure 7e) crosscuts the gas cloud, whose geometry has been nicely refined by FWI. We also show a inline vertical section at  $x = 5250$  m near the periphery of the gas cloud (Figure 7b), which highlights some low-velocity subvertical structures also identifiable in the 1 km depth slice (Figure 6g). The depth slice at 3.35 km depth crosscuts the base cretaceous reflector ([Barkved et al., 2010](#)) whose geometry is highlighted by the white arrows in the vertical sections (Figure 7b and 7e).

The final FWI model obtained with the BLR solver ( $\epsilon = 10^{-3}$ ) shown in Figures 6i–6l, 7c, and 7f does not show any obvious differences from the one obtained with the FR solver (Figures 6e–6h, 7b, and 7e). This is indeed also the case when the BLR solver is used with  $\epsilon = 10^{-4}$  and  $10^{-5}$  (not shown here).

The data fit achieved with the BLR solver ( $\epsilon = 10^{-3}$ ) is illustrated in Figure 11 for the receiver, the position of which is given in Figure 5a. The figure shows the real 5, 7, and 10 Hz monochromatic receiver gathers, the modeled ones computed in the FWI model inferred from the inversion of the frequency in question and the difference between the two. We also show a direct comparison between the recorded and modeled wavefields along the dip and cross profiles intersecting the position of the receiver. The data fit is

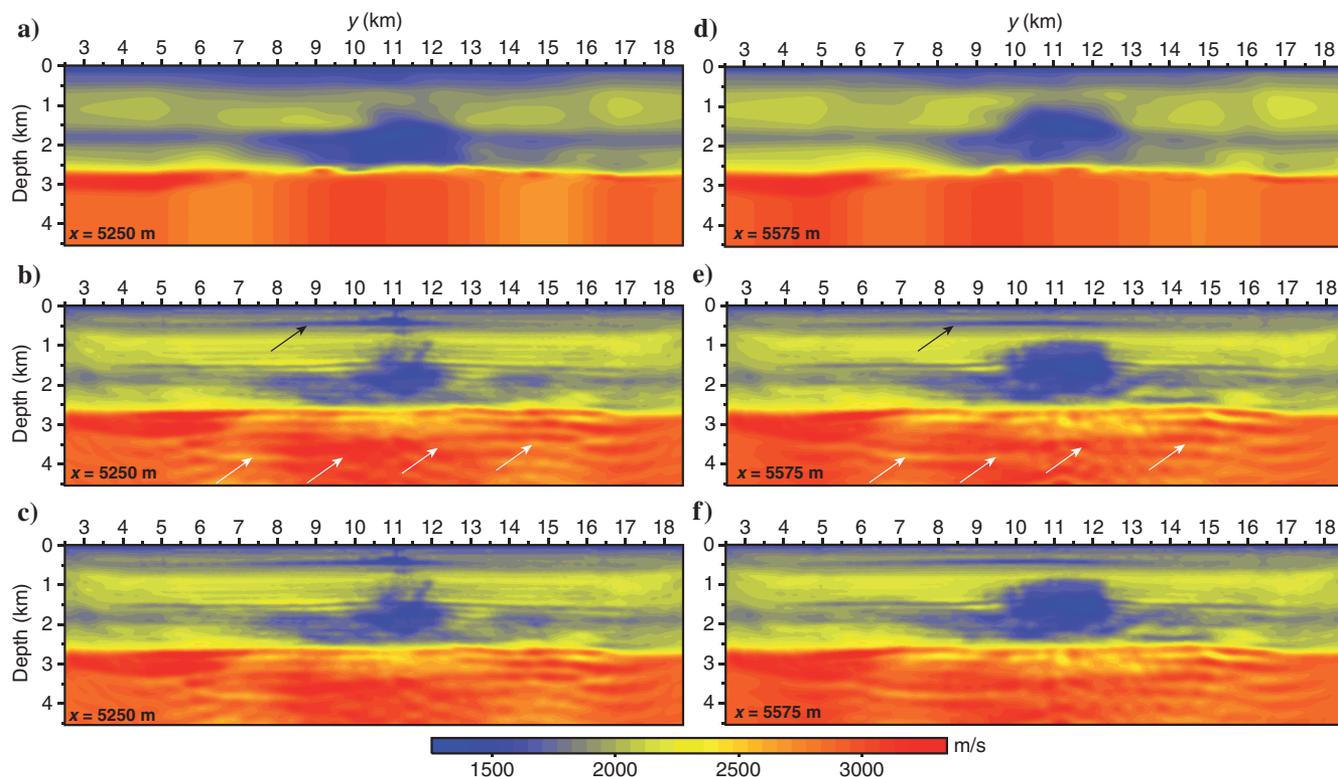


Figure 7. North Sea case study. (a and d) Vertical slices extracted from the initial model at (a)  $x = 5250$  m and (d) 5575 m. The slice in panel (d) crosscuts the gas cloud, whereas the slice in panel (a) is located in its periphery (see Figure 6). (b and e) Same as panels (a and d) for the final FWI model obtained with the BLR solver. (c and f) Same as panels (b and e) for the final FWI model obtained with the FR solver ( $\epsilon = 10^{-3}$ ). The back arrow points to a low-velocity reflector at 500 m depth (see Figure 6f and 6j for the lateral extension of this reflector at 500 m depth), whereas the white arrows point to the base Cretaceous reflector.

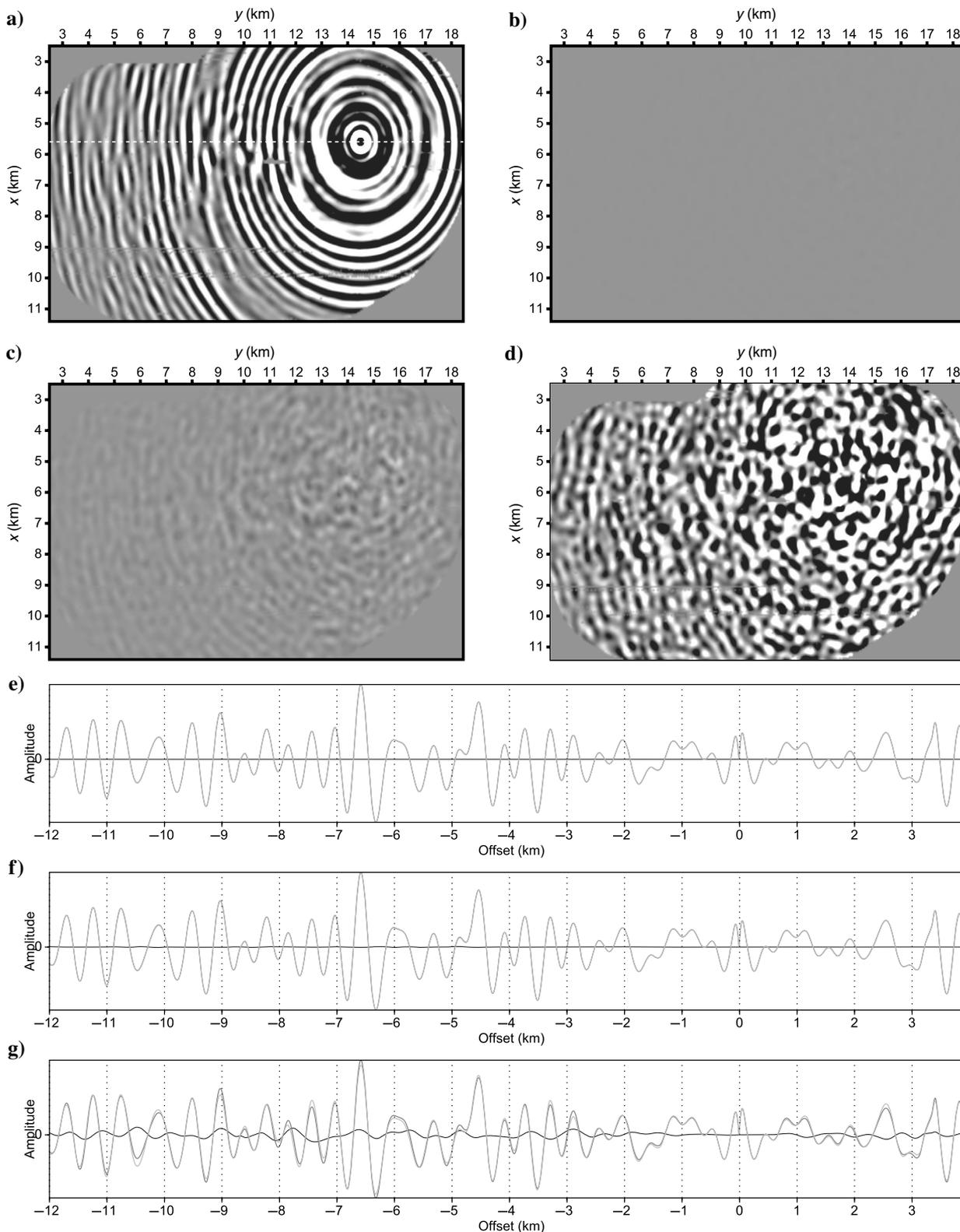


Figure 8. North Sea case study. The BLR modeling errors. (a) 5 Hz receiver gather (real part) computed with the FR solver. (b) Difference between the receiver gathers computed with the BLR ( $\epsilon = 10^{-5}$ ) and the (a) FR solvers. (c) Same as panel (b) for  $\epsilon = 10^{-4}$ . (d) Same as panel (b) for  $\epsilon = 10^{-3}$ . Residual wavefields in panels (b-d) are multiplied by a factor of 10 before the plot. The (a) FR wavefield and (b-d) the residual wavefields after multiplication by a factor of 10 are plotted with the same amplitude scale defined by a percentage of clip equal to 85 of the (a) FR-wavefield amplitudes. (e-g) Direct comparison between the wavefields computed with the FR (dark gray) and the BLR solvers (light gray) for (e)  $\epsilon = 10^{-5}$ , (f)  $10^{-4}$ , and (g)  $10^{-3}$  along an  $x$  profile intersecting the receiver position (dash line in panel (a)). The difference is shown by the thin black line. Amplitudes are scaled by a linear gain with offset.

