UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

Stochastic inversion of FWI images for reducing structural uncertainty

PhD Thesis

presented and defended on January 30th, 2026 for the degree of

Doctor of the Université de Lorraine

Geosciences Specialization

by

Giusi Ruggiero

Composition du jury:

Présidente:

Alessandra Ribodetti	Université Côte d'Azur

Rapporteurs:

Florian Wellmann	RWTH Aachen (Allemagne)
Andreas Fichtner	ETH Zurich (Suisse)

Examinateurs:

Alexandrine Gesret	Université PSL
Thomas Bodin	ICM-CSIC (Barcelone, Espagne)

Directeur de thèse:

Paul Cupillard	Université de Lorraine

Co-encadrant de thèse:

Guillaume Caumon	Université de Lorraine

GeoRessources - UMR 7359
Université de Lorraine – CNRS
École Nationale Supérieure de Géologie
2 rue du Doyen Marcel Roubault - BP 10162
54505 Vandœuvre-lès-Nancy, FRANCE

Acknowledgments

I am extremely grateful to the many people who have contributed to this project through their work, support, encouragement and feedback.
Firstly, I would like to thank my PhD jury composed of Thomas Bodin, Andreas Fichtner, Alessandra Ribodetti, Alexandrine Gesret and Florian Wellmann for accepting the invitation to evaluate my work, for taking the time to read this thesis and for the valuable feedback.
I'm also grateful to my PhD advisory committee composed of Yann Capdeville, Ludovic Métivier and Pierre Thore. Thank you for the useful suggestions and discussions that took place during our annual meetings - this work has certainly benefited from them!
Above all, I am deeply grateful to my supervisors, Paul Cupillard and Guillaume Caumon, for their trust and for giving me the opportunity not only to work on this project, but also to teach and supervise students. I honestly feel fortunate to have had the chance to collaborate with two professors who never imposed a single path on me, but instead encouraged me to make my own choices and believe in my potential. Thank you for the scientific knowledge and encouragement you have provided over the past three years.
One of the most beautiful and inspiring aspects of this PhD has been the opportunity to meet and engage with several researchers and professors, including those mentioned above, whose work I had studied and admired. I was truly surprised and inspired by their humility and their genuine openness to conversation, advice and intellectual exchange. I hope that one day I will be able to follow in their footsteps.
I also acknowledge the RING-GOCAD Consortium managed by ASGA, for funding my PhD, and all the RING team members with whom I shared these three years. I would like to thank Maryse Villeneuve, Armelle Faynot and Marie Gallaire, who helped me from the outset to navigate the French bureaucracy and certainly made my move to France much easier; Pauline Collon and Christine Fay Varnier for their sympathy and the enjoyable chats we've shared over the years; Christophe Antoine and Mustapha Zakaria for always being available to help and teach me so much from a technical point of view, you're a great asset to the team!
A huge thank to all my collegues: Enrico Scarpa, Fabrice Taty Moukati, Augustin Gouy, Amandine Fratani, Marius Rapenne, Paul Marchal, Julien Herrero, Jeremie Giraud, Gloria Arienti, François Bonneau, Loïs Letellier, Bastien Morin, Ayoub Bel Hachmi, Roberto

Cilli, Zakaria Riane Chergui, Fethi Fliti and all the new entries. Thank you guys for lightening up the long working days with a great atmosphere. And thank you especially for your patience in teaching me French, how to play tarot and the other board games, and so much more.
I would also like to thank all the Italian friends whom I have met over the past three years and who have kept me company on weekends with aperitifs, dinners and music. Nancy wouldn't have been as fun without you!
Finally and above all, I would like to thank Alessandro, my family and my long-standing friends for their love, support in my choices and their constant presence in my life. I hope you know how grateful I am, and how fortunate and complete I feel to have you by my side.

Table of contents

Extended abstract **1**

Résumé étendu **11**

1 General introduction **22**
1.1 Seismic imaging and structural geologic interpretations 23
1.1.1 Limitations for high-resolution imaging 24
1.1.2 Uncertainty in geological interpretations . . . 27
1.2 FWI and smooth equivalent models 29
1.3 Downscaling FWI to high resolution structural models with inverse homogenization 31
1.4 Thesis outline 32

2 Inverse problem theory **34**
2.1 Ill-posedness of inverse problems 36
2.2 The deterministic framework to inverse problems . . 38
2.2.1 Local optimization algorithms 38
2.2.1.1 BFGS methods 41
2.3 The Bayesian framework to statistical inverse problems 43
2.3.1 Relation between the Hessian and the posterior covariance 44
2.3.2 Statistical inversion algorithms 47
2.3.2.1 Markov Chain Monte Carlo 48
2.3.2.2 Properties of Markov chains 48
2.3.2.3 Metropolis-Hastings algorithm 50

3 Homogenization and Full Waveform Inversion overview 53
3.1 Homogenization 55
3.1.1 1-D periodic homogenization 55
3.1.1.1 Summary of results 57
3.1.2 1-D non-periodic homogenization 58
3.1.2.1 Summary of results 60
3.1.3 2-D non-periodic homogenization 61