very similar to the one achieved with the FR solver (not shown here, see [Operto et al. \[2015\]](#), their Figures 15–17) and is quite satisfactory, in particular in terms of phase. We already noted that the modeled amplitudes tend to be overestimated at long offsets when the wavefield has propagated through the gas cloud in the dip direction, unlike in the cross direction (Figure 11b, ellipse). In [Operto et al. \(2015\)](#), we interpret these amplitude mismatches as the footprint of attenuation, whose absorption effects have been underestimated during seismic modeling with a uniform  $Q$  equal to 200. The misfit functions versus the iteration number obtained with the FR and BLR ( $\epsilon = 10^{-5}$ ,  $10^{-4}$ , and  $10^{-3}$ ) solvers for the six frequencies are shown in Figure 12. The convergence curves obtained with the FR and the BLR solvers for  $\epsilon = 10^{-4}$  and  $10^{-5}$  are very similar. In contrast, we show that the convergence achieved by the BLR solver with  $\epsilon = 10^{-3}$  is alternatively better (4 and 5 Hz) and worse (3.5 and 4.5 Hz) than for the three other FWI run when the inversion jumps from one frequency to the next within the 3.5–5 Hz frequency band. However, for the last two frequencies (7 and 10 Hz) that have the best signal-to-noise ratio, all of the four convergence curves show a similar trend and reach a similar misfit function value at the last iteration. This suggests that the crude stopping criterion of iteration that is used in this study by fixing a common maximum iteration count is reasonable for a fair comparison of the computational cost of each FWI run. The different convergence behavior at low frequencies shown for  $\epsilon = 10^{-3}$  probably reflects the sensitivity of the inversion to the noise introduced by the BLR approximation at low frequencies (Figure 8), although this noise remains sufficiently weak to not alter the FWI results. The higher footprint of the BLR approximation at low frequencies during FWI is consistent with the former analysis of the relative modeling errors, which decrease as frequency increases (Table 2, ratio  $\delta_{\text{BLR}}/\delta_{\text{FR}}$ ).

#### Computational cost

The reduction of the size of the LU factors, operation count, and factorization time obtained with the BLR approximation are outlined in Table 3 for the 5, 7, and 10 Hz frequencies, respectively. Compared with the FR factorization, the size of the LU factors obtained with the BLR solver ( $\epsilon = 10^{-3}$ ) decreases by a factor of 2.6, 3.0, and 3.5 for the 5, 7, and 10 Hz frequencies, respectively (field  $FS_{\text{LU}}$  in Table 3). This can be converted to a reduction of the memory demand (ongoing work). Moreover, the number of flops during the LU factorization (field  $F_{\text{LU}}$  in Table 3)

decreases by factors of 8, 10.7, and 13.3 when the BLR solver is used. The increase of the computational saving achieved by the BLR solver, when the problem size grows, is further supported by the weak scalability analysis shown in the next subsection. When the BLR solver ( $\epsilon = 10^{-3}$ ) is used, the LU factorization time is decreased by a factor of 1.9, 2.7, and 2.7 with respect to the FR solver for the 5, 7, and 10 Hz frequencies, respectively (field  $T_{\text{LU}}$  in Table 3). The time reduction achieved by the BLR solver tends to increase with the frequency. This trend is also due to the increase of the workload per MPI process as the frequency increases. Increasing the workload per processor, which in our case was guided by memory constraints, tends to favor the parallel performance of the BLR solver.

As mentioned above, the BLR solver does not fully exploit the compression potential when multithread BLAS kernels are used, because low-rank matrices are smaller by nature than original FR blocks. This was confirmed experimentally on the 7 Hz problem with  $\epsilon = 10^{-3}$  by a computational time reduced by a factor 2.3 on 160 cores even though the flops are reduced by a factor of 11.1 (this is not reported in the tables). Adding OpenMP-based parallelism allows us to retrieve a substantial part of this potential, reaching a speedup of 3.8. In comparison with the timings reported by [Amestoy et al. \(2015c\)](#), the performance gains obtained by exploiting the sparsity of the right sides and the improved multithreading are approximately equal to a factor of 2.3. The elapsed time to compute one wavefield once the LU factorization has been performed is small (field  $T_s$  in Table 3). The two numbers provided in Table 3 for  $T_s$  are associated with the computation of the incident and adjoint wavefields. In the latter case, the source vectors are far less sparse, which leads to a computational overhead during the solution phase. These results in an elapsed time of, respectively, 262, 598, and 1542 s to compute the 4604 wavefields required for the computation of one FWI gradient (field  $T_{ms}$  in Table 3). Despite the high

**Table 2. North Sea case study. Modeling error introduced by BLR for different low-rank threshold  $\epsilon$  and different frequencies  $F$ . Here,  $\delta$ : scaled residuals defined as  $\|A_h \tilde{p}_h - b\|_{\infty} / \|A_h\|_{\infty} \|\tilde{p}_h\|_{\infty}$ , for  $b$  being for one of the RHS in  $B$ . The numbers in parentheses are  $\delta_{\text{BLR}}/\delta_{\text{FR}}$ . Note that, for a given  $\epsilon$ , this ratio decreases as frequency increases.**

$F(\text{Hz})/h(\text{m})$	$\delta(\text{FR})$	$\delta(\text{BLR}, \epsilon = 10^{-5})$	$\delta(\text{BLR}, \epsilon = 10^{-4})$	$\delta(\text{BLR}, \epsilon = 10^{-3})$
5 Hz/70 m	$2.3 \times 10^{-7}(1)$	$4.6 \times 10^{-6}(20)$	$6.7 \times 10^{-5}(291)$	$5.3 \times 10^{-4}(2292)$
7 Hz/50 m	$7.5 \times 10^{-7}(1)$	$4.6 \times 10^{-6}(6)$	$6.9 \times 10^{-5}(92)$	$7.5 \times 10^{-4}(1000)$
10 Hz/35 m	$1.3 \times 10^{-6}(1)$	$2.9 \times 10^{-6}(2.3)$	$3.0 \times 10^{-5}(23)$	$4.3 \times 10^{-4}(331)$

**Table 1. North Sea case study. Problem size and computational resources (for the results of Tables 3 and 4 only);  $h(m)$ , grid interval;  $n_{\text{PML}}$ , number of grid points in absorbing perfectly matched layers;  $u(10^6)$ , number of unknowns;  $n$ , number of computer nodes; MPI, number of MPI process; th, number of threads per MPI process;  $c$ , number of cores; and RHS, number of right sides processed per FWI gradient.**

Frequencies (Hz)	$H(m)$	Grid dimensions	$n_{\text{PML}}$	$u$	$n$	MPI	th	$c$	RHS
3.5, 4, 4.5, 5	70	$66 \times 130 \times 230$	8	2.9	12	24	10	240	4604
7	50	$92 \times 181 \times 321$	8	7.2	16	32	10	320	4604
10	35	$131 \times 258 \times 458$	4	17.4	34	68	10	680	4604

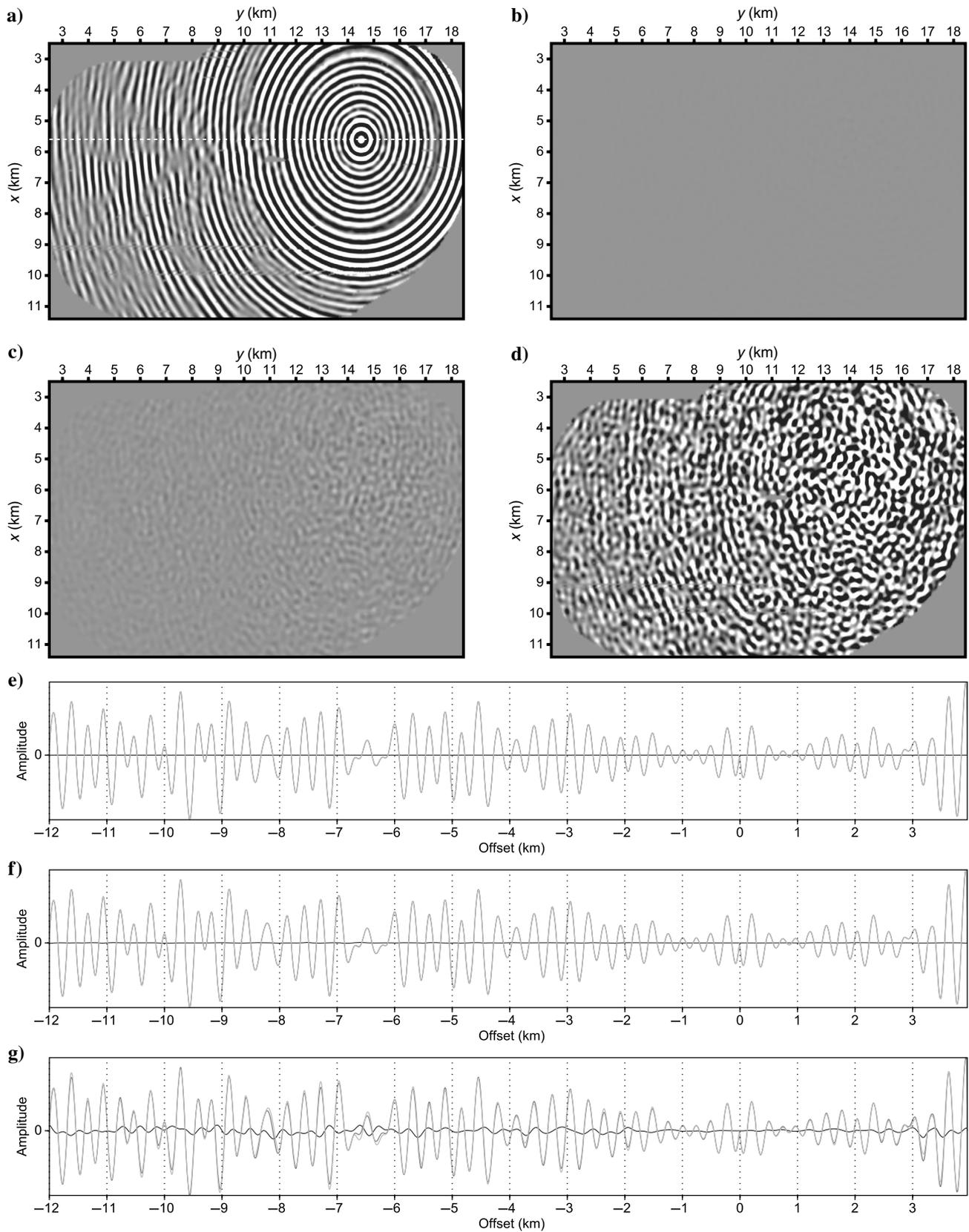


Figure 9. North Sea case study. The BLR modeling errors. Same as Figure 8 for the 7 Hz frequency. The simulations are performed in the same subsurface model obtained after a 7 Hz inversion (not shown here). The same percentage of clip (85%) of the FR-wavefield amplitudes and the same amplitude scaling of the residuals wavefields (multiplication by a factor of 10 before plot) as those used in Figure 8 are used for plot.

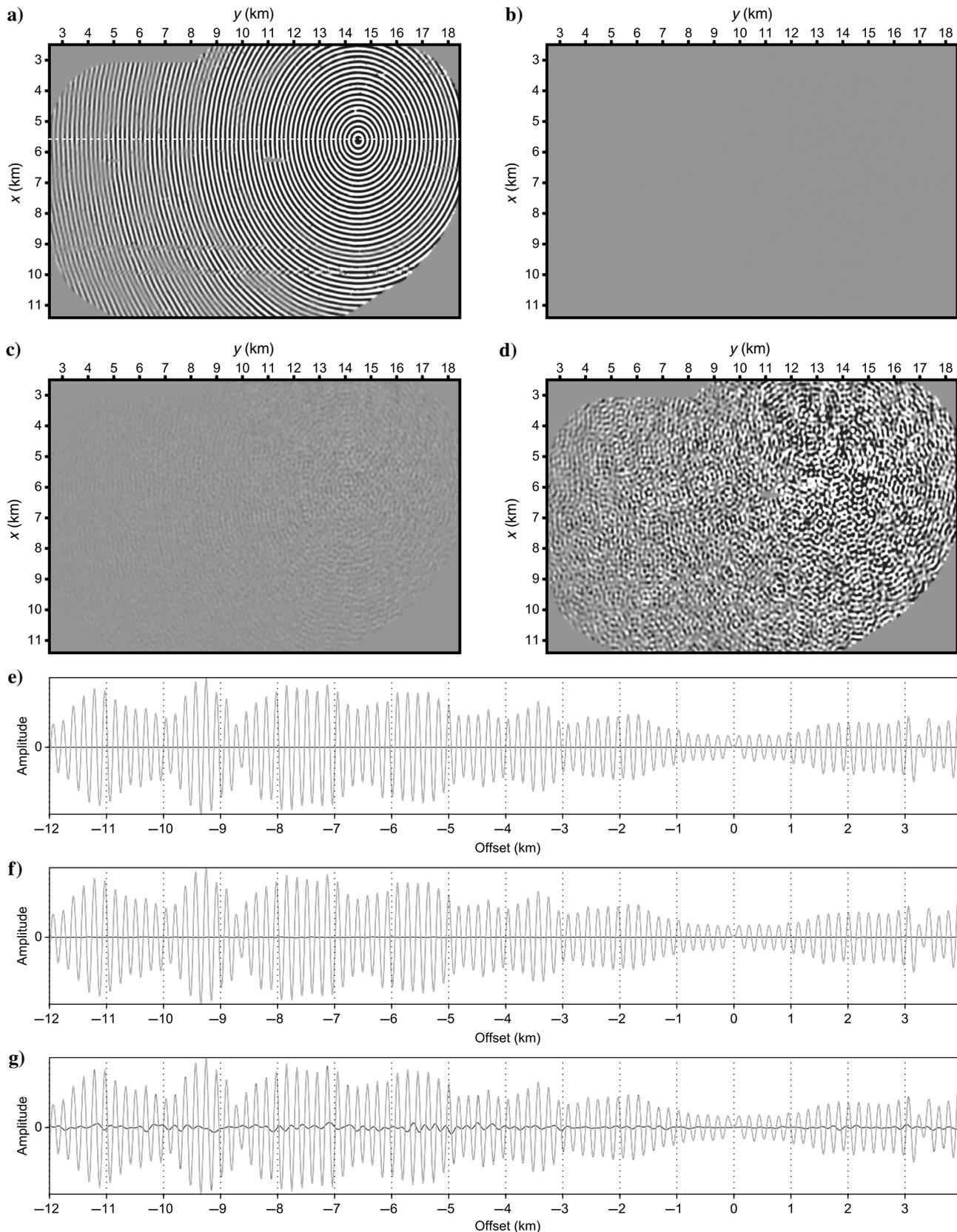


Figure 10. North Sea case study. The BLR modeling errors. Same as Figure 8 for the 10 Hz frequency. The simulations are performed in the same subsurface model obtained after a 10 Hz inversion. The same percentage of clip (85%) of the FR-wavefield amplitudes and the same amplitude scaling of the residuals wavefields (multiplication by a factor of 10 before plot) as those used in Figure 8 are used for plot.

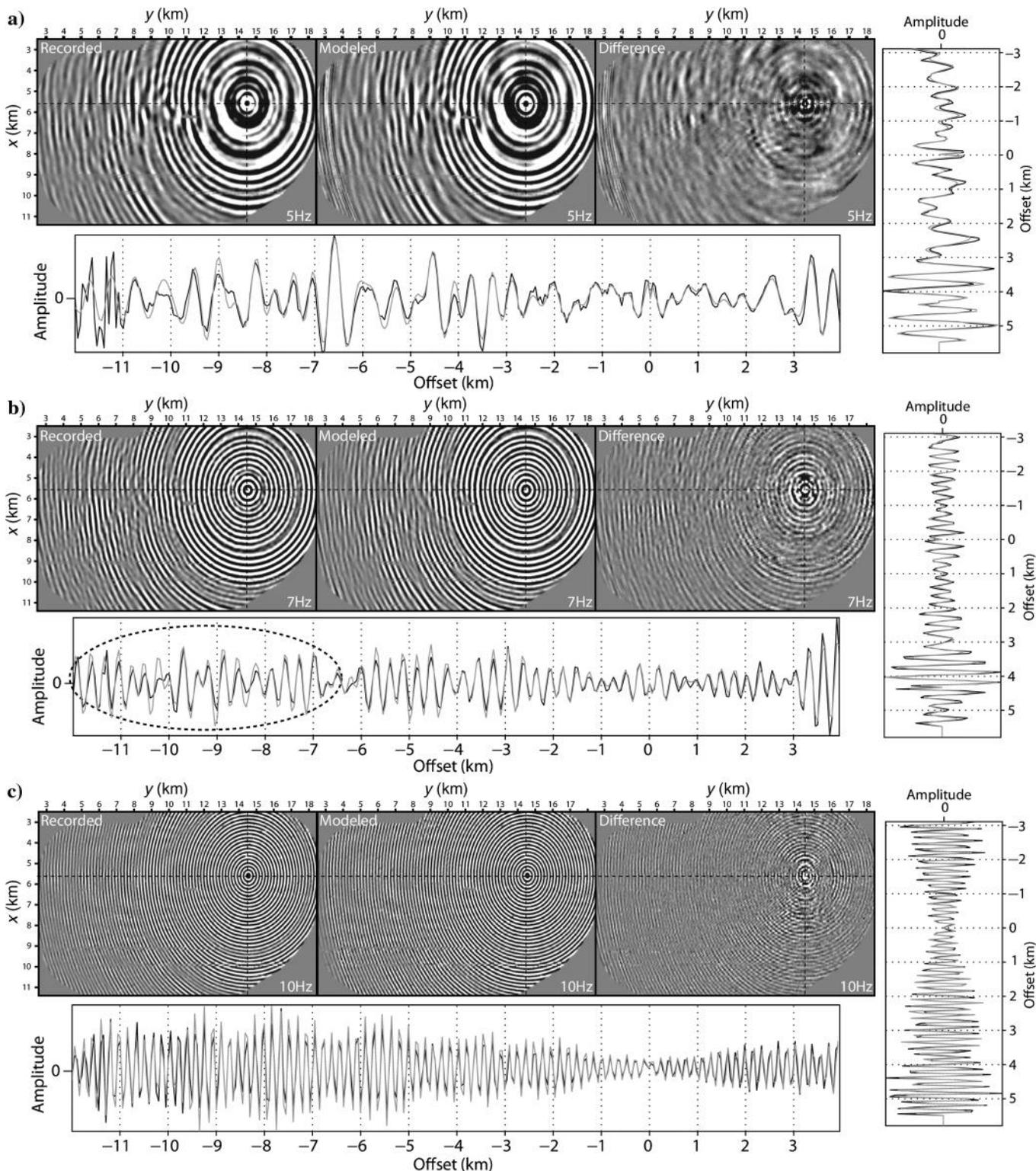


Figure 11. North Sea case study. Data fit achieved with the BLR solver ( $\epsilon = 10^{-3}$ ): (a) 5, (b) 7, and (c) 10 Hz frequencies. Left ( $x$ - $y$ ) panel shows the recorded data. Middle panel shows the modeled data computed in the FWI model inferred from the inversion of the current frequency. Right panel is the difference. Bottom and right amplitude-offset panels show a direct comparison between the recorded (black) and the modeled (gray) data (real part) along a dip and across profiles intersecting the receiver position (dash lines in  $(x$ - $y$ ) panels). Amplitudes are corrected for geometric spreading. The ellipse delineates an offset range, for which modeled amplitudes are overestimated (see text for interpretation).