3.1.3.1 Construction of the effective elastic tensor and density 62
3.1.4 Numerical example 63
3.2 Full waveform inverse problem 68
3.2.1 Local uncertainty estimation 69
3.2.2 Homogenized full waveform inverse problem . 71
3.2.3 Numerical example 71
3.2.3.1 Forward modeling and inversion setup 72
3.2.3.2 Velocity model and acquisition geometry 74
3.2.3.3 Inversion results 75
3.2.3.4 Standard deviation maps 76
3.2.3.5 Homogenized FWI results 78
3.3 Summary and discussion 79

4 Downscaling inversion for fault-related uncertainty quantification 82
4.1 Introduction 82
4.2 Geological Parametrization 83
4.3 Bayesian inference with MCMC 84
4.4 Examples 85
4.4.1 Application with synthetic HFWI observations 87
4.4.2 Application with homogenized FWI observations 90
4.5 Discussion 94
4.5.1 Geological complexity 95
4.5.2 HFWI as observations 97
4.5.3 Computational aspects 98

5 Downscaling inversion for stratigraphic and fault-related uncertainty quantification 100
5.1 Introduction 100
5.2 Model Parametrization 101
5.3 Bayesian framework with MCMC 103
5.4 Example 104
5.4.1 Results 107
5.5 Discussion 108
5.5.1 Geological complexity 109
5.5.2 Homogenization and geological heterogeneities scales 112
5.5.3 Computational aspects 113

6 Conclusions **114**
6.1 Motivations for this thesis 114
6.2 Thesis achievements 115
6.3 Critical Assessment and Future Perspectives 115

a Single-pass randomized SVD algorithm **118**

b Turning Band Simulation **120**

c Empirical and theoretical variograms **121**

List of Figures **132**

Bibliography **133**

Extended abstract

Introduction

In several scientific and industrial applications, we are interested in knowing the structures of the Earth's subsurface. Conventionally, subsurface structural models are built from structural interpretations derived from reflection seismic images. Thanks to ever-growing datasets and computational power in the last two decades, many progresses have been made to increase the resolution of seismic images, and consequently improve the accuracy of structural interpretations. Notably, seismic image resolution has improved as a result of accurate velocity models generated via full waveform inversion (FWI). This technique has been applied extensively in recent years at regional scales (e.g., Tape et al., 2010, French et al., 2013, Colli et al., 2013) and even global scale (Bozdağ et al., 2016), revealing more detailed and reliable structural features than previous studies. However, due to the finite-frequency nature of seismic waves used to probe the Earth's interior, the resolved Earth model is more likely a smooth representation of the true one, or an 'effective' Earth (e.g., Bodin et al., 2015). Indeed, as heterogeneities at scales smaller than the shortest wavelength of seismic waves cannot be resolved, this image is a smoothed view of the real complex structures seen by long-period seismic waves. Inability or uncertainties in the detection of small-scale features (such as faults, fractures, or thin beds) can lead to oversimplified subsurface structural models and unreliable model-based forecasts.

This challenging problem is tackled in this thesis by developing an approach aimed at reducing uncertainty on small-scale geological structures from smooth tomographic models. In particular, the proposed approach employs the non-periodic elastic homogenization to compute the effective elastic properties of a subsurface model containing small-scale heterogeneities. The fundamental idea is that of

decomposing subsurface imaging into two inverse problems: the first aims to derive a smooth equivalent (effective) version of the complex heterogeneous subsurface with the homogenized FWI technique (HFWI), where the non-periodic homogenization is introduced in the FWI inversion scheme (Capdeville & Métivier, 2018); the second inverse problem, referred to as downscaling, or inverse homogenization, is addressed within a probabilistic framework to interpret the smooth HFWI solution as an ensemble of plausible small-scale models of the subsurface.
The inverse homogenization has already been tested in the layered model case (Bodin et al., 2015) and in the case of a 2-D homogeneous medium containing cavities (Hedjazian et al., 2021). In this work, this technique is applied to a 2-D elastic model of the subsurface in order to reduce fault-related uncertainty (in fault geometry and displacement) and uncertainty in small-scale elastic property variations (below seismic resolution), which serve as a proxy for uncertainty in subsurface stratigraphy. A central aspect of this work is therefore the development of a downscaling probabilistic framework that incorporates stochastic structural modeling, wherein fault displacement is produced by using a direct kinematic modeling approach and fine-scale stratigraphy is modeled with a geostatistical approach. In the Bayesian formulation of the downscaling inverse problem, prior knowledge about model parameters (i.e., structural and associated small-scale elastic parameters) is updated with information from the observed data (i.e., HFWI result) to calculate a posterior probability distribution over possible fine-scale structural models (Fig. 1).

Inverse problem theory

Chapter 2 examines the fundamentals of inverse problem theory and describes the two principal formulations: the deterministic and the probabilistic Bayesian framework to inverse problems, highlighting the connection between them and how these approaches account for the uncertainty in the solution. Furthermore, an overview of inversion algorithms employed within each formulation is provided.