efficiency of the substitution step, the elapsed time required to compute the wavefield solutions by substitution (262, 598, and 1542 s) is significantly higher than the time required to perform the LU factorization (41, 121, and 424 s with the BLR solver  $\epsilon = 10^{-3}$ ), when all the (reciprocal) sources are processed at each FWI iteration. However, the rate of increase of the solution step is smaller than the linear increase of the factor size and factorization time when increasing the frequency. In other words, the real-time complexity of the LU factorization is higher than that of the solution phase, although the theoretical time complexities are the same for  $N^2$  right sides ( $\mathcal{O}(N^6)$ ). This is shown by the decrease of the ratio  $T_{ms}/T_{LU}$  as the problem size increases (3.4, 1.86, and 1.34 for the 70, 50, and 35 m grids, respectively, when the FR solver is used). The fact that the speedup of the factorization phase achieved by the BLR approximation increases with the problem size, will balance the higher complexity of the LU factorization relative to the solution phase and help to address large-scale problems. As reported in col-

umn  $T_g$ (mn) of Table 3, the elapsed times required to compute one gradient with the FR solver are of the order of 9.9, 21.2, and 69 min for the 5, 7, and 10 Hz frequencies, respectively, whereas those with the BLR solver are of the order of 9.1, 17.1, and 49.5 min, respectively. Please note that FR and BLR timings already include the acceleration due to the exploitation of the sparsity of the right sides. Thus, the difference between these two times reflects the computational saving achieved by the BLR approximation during the LU factorization, because the BLR approximation is currently exploited only during this task.

For a fair assessment of the FWI speedup provided by the BLR solver, it is also important to check the impact of the BLR approximation on the line search, and hence the number of FWI gradients computed during FWI (Table 4). On the 70 m grid where the impact of the BLR errors is expected to be the largest one (as indicated in Table 2), the inversion with the BLR solver ( $\epsilon = 10^{-3}$ ) computes 82 gradients against 77 gradients for the three other settings

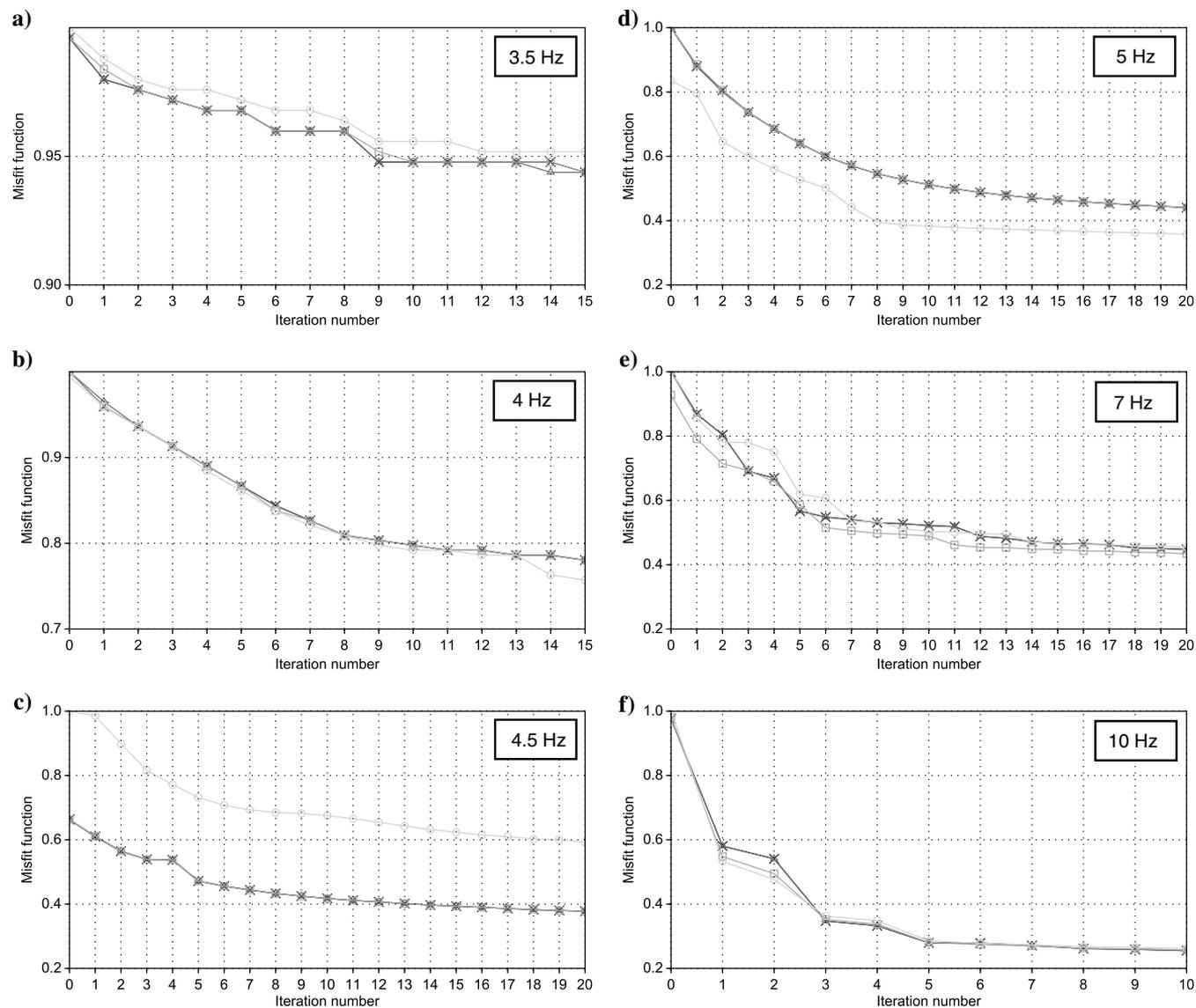


Figure 12. North Sea case study. Misfit function versus iteration number achieved with the FR (black cross) and the BLR solvers with  $\epsilon = 10^{-5}$  (dark gray triangle),  $10^{-4}$  (gray square), and  $10^{-3}$  (light gray circle): (a) 3.5, (b) 4, (c) 4.5, (d) 5, (e) 7, and (f) 10 Hz inversion.

(FR and BLR solvers with  $\epsilon = 10^{-4}$  and  $10^{-5}$ ), for a total of 70 FWI iterations. On the 7 Hz grid, the FR and the BLR solvers with  $\epsilon = 10^{-5}$  compute 39 gradients against 30 gradients with the BLR solver with  $\epsilon = 10^{-4}$  and  $10^{-3}$  for a total of 20 FWI iterations. On the 10 Hz grid, the four inversions compute 16 gradients for a total of 10 FWI iterations.

The elapsed time to perform the FWI is provided for each grid in Table 4. The entire FWI application takes 49.2, 39.4, 36, and 37.8 h with the FR solver and the BLR solver with  $\epsilon = 10^{-5}$ ,  $10^{-4}$ , and  $10^{-3}$ , respectively. We remind that the FWI application performed with the BLR solver with  $\epsilon = 10^{-3}$  takes more time than with  $\epsilon = 10^{-4}$  because more gradients were computed on the 70 m grid. We conclude from this analysis that, for this case study, the BLR solver with  $\epsilon = 10^{-4}$  provides the best trade-off between the number of FWI gradients required to reach a given value of the misfit function and the computational cost of one gradient computation at low frequencies. At the 7 and 10 Hz frequencies, the BLR solver with  $\epsilon = 10^{-3}$  provides the smaller computational cost without impacting the quality of the FWI results.

### Strong and weak scalability of the block low-rank multifrontal solver

We now present a weak and strong scalability analysis of the FR and BLR solvers using the FWI subsurface models described in the previous sections.

For the strong scalability analysis, we use a subsurface model that has been obtained by FWI for the 7 Hz frequency. It is reminded that the grid spacing is 50 m leading to 7.2 million unknowns in the linear system (Table 1). We perform seismic modeling in this model with the FR and BLR ( $\epsilon = 10^{-3}$ ) solvers by multiplying by a factor of two, the number of cores from one run to the next: 40, 80, 160, 320, and 640 cores (Table 5 and Figure 13). Therefore, the ideal acceleration from one configuration to the next is two (represented

by the dashed lines in Figure 13). We obtain an average acceleration of 1.5 and 1.4 with the FR and BLR solvers, respectively. The scalability of the FR solver is satisfactory on 320 cores (speedup of 1.8 from 160 to 320 cores, Table 5). The 7 Hz matrix is, however, too small to provide enough parallelism, and thus we only reach a speedup of 1.4 from 320 to 640 cores. Moreover, we show in Figure 13 that the difference between the FR and BLR execution times decreases as the number of cores increases because the BLR solver performs fewer flops with a smaller granularity than the FR solver, and thus the relative weight of communications and noncomputational tasks (e.g., memory copies and assembly) becomes more important in BLR. Indeed, although the elapsed time for the FR LU factorization is 3.9 times higher than the one for the BLR LU factorization when 40 cores are used, this ratio decreases to 3 on 320 cores and 2.7 on 640 cores. Therefore, the strong scalability of the FR solver and the efficiency of the BLR solver relative to the FR solver should be taken into account when choosing the computational resources that are used for an application. Finally, the flop compression rate achieved by BLR is provided in Figure 13 on top of each point. It has the desirable property to remain roughly constant, when the number of processes grows; i.e., the BLR compression does not degrade on higher core counts.

A weak scalability analysis is shown in Table 6 to assess the computational saving achieved by the BLR solver, when the problem size (i.e., the frequency) grows. We used three matrices generated from subsurface models obtained by FWI for the 5, 7, and 10 Hz frequencies. For these three frequencies, the grid interval is 70, 50, and 35 m leading to 2.9, 7.2, and 17.4 million of unknowns, respectively (Table 1). The BLR threshold is set to  $\epsilon = 10^{-3}$ . The number of MPI processes is chosen to keep the memory demand of the LU factorization per processor of the order of 15 GB. This leads to 6, 20, and 64 MPI processes and to  $1.1 \times 10^{13}$ ,  $2.1 \times 10^{13}$ , and  $4.1 \times 10^{13}$  flops/MPI process (for the FR solver) for the 5, 7, and 10 Hz matrices, respectively. Because the number of flops per MPI process

**Table 3. North Sea case study. Computational savings provided by the BLR solver during the factorization step. Factor of improvement due to BLR is indicated between parentheses. The elapsed times required to perform the multi-RHS substitution step and to compute the gradient are also provided.  $FS_{LU}$ (GB): size of LU factors (GigaBytes);  $F_{LU}$ : flops for the LU factorization;  $T_{LU}$ (s): elapsed time for the LU factorization;  $T_s$ (s): average time for one solution. The first and second numbers are related to the incident and adjoint wavefields, respectively. The acceleration from the exploitation of the RHS sparsity is included. Also,  $T_{ms}$ (s): elapsed time for 4604 solutions (incident + adjoint wavefields);  $T_g$ (mn): elapsed time to compute the FWI gradient. This time also includes the IO tasks.**

$F$ (Hz)/ $h$ (m)	$\epsilon$	$FS_{LU}$ (GB)	$F_{LU}(\times 10^{12})$	$T_{LU}$ (s)	$T_s$ (s)	$T_{ms}$ (s)	$T_g$ (mn)
5 Hz/70 m (240 cores)	FR	62 (1.0)	66 (1.0)	78 (1.0)	0.051/0.063	262	9.9
	$10^{-5}$	35 (1.8)	17 (3.8)	48 (1.6)			9.4
	$10^{-4}$	30 (2.1)	12 (5.3)	46 (1.7)			9.1
	$10^{-3}$	24 (2.6)	8 (8.0)	41 (1.9)			9.1
7 Hz/50 m (320 cores)	FR	211 (1.0)	410 (1.0)	322 (1.0)	0.12/0.14	598	21.2
	$10^{-5}$	90 (2.3)	90 (4.5)	157 (2.1)			17.7
	$10^{-4}$	88 (2.4)	63 (6.5)	136 (2.4)			17.3
	$10^{-3}$	70 (3.0)	38 (10.7)	121 (2.7)			17.1
10 Hz/35 m (680 cores)	FR	722 (1.0)	2600 (1.0)	1153 (1.0)	0.26/0.41	1542	69.0
	$10^{-5}$	333 (2.2)	520 (4.9)	503 (2.3)			48.6
	$10^{-4}$	271 (2.7)	340 (7.5)	442 (2.6)			48.9
	$10^{-3}$	209 (3.5)	190 (13.3)	424 (2.7)			49.5

grows in this parallel configuration, the execution time increases with the matrix size accordingly (Table 6). This allows the FR factorization to maintain a high Gflops/s/core on higher core counts (5.2, 4.1, and 3.5 Gflops/s/core, respectively, to compare with a peak of 10 Gflops/s/core), which corresponds to a parallel efficiency (normalized with respect to the 5 Hz problem) of 0.80 and 0.69 for the 7 and 10 Hz problems. In comparison, the normalized efficiency of the BLR factorization is 0.72 and 0.52 for the 7 and 10 Hz problems. Even though the efficiency decreases slightly faster in BLR than in FR, the flop reduction due to BLR also increases with the frequency (12.4%, 9.0%, and 7.3%, respectively), which in the end leads to a time reduction by a factor of 2.7, 3.4, and 3.5, respectively. Thus, even though the flop reduction is not fully translated into time, this weak scaling study shows that the gain due to BLR can be maintained on higher core counts with a significant reduction factor of the order of three in time for the current BLR implementation.

## DISCUSSION

We have shown that 3D viscoacoustic VTI frequency-domain FWI allows for efficiently building a subsurface model from stationary recording systems, such as OBC with limited computational resources, when the linear system resulting from the discretization of the time-harmonic wave equation is solved with a sparse direct solver. Two key ingredients in the direct solver were implemented to achieve high computational performance for FWI application: The first one exploits the low-rank properties of the elliptic partial differential operators embedded in the time-harmonic wave equation that allows us to reduce the cost of the LU factorization by a factor of approximately three in terms of computation time and factor size; the second one exploits the sparsity of the seismic source vectors to speed up the forward elimination step during the computation of the solutions. Note that the BLR approximation has so far been implemented in the LU factorization only; i.e., the FR uncompressed LU factors are still used to compute the wavefield solutions by substitution. Therefore, there is a potential to accelerate the solution step, once the compression of the LU factors achieved by the BLR approximation is exploited. More generally, we believe that in the context of the solution phase with large numbers of right-hand sides, there is still much scope for improvement to better exploit the parallelism of our target computers. We will investigate in priority this issue because it is critical. In parallel to the work on the theoretical complexity bounds of the BLR factorization (Amestoy et al., 2016b), we are investigating variants of the BLR factorization improving its performance and complexity. Another possible object of research is the use of BLR as a preconditioner rather than a direct solver. Preliminary experiments in Weisbecker (2013) show promising results, although the time to solve

( $T_s$  in Table 3), which is the bottleneck in this applicative context, would be multiplied by the number of iterations.

The choice of the low-rank threshold is performed by trial and error. However, the analysis of the modeling errors introduced by the BLR approximation and their impact on the convergence behavior of the FWI support that the BLR approximation has a stronger footprint at low frequencies for a given subsurface target. Therefore, a general strategy might be to use a slightly more accurate factorization (smaller  $\epsilon$ ) at low frequencies, when the FWI is not expensive and use a slightly more aggressive thresholding (higher  $\epsilon$ ) at higher frequencies. For the case study presented here, a value of  $\epsilon = 10^{-3}$  does not impact upon the quality of the FWI results at least at the 7 and 10 Hz frequencies. Another case study (not shown here) performed with a smaller target, a portion of the 3D SEG/EAGE land overthrust model, shows a stronger footprint of the BLR approximation in the FWI results for  $\epsilon = 10^{-3}$ . Additional work is still necessary to estimate in a more automated way the optimal choice of  $\epsilon$ . Furthermore, the influence of the frequency on the error (FR and

**Table 4. North Sea case study. The FWI cost;  $it$  and  $g$  are the number of FWI iterations and the number of gradients computed on each grid;  $T_{FWI}$  is the elapsed time for FWI on each grid. The total times for the FWI application are 49.2, 39.4, 36, and 37.8 h for the FR, BLR( $10^{-5}$ ), BLR( $10^{-4}$ ), and BLR( $10^{-3}$ ) solvers, respectively.**

$h$ (m)	Frequency (Hz)	$c$	$\epsilon$	$it$	$g$	$T_{FWI}$ (h)
70	3.5, 4, 4.5, 5	240	FR	70	77	14.0
			$10^{-5}$	70	77	12.9
			$10^{-4}$	70	77	12.7
			$10^{-3}$	70	82	14.0
50	7	320	FR	20	39	14.5
			$10^{-5}$	20	39	12.0
			$10^{-4}$	20	30	9.1
			$10^{-3}$	20	30	9.0
35	10	680	FR	10	16	20.7
			$10^{-5}$	10	16	14.5
			$10^{-4}$	10	16	14.2
			$10^{-3}$	10	16	14.8

**Table 5. North Sea case study. Strong scaling analysis of the factorization of the 7 Hz problem. For BLR, the threshold is set to  $\epsilon = 10^{-3}$ ; AccFR/LR ( $X \rightarrow 2X$ ) is the acceleration factor (i.e., ratio of time  $X$  and over time  $2X$ ) obtained by doubling the number of processes. See also Figure 13.**

	Time (s) FR	Time (s) BLR	Ratio FR/BLR	AccFR ( $X \rightarrow 2X$ )	AccLR ( $X \rightarrow 2X$ )
$4 \times 10$	1323	336	3.9	1.5	1.4
$8 \times 10$	865	241	3.6	1.5	1.6
$16 \times 10$	574	151	3.8	1.8	1.4
$32 \times 10$	327	110	3.0	1.4	1.3
$64 \times 10$	237	87	2.7	—	—

with BLR) deserves to be more thoroughly investigated, from an experimental and a theoretical standpoint.