Inverse problems arise naturally in many physical applications where the aim is to reconstruct a model from a set of indirect measurements. Geophysical inverse problems, including FWI, are mostly nonlinear and ill-posed. The ill-posedness arises because observations are sparse, noisy, and indirect, leading to non-unique and pos-

sibly unstable solutions. These problems can be addressed using either a Bayesian framework or a deterministic one:

- In Bayesian inference, the goal is to estimate an unknown model and quantify its uncertainty by combining prior knowledge with information from the measurement process. The inverse problem is defined by three components: the *prior*, representing all knowledge about the model before measurements; the *likelihood*, describing the probability of observing the measurements given the model; and the *posterior*, providing the statistical solution of the inverse problem as a probability density of the model parameters.
  As presented in Chapter 2, statistical inversion methods address inverse problems by characterizing the posterior probability distribution. Because exact analytical solutions are rarely attainable — particularly in nonlinear or high-dimensional settings — computational techniques and approximations are required to extract useful information from the posterior. One common strategy is to employ sampling methods, which aim to approximate or explore the full posterior distribution and thereby capture nonlinearities and multimodal behaviors.
  In this chapter we mainly focus on Markov Chain Monte Carlo (MCMC) methods, and particularly on the well known Metropolis–Hastings (MH) algorithm, which is used to solve the downscaling inverse problem presented in Chapter 4.

- The deterministic formulation of the inverse problem aims to identify the model that minimizes the misfit between observed and predicted data through an iterative procedure. In this framework, the inversion produces a single estimate of the quantities of interest based on the available data. To address the ill-posedness of the inverse problem and guarantee a unique and stable solution, an appropriate regularization term is often introduced in the misfit function.
  In Chapter 2, some iterative gradient-based optimization algorithms are introduced. Particular attention is given to the BFGS optimization scheme, as it will be employed in the numerical examples of FWI presented in this thesis.
  Many of the studies concerning uncertainty quantification in deterministic methods are based on the approximation of the inverse Hessian matrix at the solution as the covariance matrix of the posterior density within the Bayesian inference framework. The theoretical justification for this approximation, under the assumption of linearizable forward modeling and Gaus-

sian model priors, is given in Section 2.3.1. In Chapter 3 we investigate this approximation in relation to local uncertainty estimations within the context of the FWI inverse problem.

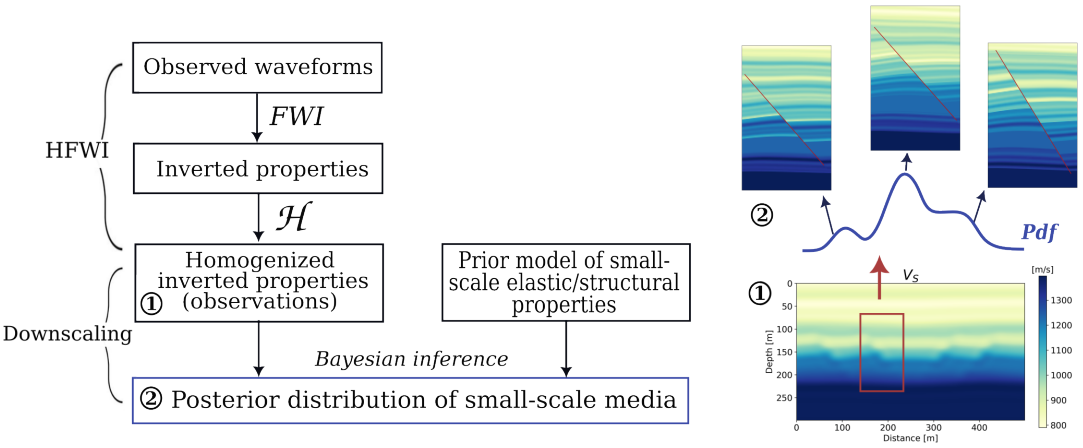

Figure 1: Overview of the methodology. Downscaling is cast into a Bayesian framework where information coming from the observed data, a smooth effective medium derived from HFWI, is combined with prior knowledge on the small-scale elastic and structural properties to retrieve the posterior distribution of all possible small-scale models that fit the data set.

Homogenization and FWI overview

Chapter 3 starts with a description of the basic theoretical foundations of the non-periodic homogenization for elastic waves, which represents the forward problem in the downscaling inversion. Next, the chapter describes important features of the FWI inverse problem in the frame of local optimization and introduces results from Capdeville & Métivier (2018) on HFWI, whose solution represents the observed data in the Bayesian framework of the downscaling.

The theory of homogenization deals with extracting homogeneous effective parameters from systems with microscopic periodic structures. In seismology, this theory has been extended to non-periodic media, relying on the fact that heterogeneous structures of scales much smaller than the minimum wavelength of seismic waves have very limited influence on waves propagating in the medium. Accordingly, the goal of homogenization is to replace the complex heterogeneous Earth by a simpler effective medium such that the resulting wavefield generated by a seismic source remains approximately (i.e.,

up to a controlled accuracy) unchanged for a given minimum wavelength.
Because FWI based on limited frequency band data can retrieve, at best (provided that data is not suffering from noise or poor coverage), the smooth homogenized version of the reference subsurface model (Capdeville & Métivier, 2018), the idea behind HFWI is that of introducing the homogenization operator in the FWI inverse problem to directly invert for smooth effective media rather than invert for sharp discontinuities.
Assuming that HFWI provides the effective properties of a true subsurface model, the inverse homogenization problem is then applied to downscale these properties. A discussion on addressing HFWI uncertainty in the Bayesian framework of the downscaling is also included in this chapter.
In a simple numerical example we applied the non-periodic homogenization to a 2-D subsurface model containing small-scale heterogeneities to show the effect of the homogenization operator, for a given user-defined scale-separation wavelength, on the elastic medium. The resulting medium is smooth and presents effective anisotropy induced by faults and velocity discontinuities within major layers (Fig. 2). With a comparison of seismic wavefield simulations in the original and homogenized medium we have assessed the poor wavefield sensitivity to the small-scale heterogeneities that have been smoothed out by homogenizing the medium.