## CONCLUSION

Although 3D frequency-domain FWI based on sparse direct methods is generally considered intractable, we have shown in this study its high computational efficiency to process OBC data in the viscoacoustic VTI approximation with quite limited computational resources. The computational efficiency of the frequency-domain FWI relies on a suitable choice of a few discrete frequencies and recent methodological advances in the development of massively parallel sparse multifrontal solver. The first advance exploits some low-rank properties of the elliptic partial differential operators embedded in the time harmonic wave equation, which allows for a significant reduction in operation count and computation time during the LU factorization. This compression procedure, based on the so-called BLR format and the related approximation, should lead to a speedup of the solution step too, although this is not yet implemented. The second advance exploits the sparsity of the right sides, namely, the seismic source vectors, to speed up the forward elimi-

nation step during the wavefield computations. All these features lead to a quite efficient computation of wavefield solutions by substitution once the LU factorization has been performed at a very reasonable cost. This offers a suitable framework to preserve the fold resulting from dense seismic acquisitions during the stack procedure underlying FWI and hence build subsurface model with a high signal-to-noise ratio. Although the FWI was limited to a maximum frequency of 10 Hz, it is probably reasonable to try to push the inversion up to a frequency of 15 Hz in the near future. For the case study presented here, this would require managing computational grids with up to 60 million of unknowns. On the other hand, although the frequency decimation underlying efficient frequency-domain FWI is relatively neutral for monoparameter FWI, the impact of this decimation will have to be assessed with care in the framework of multiparameter reconstruction. This comment, however, also applies to time-domain FWI, which is generally applied to a quite narrow frequency bandwidth. Frequency-domain FWI remains limited to a relatively narrow range of applications in terms of wave physics and acquisition geometries. Extension of our finite-difference stencil to tilted transverse isotropy (TTI) is not straightforward and should lead to a significant computational overhead. Application to short-spread narrow-azimuth streamer data might not be beneficial because the cost of the LU factorizations might become prohibitive relative to one of the solution steps and the number of frequencies to be managed should be increased to prevent wraparound artifacts. Despite these limitations, 3D frequency-domain FWI on OBC data based on a sparse direct solver can also be viewed as an efficient tool to build an initial viscoacoustic VTI subsurface model for subsequent elastic FWI of multicomponent data. Feasibility of frequency-domain viscoelastic modeling based on sparse direct solver for multicomponent/multiparameter FWI applications at low frequencies needs to be assessed and will be the aim of future work.

## ACKNOWLEDGMENTS

This study was partly funded by the SEISCOPE consortium <http://seiscope2.osug.fr>. This study was granted access to the high-performance computing facilities of the SIGAMM (Observatoire de la Côte d'Azur) and the HPC resources of CINES/IDRIS under the allocation 046091 made by GENCI. We also thank BP Norge AS and their Valhall partner Hess Norge AS for allowing access to the Valhall data set as well as the well-log velocities. We thank J. Virieux (ISTERRE) for his reading of the manuscript. We thank the associate editor A. Guitton, J. de la Puente, J. D. De

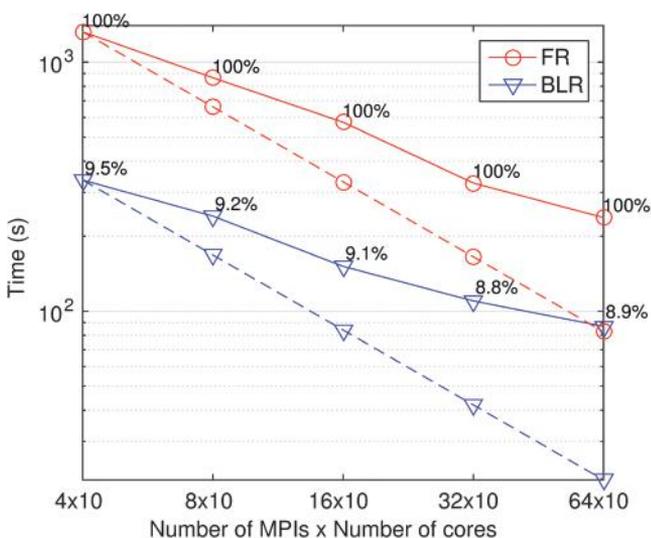


Figure 13. North Sea case study. Strong scaling analysis on the 7 Hz problem. For BLR, the threshold is set to  $\epsilon = 10^{-3}$ . The values next to the markers show the flops compression rate with respect to the FR factorization. See also Table 5.

Table 6. North Sea case study. Weak scaling analysis of the factorization on the 5, 7, and 10 Hz problems. For BLR, the threshold is set to  $\epsilon = 10^{-3}$ . The normalized efficiency is the parallel efficiency normalized with respect to the 5 Hz problem run on 6 x 10 cores. The peak is 10 Gflops/s/core.

	Time (s)		Ratio FR/BLR	Flops FR ( $\times 10^{12}$ )	Flops BLR/FR	Gflops/s/core		Normalized efficiency	
	FR	BLR				FR	BLR	FR	BLR
5 Hz(6 x 10)	210	77	2.7	66	12.4%	5.2	1.8	1	1
7 Hz(20 x 10)	489	144	3.4	420	9.0%	4.1	1.3	0.80	0.72
10 Hz(64 x 10)	1132	324	3.5	2624	7.3%	3.5	0.9	0.69	0.52

Basabe, and three anonymous reviewers for their constructive comments, which contributed to the improvement of the manuscript.

## APPENDIX A

### ALGORITHMIC ASPECTS OF FREQUENCY-DOMAIN FWI

The frequency-domain FWI is implemented with the SEISCOPE optimization toolbox, which relies on a reverse-communication interface with the user (Métivier and Brossier, 2015). The SEISCOPE optimization toolbox contains several optimization algorithms such as steepest-descent, conjugate gradient, limited-memory quasi Newton, and truncated Newton optimization as well as a line search that guarantees that the Wolfe conditions are satisfied during the estimation of the step length (Nocedal and Wright, 2006). The descent direction provided by each of these optimization algorithms can be improved by a preconditioner provided by the user. For the most basic steepest-descent optimization, the user provides to the optimization toolbox the current subsurface model, the misfit function computed in this model, and its gradient. In this study, the misfit function is the usual least-squares norm of the difference between the recorded and the modeled monochromatic data. The gradient is computed with the adjoint-state method that requires seismic modeling twice per (reciprocal) source (Plessix, 2006). The gradient of the misfit function for the VTI time-harmonic wave equation is developed in Operto et al. (2015), and is given by

$$\nabla C_m = \Re \left\{ \left( \frac{\partial \mathbf{A}_h}{\partial m} \mathbf{p}_h \right)^\dagger \mathbf{a}_1 \right\}, \quad (\text{A-1})$$

where the dagger indicates adjoint operator and adjoint wavefield  $a_1$  satisfies

$$\mathbf{A}_h \mathbf{a}_1 = \frac{1}{3} (\mathbf{A}_v + 2I) R^t \Delta \mathbf{d}. \quad (\text{A-2})$$

In equation A-2,  $\Delta \mathbf{d}$  are the data residuals,  $\Re$  denotes the real part of a complex number,  $I$  is the identity matrix, and  $R^t$  is a prolongation operator, which augments with zeros the data residual vector in the full computational domain (Pratt et al., 1998).

After the LU factorization of  $A_h$ , the LU factors are stored in main memory in distributed form. We then sequentially process  $N_a$  partitions of  $N_s$  seismic sources ( $N_s$  is the number of columns in matrix  $B$ , equation 7). The  $N_s$  right-hand sides are built in parallel by distributing them over the processors before their centralization on the master process with sparse storage. The seismic sources are positioned in the finite-difference grids with a windowed sinc parameterization (Hicks, 2002). The solution step performed by MUMPS returns the  $N_s$  monochromatic wavefields in distributed form following a domain decomposition driven by the distribution of the LU factors: each processor stores a subdomain of the  $N_s$  wavefields. Because the domain decomposition is driven by the LU factorization, it is unstructured with subdomains of quite heterogeneous sizes (Soubrier et al. [2009], their Figure 2). Therefore, it is inefficient for parallel gradient computation as it would lead to an unbalanced workload. Moreover, a domain decomposition of the gradient would complicate unnecessarily the implementation of the radiation pattern matrices  $\partial \mathbf{A}_h / \partial m$  (equation A-1) for secondary parameters as density and Thomsen's parameters as well as regu-

larizations. We therefore chose to redistribute the  $N_s$  wavefields over the processors according to the source index to implement an embarrassing MPI parallelism over sources: Each processor stores the full domain of  $N_s/N_p$  wavefields, where  $N_p$  denotes the number of MPI process.

Next, we interpolate the values of the incident wavefields at the receiver positions with the windowed sinc parameterization (Hicks, 2002), followed by the estimation of the source signature (Pratt, 1999) and the update of the misfit function. The source signature is averaged over the sources involved in the current partition. The same modeling workflow is applied to process the adjoint wavefields. All along the process, the  $N_s$  incident and adjoint wavefields remain in random-access memory (RAM) in distributed form according to the source index. The gradient of the misfit function can then be updated quite efficiently by combining the embarrassing parallelism over sources with shared-memory parallelism to process the degrees of freedom of the gradient grid. The gradient algorithm is above-described for one frequency. When several frequency components are simultaneously inverted, the multifrequency gradient is the sum of each monofrequency gradient. The outer loop over frequencies in the FWI algorithm can be either processed sequentially or in parallel with an additional MPI communication if enough computational resources are available.