Next, a synthetic FWI experiment is carried out to infer the elastic properties of this subsurface model and investigate its ability in recovering information about the subsurface geological structures. The results show that even under an idealized FWI setup — with noise-free data and good illumination — the exact extent and displacement of individual faults remain uncertain and, especially in areas where the geology is more complex, multiple alternative fault configurations are plausible. An approach to evaluate uncertainty in the FWI solution is then proposed. In particular, uncertainty is assessed by calculating an approximate posterior covariance matrix based on a low-rank approximation of the data-misfit inverse Hessian matrix.
Finally, the homogenization operator is applied to the elastic medium recovered by FWI to obtain the homogenized FWI solution and investigate the amount of effective anisotropy, containing the information on the small-scale structures seen in the waveform data, that has been recovered.

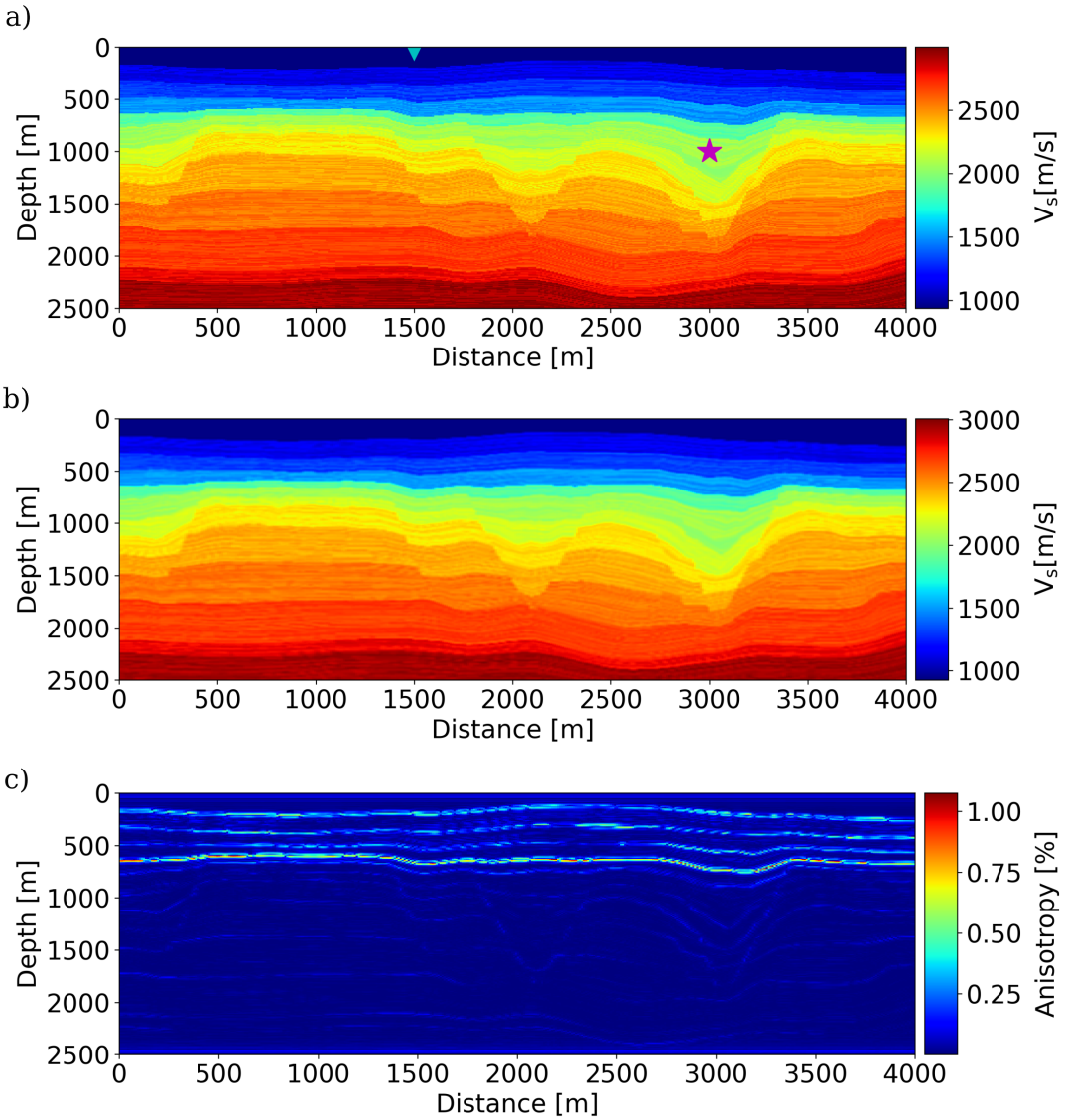

Figure 2: a) S-wave velocity of the original fine-scale elastic medium. For the seismic wave simulation, a receiver (green triangle) is positioned on the surface, while the seismic source (blue star) is located within the medium. b) Effective S-wave velocity of the homogenized medium. c) Effective anisotropy of the homogenized medium.