A suitable architecture for frequency-domain FWI should take advantage of multicore central processing units (CPUs) with a significant amount of RAM. This significant amount of RAM allows for the simultaneous processing of a large number of right-hand sides in the matrix  $B$  (equation 7) taking advantage of multithreaded BLAS3 kernels and the sparsity of the source vectors. Note that because we can afford to keep in the core the incident wavefields all along the computation, they do not need to be recomputed during the adjoint simulation and the gradient computation, a distinct advantage compared with the time-domain implementation. To limit memory overheads and volume of communication during the LU factorization, we typically assign only one MPI process per socket or CPU, each MPI process being multithreaded over the cores of its CPU. This matches the architecture best, ensuring a uniform memory access within each MPI process, and in our experience, provides, in general, comparable or better performance during LU factorization and solving than using a single MPI process per node.

## REFERENCES

- Agullo, E., P. R. Amestoy, A. Buttari, A. Guermouche, J.-Y. L'Excellent, and F.-H. Rouet, 2016, Robust memory-aware mappings for parallel multifrontal factorizations: SIAM Journal on Scientific Computing, **38**, C256–C279.
- Amestoy, P. R., J. Anton, C. Ashcraft, A. Buttari, P. Ghysels, J.-Y. L'Excellent, X. S. Li, T. Mary, F.-H. Rouet, and C. Weisbecker, 2016a, A comparison of parallel rank-structured solvers: Presented at the SIAM Conference on Parallel Processing.
- Amestoy, P. R., C. Ashcraft, O. Boiteau, A. Buttari, J.-Y. L'Excellent, and C. Weisbecker, 2015a, Improving multifrontal methods by means of block low-rank representations: SIAM Journal on Scientific Computing, **37**, A1451–A1474, doi: 10.1137/120903476.
- Amestoy, P. R., A. Buttari, J.-Y. L'Excellent, and T. Mary, 2016b, Complexity and performance of the block low-rank multifrontal factorization: Presented at the SIAM Conference on Parallel Processing, SIAM PP16.
- Amestoy, P. R., R. Brossier, A. Buttari, J.-Y. L'Excellent, T. Mary, L. Métivier, A. Miniussi, S. Operto, J. Virieux, and C. Weisbecker, 2015b, 3D frequency-domain seismic modeling with a parallel BLR multifrontal direct solver: 85th Annual International Meeting, SEG, Expanded Abstracts, 3606–3611.
- Amestoy, P. R., R. Brossier, A. Buttari, J.-Y. L'Excellent, T. Mary, L. Métivier, A. Miniussi, S. Operto, A. Ribodetti, J. Virieux, and C. Weisbecker, 2015c, Efficient 3D frequency-domain full-waveform inversion of ocean-

- bottom cable data with sparse block low-rank direct solver: A real data case study from the North Sea: 85th Annual International Meeting, SEG, Expanded Abstracts, 1303–1308.
- Amestoy, P. R., I. S. Duff, J.-Y. L'Excellent, and F.-H. Rouet, 2015d, Parallel computation of entries of  $A^{-1}$ : *SIAM Journal on Scientific Computing*, **37**, C268–C284, doi: [10.1137/120902616](https://doi.org/10.1137/120902616).
- Amestoy, P. R., I. S. Duff, J. Koster, and J.-Y. L'Excellent, 2001, A fully asynchronous multifrontal solver using distributed dynamic scheduling: *SIAM Journal on Matrix Analysis and Applications*, **23**, 15–41, doi: [10.1137/S0895479899358194](https://doi.org/10.1137/S0895479899358194).
- Amestoy, P. R., A. Guermouche, J.-Y. L'Excellent, and S. Pralet, 2006, Hybrid scheduling for the parallel solution of linear systems: *Parallel Computing*, **32**, 136–156, doi: [10.1016/j.parco.2005.07.004](https://doi.org/10.1016/j.parco.2005.07.004).
- Aminfar, A., and E. Darve, 2016, A fast, memory efficient and robust sparse preconditioner based on a multifrontal approach with applications to finite-element matrices: *International Journal for Numerical Methods in Engineering*, **107**, 520–540.
- Barkved, O., U. Albertin, P. Heavey, J. Kommedal, J. van Gestel, R. Synove, H. Pettersen, and C. Kent, 2010, Business impact of full waveform inversion at Valhall: 91st Annual International Meeting and Exposition, SEG, Expanded Abstracts, 925–929.
- Barnes, C., and M. Charara, 2009, The domain of applicability of acoustic full-waveform inversion for marine seismic data: *Geophysics*, **74**, no. 6, WCC91–WCC103, doi: [10.1190/1.3250269](https://doi.org/10.1190/1.3250269).
- Bebendorf, M., 2004, Efficient inversion of the Galerkin matrix of general second-order elliptic operators with nonsmooth coefficients: *Mathematics of Computation*, **74**, 1179–1200, doi: [10.1090/S0025-5718-04-01716-8](https://doi.org/10.1090/S0025-5718-04-01716-8).
- Ben Hadj Ali, H., S. Operto, and J. Virieux, 2008, Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data: *Geophysics*, **73**, no. 5, VE101–VE117, doi: [10.1190/1.2957948](https://doi.org/10.1190/1.2957948).
- Béranger, J. P., 1996, Three-dimensional perfectly matched layer for the absorption of electromagnetic waves: *Journal of Computational Physics*, **127**, 363–379, doi: [10.1006/jcp.1996.0181](https://doi.org/10.1006/jcp.1996.0181).
- Brossier, R., V. Etienne, S. Operto, and J. Virieux, 2010, Frequency-domain numerical modeling of visco-acoustic waves based on finite-difference and finite-element discontinuous Galerkin methods, in D. W. Dissanayake, ed., *Acoustic waves: INTECH*, 125–158.
- Castagna, J. P., M. L. Batzle, and T. K. Kan, 1993, Rock physics — The link between rock properties and AVO response, in J. P. Castagna and M. M. Backus, eds., *Offset-dependent reflectivity-theory and practice of AVO analysis*: SEG.
- Coleman, T., and C. vanLoan, 1988, *Handbook of matrix computations*: SIAM.
- Duff, I. S., and J. K. Reid, 1983, The multifrontal solution of indefinite sparse symmetric linear systems: *ACM Transactions on Mathematical Software*, **9**, 302–325, doi: [10.1145/356044.356047](https://doi.org/10.1145/356044.356047).
- Erlangga, Y. A., and R. Nabben, 2008, On a multilevel Krylov method for the Helmholtz equation preconditioned by shifted Laplacian: *Electronic Transactions on Numerical Analysis*, **31**, 403–424.
- George, A., and J. W. Liu, 1981, *Computer solution of large sparse positive definite systems*: Prentice-Hall Inc.
- Ghysels, P., X. S. Li, F. Rouet, S. Williams, and A. Napov, 2015, An efficient multi-core implementation of a novel HSS-structured multifrontal solver using randomized sampling: *CoRR*, <http://arxiv.org/abs/1502.07405v1>, accessed 29 August 2016.
- Gilbert, J. R., and J. W. Liu, 1993, Elimination structures for unsymmetric sparse LU factors: *SIAM Journal on Matrix Analysis and Applications*, **14**, 334–352, doi: [10.1137/0614024](https://doi.org/10.1137/0614024).
- Gosselin-Cliche, B., and B. Giroux, 2014, 3D frequency-domain finite-difference viscoelastic-wave modeling using weighted average 27-point operators with optimal coefficients: *Geophysics*, **79**, no. 3, T169–T188, doi: [10.1190/geo2013-0368.1](https://doi.org/10.1190/geo2013-0368.1).
- Guermouche, A., J.-Y. L'Excellent, and G. Utard, 2003, Impact of reordering on the memory of a multifrontal solver: *Parallel Computing*, **29**, 1191–1218, doi: [10.1016/S0167-8191\(03\)00099-1](https://doi.org/10.1016/S0167-8191(03)00099-1).
- Hackbusch, W., 1999, A sparse matrix arithmetic based on  $H$ -matrices. Part I: Introduction to  $H$ -matrices: *Computing*, **62**, 89–108, doi: [10.1007/s006070050015](https://doi.org/10.1007/s006070050015).
- Hicks, G. J., 2002, Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions: *Geophysics*, **67**, 156–165, doi: [10.1190/1.1451454](https://doi.org/10.1190/1.1451454).
- Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: *Geophysical Journal International*, **157**, 1269–1296, doi: [10.1111/j.1365-246X.2004.02289.x](https://doi.org/10.1111/j.1365-246X.2004.02289.x).
- Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator: *Geophysics*, **61**, 529–537, doi: [10.1190/1.1443979](https://doi.org/10.1190/1.1443979).
- Krebs, J., J. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M. D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, **74**, no. 6, WCC105–WCC116, doi: [10.1190/1.3230502](https://doi.org/10.1190/1.3230502).
- Li, Y., L. Métivier, R. Brossier, B. Han, and J. Virieux, 2015, 2D and 3D frequency-domain elastic wave modeling in complex media with a parallel iterative solver: *Geophysics*, **80**, no. 3, T101–T118, doi: [10.1190/geo2014-0480.1](https://doi.org/10.1190/geo2014-0480.1).
- Marfurt, K., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations: *Geophysics*, **49**, 533–549, doi: [10.1190/1.1441689](https://doi.org/10.1190/1.1441689).
- Métivier, L., and R. Brossier, 2015, The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: *Geophysics*, **81**, no. 2, F1–F15, doi: [10.1190/geo2015-0031.1](https://doi.org/10.1190/geo2015-0031.1).
- Min, D. J., C. Sin, B.-D. Kwon, and S. Chung, 2000, Improved frequency-domain elastic wave modeling using weighted-averaging difference operators: *Geophysics*, **65**, 884–895, doi: [10.1190/1.1444785](https://doi.org/10.1190/1.1444785).
- MUMPS team, 2015, MUMPS — Multifrontal massively parallel solver users' guide — version 5.0.0, <http://www.mumps-solver.org>, accessed 20 February 2015.
- Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD): *Geophysical Journal International*, **169**, 1069–1078, doi: [10.1111/j.1365-246X.2006.03262.x](https://doi.org/10.1111/j.1365-246X.2006.03262.x).
- Nocedal, J., and S. J. Wright, 2006, *Numerical optimization*, 2nd ed.: Springer.
- Operto, S., R. Brossier, L. Combe, L. Métivier, A. Ribodetti, and J. Virieux, 2014, Computationally efficient three-dimensional visco-acoustic finite-difference frequency-domain seismic modeling in vertical transversely isotropic media with sparse direct solver: *Geophysics*, **79**, no. 5, T257–T275, doi: [10.1190/geo2013-0478.1](https://doi.org/10.1190/geo2013-0478.1).
- Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain monoparameter full-waveform inversion of ocean-bottom cable data: Application to Valhall in the viscoacoustic vertical transverse isotropic approximation: *Geophysical Journal International*, **202**, 1362–1391, doi: [10.1093/gji/ggv226](https://doi.org/10.1093/gji/ggv226).
- Operto, S., J. Virieux, P. R. Amestoy, J.-Y. L'Excellent, L. Giraud, and H. Ben Hadj Ali, 2007, 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study: *Geophysics*, **72**, no. 5, SM195–SM211, doi: [10.1190/1.2759835](https://doi.org/10.1190/1.2759835).
- Petrov, P. V., and G. A. Newman, 2012, 3D finite-difference modeling of elastic wave propagation in the Laplace-Fourier domain: *Geophysics*, **77**, no. 4, T137–T155, doi: [10.1190/geo2011-0238.1](https://doi.org/10.1190/geo2011-0238.1).
- Petrov, P. V., and G. A. Newman, 2014, Three-dimensional inverse modeling of damped elastic wave propagation in the Fourier domain: *Geophysical Journal International*, **198**, 1599–1617, doi: [10.1093/gji/ggu222](https://doi.org/10.1093/gji/ggu222).
- Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503, doi: [10.1111/j.1365-246X.2006.02978.x](https://doi.org/10.1111/j.1365-246X.2006.02978.x).
- Plessix, R. E., 2007, A Helmholtz iterative solver for 3D seismic-imaging problems: *Geophysics*, **72**, no. 5, SM185–SM194, doi: [10.1190/1.2738849](https://doi.org/10.1190/1.2738849).
- Plessix, R. E., 2009, Three-dimensional frequency-domain full-waveform inversion with an iterative solver: *Geophysics*, **74**, no. 6, WCC53–WCC61, doi: [10.1190/1.3211198](https://doi.org/10.1190/1.3211198).
- Plessix, R.-E., and C. Perez Solano, 2015, Modified surface boundary conditions for elastic waveform inversion of low-frequency wide-angle active land seismic data: *Geophysical Journal International*, **201**, 1324–1334, doi: [10.1093/gji/ggv087](https://doi.org/10.1093/gji/ggv087).
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain. Part I: Theory and verification in a physics scale model: *Geophysics*, **64**, 888–901, doi: [10.1190/1.1444597](https://doi.org/10.1190/1.1444597).
- Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi: [10.1046/j.1365-246X.1998.00498.x](https://doi.org/10.1046/j.1365-246X.1998.00498.x).
- Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method: *Geophysical Prospecting*, **38**, 287–310, doi: [10.1111/j.1365-2478.1990.tb01846.x](https://doi.org/10.1111/j.1365-2478.1990.tb01846.x).
- Prieux, V., R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. Barkved, and J. Kommedal, 2011, On the footprint of anisotropy on isotropic full waveform inversion: The Valhall case study: *Geophysical Journal International*, **187**, 1495–1515, doi: [10.1111/j.1365-246X.2011.05209.x](https://doi.org/10.1111/j.1365-246X.2011.05209.x).
- Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013a, Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 1: Imaging compressional wavespeed, density and attenuation: *Geophysical Journal International*, **194**, 1640–1664, doi: [10.1093/gji/ggt177](https://doi.org/10.1093/gji/ggt177).
- Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013b, Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 2: Imaging compressional and shear-wave velocities: *Geophysical Journal International*, **194**, 1665–1681, doi: [10.1093/gji/ggt178](https://doi.org/10.1093/gji/ggt178).
- Riyanti, C. D., Y. A. Erlangga, R. E. Plessix, W. A. Mulder, C. Vukic, and C. Oosterlee, 2006, A new iterative solver for the time-harmonic wave equation: *Geophysics*, **71**, no. 5, E57–E63, doi: [10.1190/1.2231109](https://doi.org/10.1190/1.2231109).