Downscaling inversion for fault-related uncertainty quantification

Objective and parametrization

Chapter 4 presents an application in which downscaling aims to quantify structural uncertainty. Specifically, from the smooth effective elastic properties of the subsurface, derived via HFWI and

representing the observed data, the downscaling inverse problem aims to infer fault-related model parameters describing the fault geometry and displacement. A Bayesian formalism is adopted to address the ill-posed nature of this inverse problem by incorporating prior geological knowledge.
To guarantee that model realizations are structurally compatible, fault displacement is generated using a kinematic modeling approach (Wellmann et al., 2016, Godefroy et al., 2018). The significant advantage of kinematic fault modeling is that it is easy to implement and less computationally demanding than fully dynamic or geomechanical modeling, since modeling is purely geometric and does not account for full stress-strain mechanics.
In this work, the kinematic fault model is characterized by a discontinuous displacement field defined within an elliptic shape centered on the fault. The location of the fault in the domain (defined by the coordinates of the center of the ellipse), the fault dip (given by the rotation angle of the ellipse), the spatial extent of the fault-induced displacement (provided by the ellipse's axes semi-length) and a maximum amount of slip, assumed to be at the fault center, are treated as unknowns in the downscaling inverse problem. These fault-related variables are sampled to retrieve an ensemble of possible fault configurations, consistent with the HFWI observation and prior information.

Application

The methodology is applied to a specific region in a reference fine-layered subsurface model where seismic resolution is insufficient to accurately interpret a fault structure. Downscaling is carried out within a stochastic Bayesian framework, in which an MCMC sampling scheme is employed to characterize the posterior distribution of the unknown model parameters describing the fault structure.
We conducted two distinct experiments to investigate the benefits of using as observations the full effective elastic tensor — which encodes anisotropy information — of the HFWI solution, rather than relying solely on its first component. In the first example, data are synthetically generated by homogenizing the reference elastic medium, to which noise has been introduced to make the inversion more challenging. In this case, the algorithm was highly effective in recovering the exact fault structure, owing to the informative content of the data, particularly the effective anisotropy information provided by the off-diagonal components of the full elastic tensor. In a second experiment, we tested the same algorithm with less con-

straining observed data: the first component of the effective elastic tensor obtained by homogenizing FWI-derived elastic properties. In this case, even with greater uncertainty, the mean model reflects the target fault structure (Fig. 3), demonstrating the effectiveness of the methodology. Assumptions and simplifications made in the synthetic tests, as well as the additional challenges to address in the case of real data application, are also discussed in this chapter.

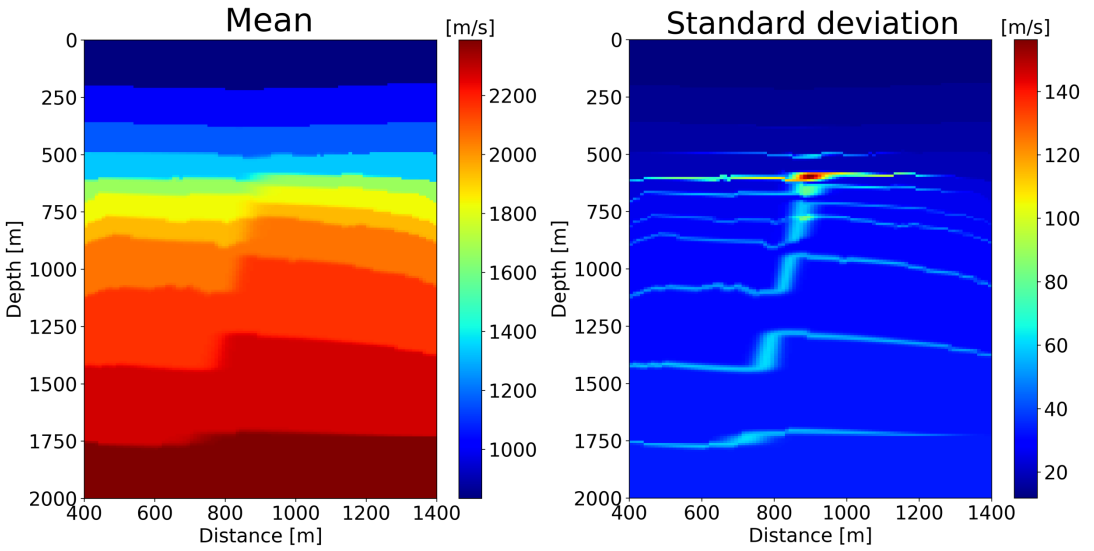

Figure 3: Posterior statistics of the downscaling inversion with homogenized FWI observations: mean and standard deviation model in which the fault structure is derived from posterior samples post-burn-in of 4 MCMC chains.