- Rouet, F., X. S. Li, P. Ghysels, and A. Napov, 2015, A distributed-memory package for dense hierarchically semi-separable matrix computations using randomization: CoRR.
- Schreiber, R., 1982, A new implementation of sparse Gaussian elimination: *ACM Transactions on Mathematical Software*, **8**, 256–276, doi: [10.1145/356004.356006](https://doi.org/10.1145/356004.356006).
- Shin, C., S. Jang, and D. J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: *Geophysical Prospecting*, **49**, 592–606, doi: [10.1046/j.1365-2478.2001.00279.x](https://doi.org/10.1046/j.1365-2478.2001.00279.x).
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248, doi: [10.1190/1.1649391](https://doi.org/10.1190/1.1649391).
- Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: *First Break*, **28**, 65–70, doi: [10.3997/1365-2397.2010012](https://doi.org/10.3997/1365-2397.2010012).
- Sirgue, L., J. T. Etgen, and U. Albertin, 2008, 3D frequency domain waveform inversion using time domain finite difference methods: 70th Annual International Conference and Exhibition, EAGE, Extended Abstracts, F022.
- Soubrier, F., A. Haiddar, L. Giraud, H. Ben-Hadj-Ali, S. Operto, and J. Virieux, 2011, Three-dimensional parallel frequency-domain visco-acoustic wave modelling based on a hybrid direct/iterative solver: *Geophysical Prospecting*, **59**, 834–856, doi: [10.1111/j.1365-2478.2011.00966.x](https://doi.org/10.1111/j.1365-2478.2011.00966.x).
- Soubrier, F., S. Operto, J. Virieux, P. R. Amestoy, and J.-Y. L'Excellent, 2009, FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data — Part 1: Algorithm: *Computers & Geosciences*, **35**, 487–495, doi: [10.1016/j.cageo.2008.04.013](https://doi.org/10.1016/j.cageo.2008.04.013).
- Stekl, I., and R. G. Pratt, 1998, Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators: *Geophysics*, **63**, 1779–1794, doi: [10.1190/1.1444472](https://doi.org/10.1190/1.1444472).
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: [10.1190/1.1441754](https://doi.org/10.1190/1.1441754).
- Vigh, D., and E. W. Starr, 2008a, 3D prestack plane-wave, full waveform inversion: *Geophysics*, **73**, no. 5, VE135–VE144, doi: [10.1190/1.2952623](https://doi.org/10.1190/1.2952623).
- Vigh, D., and E. W. Starr, 2008b, Comparisons for waveform inversion, time domain or frequency domain?: 78th Annual International Meeting, SEG, Expanded Abstracts, 1890–1894.
- Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Wang, S., M. V. de Hoop, and J. Xia, 2011, On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver: *Geophysical Prospecting*, **59**, 857–873, doi: [10.1111/j.1365-2478.2011.00982.x](https://doi.org/10.1111/j.1365-2478.2011.00982.x).
- Wang, S., M. V. de Hoop, J. Xia, and X. S. Li, 2012a, Massively parallel structured multifrontal solver for time-harmonic elastic waves in 3D anisotropic media: *Geophysical Journal International*, **191**, 346–366, doi: [10.1111/j.1365-246X.2012.05634.x](https://doi.org/10.1111/j.1365-246X.2012.05634.x).
- Wang, S., J. Xia, M. V. de Hoop, and X. S. Li, 2012b, Massively parallel structured direct solver for equations describing time-harmonic qP-polarized waves in TTI media: *Geophysics*, **77**, no. 3, T69–T82, doi: [10.1190/geo2011-0163.1](https://doi.org/10.1190/geo2011-0163.1).
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80, doi: [10.1190/geo2012-0338.1](https://doi.org/10.1190/geo2012-0338.1).
- Weisbecker, C., 2013, Improving multifrontal solvers by means of algebraic block low-rank representations: Ph.D. thesis, Toulouse University.
- Xia, J., 2013, Randomized sparse direct solvers: *SIAM Journal on Matrix Analysis and Applications*, **34**, 197–227, doi: [10.1137/12087116X](https://doi.org/10.1137/12087116X).
- Xia, J., S. Chandrasekaran, M. Gu, and X. S. Li, 2009, Superfast multifrontal method for large structured linear systems of equations: *SIAM Journal on Matrix Analysis and Applications*, **31**, 1382–1411, doi: [10.1137/09074543X](https://doi.org/10.1137/09074543X).