Downscaling inversion for stratigraphic and fault-related uncertainty quantification

Objective and parametrization

Chapter 5 extends the work presented in the previous chapter to an application whose objective is to resolve uncertainty on both structural (fault-related) and small-scale stratigraphic (velocity) structures. Hence, the downscaling inverse problem is cast within a Bayesian probabilistic framework to retrieve an ensemble of possible subsurface velocity models at the small-scale consistent with the smooth HFWI solution and the a *priori* knowledge. For this purpose, the methodology combines stochastic fault modeling with geostatistical simulations for generating multiple realizations of faulted

subsurface models incorporating small-scale velocity variations.
Geostatistical simulations are commonly adopted in Earth-science inverse problems to simulate spatial correlation patterns consistent with prior geological observations. Indeed, these techniques often use data to constrain realizations, thus ensuring that the generated spatial distribution of the modeled properties honors the observations.
In our downscaling inverse problem, geostatistical simulations are conditioned on a set of pilot points where velocity values are sampled and inferred. In this way, the MCMC sampling explores the parameter space by adjusting the values at the pilot points, which in turn influences the simulated velocity in the whole domain. This parametrization is chosen to reduce the dimensionality of the model parameter space, while honoring the spatial correlation structure. In this framework, the density of the pilot-point grid is used to regulate the trade-off between velocity model accuracy and computational feasibility.

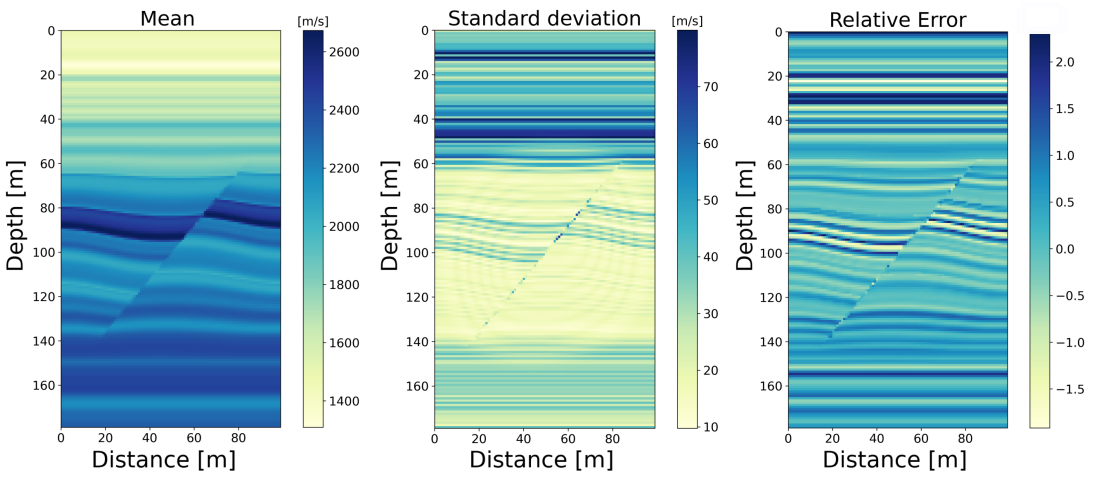

Figure 4: Posterior statistics post burn-in: mean, standard deviation and relative error of the small-scale medium $\mathbf{m}(\boldsymbol{\theta})$ obtained from posterior samples of 4 MCMC chains. The fault structure has been successfully recovered and well-constrained, whereas higher uncertainty is observed for some fine-scale velocity layers.

Application

The methodology is tested on a synthetic application, in which the HFWI medium is synthetically generated and a model involving a single fault that offsets a fine-layered stratigraphy is used to repre-

sent the reference subsurface model.
For this application, we assume that the velocity in the subsurface model can be decomposed into a macro-layered velocity field (assumed as known) and a residual velocity field that is used to reproduce thin-layering structures inside each major layer. Specifically, this residual field is simulated by generating samples from an anisotropic Gaussian random field with Turning Bands Simulation. In the downscaling inversion, both the fault-related parameters and the residual velocity field at selected pilot points are sampled via MCMC with a simple adaptive scaling M-H algorithm.
Results demonstrate the ability of the methodology to resolve uncertainty on the unknown small-scale features (fine-layering stratigraphy and fault structure), in spite of the spatially correlated noise added to the data and the large uniform prior distribution. The posterior mean, standard deviation and relative error models are displayed in Figure 4. The good performance is attributed in part to specific simplifications that were made in the synthetic application; these, together with other relevant considerations, are discusses in this chapter.

Résumé étendu

Introduction

Dans plusieurs applications scientifiques et industrielles, nous souhaitons connaître les structures du sous-sol terrestre. Traditionnellement, les modèles structuraux du sous-sol sont élaborés à partir d'interprétations structurales d'images de sismique réflexion. Grâce à l'augmentation constante des jeux de données et de la puissance de calcul au cours des deux dernières décennies, de nombreux progrès ont été réalisés pour améliorer la résolution des images sismiques et, par conséquent, améliorer la précision des interprétations structurales. La résolution des images sismiques s'est notamment améliorée grâce à des modèles de vitesse précis générés par inversion de forme d'onde complète (FWI). Cette technique a été largement appliquée ces dernières années à des échelles régionales (e.g., Tape et al., 2010, French et al., 2013, Colli et al., 2013) et même globale (Bozdağ et al., 2016), révélant des caractéristiques structurales plus détaillées et plus fiables que les études précédentes. Cependant, en raison de la nature finie des fréquences sismiques utilisées pour sonder l'intérieur de la Terre, le modèle du sous-sol obtenu est plus probablement une représentation lissée du modèle réel ; on parle alors de Terre "effective" (e.g., Bodin et al., 2015). En effet, comme les hétérogénéités à des échelles inférieures à la plus courte longueur d'onde sismique ne peuvent être résolues, chaque modèle est une vue lissée par les ondes sismiques à longue période des structures multi-échelle réelles. L'incapacité ou les incertitudes dans la détection des caractéristiques à petite échelle (telles que les failles, les fractures ou les couches fines) peuvent conduire à des modèles structuraux du sous-sol trop simples et peu prédictifs.
Ce problème difficile est abordé dans cette thèse en développant une approche visant à réduire l'incertitude sur les structures géologiques à petite échelle à partir de modèles tomographiques. En particulier,

la méthodologie proposée dans cette thèse utilise l'homogénéisation élastique non périodique pour calculer les propriétés élastiques effectives d'un modèle du sous-sol contenant des hétérogénéités à petite échelle. L'idée fondamentale derrière la méthodologie proposée est de décomposer l'imagerie du sous-sol en deux problèmes inverses : le premier est résolu par optimisation locale afin d'obtenir une version équivalente lisse du sous-sol réel à l'aide de la technique de FWI homogénéisée (HFWI), où l'homogénéisation non périodique est introduite dans le schéma d'inversion FWI (Capdeville & Métivier, 2018) ; le deuxième problème inverse, appelé downscaling ou homogénéisation inverse, est traité dans un cadre probabiliste afin d'interpréter la solution HFWI lisse comme un ensemble de modèles plausibles à petite échelle du sous-sol.
L'homogénéisation inverse a déjà été testée dans le cas d'un modèle stratifié (Bodin et al., 2015) et dans le cas d'un milieu homogène en 2-D contenant des cavités (Hedjazian et al., 2021). Dans ce travail, le problème inverse de réduction d'échelle est appliqué à un modèle élastique 2-D du sous-sol dans le but de résoudre l'incertitude liée aux failles (géométrie et déformation des couches affectées) et l'incertitude dans la stratigraphie à petite échelle (inférieures à la résolution sismique) des propriétés élastiques. Un aspect central de ce travail est donc le développement d'un cadre probabiliste de réduction d'échelle qui intègre une modélisation structurale stochastique, dans laquelle le déplacement des failles est produit à l'aide d'une approche de modélisation cinématique directe et la stratigraphie à échelle fine est modélisée à l'aide d'une approche géostatistique. Dans la formulation bayésienne du problème inverse, les connaissances a priori sur les paramètres du modèle (c'est-à-dire les paramètres structuraux et les paramètres élastiques à petite échelle associés) sont mises à jour à l'aide des informations issues des données observées (c'est-à-dire les résultats HFWI) afin de calculer une distribution de probabilité a posteriori sur les modèles structuraux à échelle fine (Fig. 5).

Théorie des problèmes inverses

Le Chapitre 2 examine les principes fondamentaux de la théorie des problèmes inverses et décrit les deux principales formulations : le cadre déterministe et le cadre probabiliste bayésien, en soulignant le lien entre eux et la manière dont ces approches tiennent compte de l'incertitude dans la solution. En outre, un aperçu des algorithmes d'inversion utilisés dans chaque formulation est fourni.

Les problèmes inverses se posent naturellement dans de nombreuses applications physiques où l'objectif est de reconstruire un modèle à partir d'un ensemble de mesures indirectes. Les problèmes inverses géophysiques, y compris la FWI, sont pour la plupart non linéaires et mal posés. Le caractère mal posé résulte du fait que les observations sont rares, bruitées et indirectes, ce qui conduit à des solutions non uniques et potentiellement instables. Ces problèmes peuvent être abordés soit via un cadre bayésien, soit via un cadre déterministe :

- l'objectif est d'estimer un modèle inconnu et de quantifier son incertitude en combinant les connaissances a priori avec les informations issues du processus de mesure. Le problème inverse est défini par trois composantes : le *prior* (a priori), représentant toutes les connaissances sur le modèle avant les mesures ; le *likelihood* (vraisemblance), décrivant la probabilité d'observer les mesures étant donné le modèle ; et le *posterior* (a posteriori), fournissant la solution statistique du problème inverse sous forme de densité de probabilité des paramètres du modèle. Comme présenté dans le Chapitre 2, les méthodes d'inversion statistique abordent les problèmes inverses en caractérisant la distribution de probabilité a posteriori. Étant donné que des solutions analytiques exactes sont rarement accessibles — en particulier dans des contextes non linéaires ou à grande dimension — des techniques computationnelles et des approximations sont nécessaires pour extraire des informations utiles de la distribution a posteriori. Une stratégie courante consiste à utiliser des méthodes d'échantillonnage, qui visent à approximer ou explorer la distribution a posteriori complète et ainsi capturer les non-linéarités et comportements multimodaux. Dans le Chapitre 2, nous nous concentrons principalement sur les méthodes de Monte Carlo par chaînes de Markov (MCMC), et plus particulièrement sur le célèbre algorithme de Metropolis–Hastings (MH), qui est utilisé pour résoudre le problème inverse de downscaling présenté dans le Chapitre 4.
- La formulation déterministe du problème inverse vise à identifier le modèle qui minimise l'écart entre les données observées et les données prédites via une procédure itérative. Dans ce cadre, l'inversion produit une seule estimation des quantités d'intérêt à partir des données disponibles. Pour traiter le caractère mal posé du problème inverse et garantir une solution unique et stable, un terme de régularisation approprié est souvent introduit dans la fonction d'écart. Dans le Chapitre 2,

certains algorithmes d'optimisation itératifs basés sur le gradient sont présentés. Une attention particulière est portée au schéma d'optimisation BFGS, qui sera utilisé dans les exemples numériques de FWI présentés dans cette thèse.
De nombreuses études concernant la quantification de l'incertitude dans les méthodes déterministes sont basées sur l'approximation de la matrice Hessienne inverse à la solution comme matrice de covariance de la densité a posteriori dans le cadre de l'inférence Bayésienne. La justification théorique de cette approximation, dans l'hypothèse d'une modélisation directe linéarisable et de modèles a priori gaussiens, est donnée dans la Section 2.3.1. Dans le chapitre 3, nous étudions cette approximation en relation avec les estimations d'incertitude locales dans le contexte du problème inverse FWI.

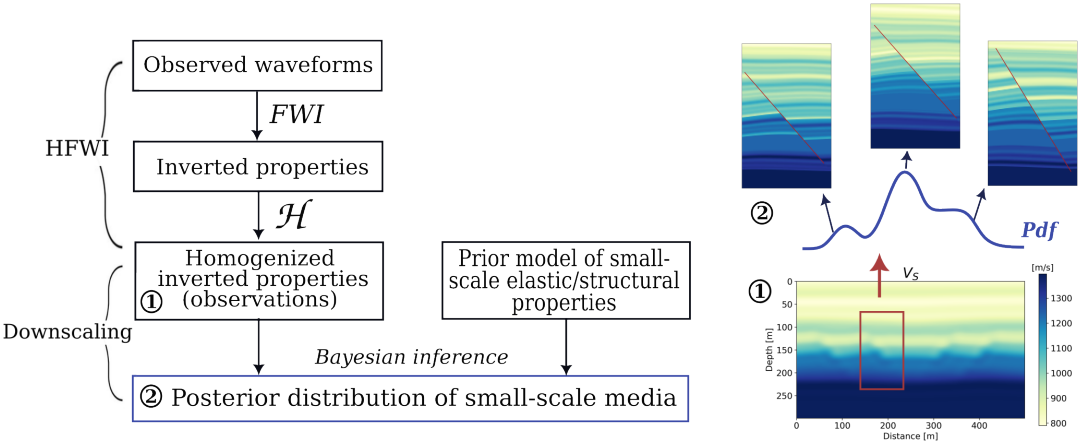

FIGURE 5 : Vue d'ensemble de la méthodologie. Le downscaling est formulé dans un cadre bayésien où les informations provenant des données observées, c'est-à-dire un milieu effectif lisse issu de la HFWI, sont combinées avec les connaissances a priori sur les propriétés élastiques et structurales à petite échelle afin d'obtenir la probabilité a posteriori de tous les modèles à petite échelle compatibles avec le jeu de données.

Aperçu de l'homogénéisation et de la FWI

Le Chapitre 3 commence par une description des fondements théoriques de base de l'homogénéisation non périodique pour les ondes élastiques, qui représente le problème direct dans l'inversion de downscaling. Ensuite, le chapitre décrit les caractéristiques importantes du problème inverse FWI dans le cadre de l'optimisation

locale et présente les résultats de Capdeville & Métivier (2018) sur la HFWI, dont la solution représente les données observées dans le cadre bayésien du downscaling.

La théorie de l'homogénéisation traite de l'extraction de paramètres effectifs homogènes à partir de systèmes possédant des structures microscopiques périodiques. En sismologie, cette théorie a été étendue aux milieux non périodiques, en s'appuyant sur le fait que les structures hétérogènes à des échelles beaucoup plus petites que la longueur d'onde minimale des ondes sismiques ont une influence très limitée sur la propagation des ondes dans le milieu. En conséquence, l'objectif de l'homogénéisation est de remplacer les propriétés hétérogènes et discontinues de la Terre par un milieu effectif plus lisse, de sorte que le champ d'ondes généré par une source sismique reste approximativement (c'est-à-dire jusqu'à une certaine précision que l'on contrôle) inchangé pour une longueur d'onde minimale donnée. Étant donné que la FWI, appliquée à des données à bande de fréquences limitée, peut au mieux (si les données ne sont pas bruitées et présentent une couverture suffisante) restituer la version homogénéisée du sous-sol (Capdeville & Métivier, 2018), l'idée à la base de la HFWI consiste à introduire l'opérateur d'homogénéisation dans le problème inverse FWI afin d'inverser directement pour des milieux effectifs lisses plutôt que pour des discontinuités nettes. En supposant que la HFWI fournisse les propriétés effectives d'un véritable modèle de sous-sol, le problème d'homogénéisation inverse est ensuite appliqué pour effectuer le downscaling de ces propriétés. Le Chapitre 3 inclut également une discussion sur la prise en compte de l'incertitude de la HFWI dans le cadre bayésien du downscaling. Dans un exemple numérique simple, nous avons appliqué l'homogénéisation non périodique à un modèle du sous-sol en 2-D contenant des hétérogénéités à petite échelle afin de montrer l'effet de l'opérateur d'homogénéisation, pour une longueur d'onde de séparation d'échelle donnée, sur le milieu élastique. Le milieu résultant est lisse et présente une anisotropie effective induite par des failles et les discontinuités de vitesse au sein des couches principales (Fig. 6). En comparant les simulations du champ d'ondes sismiques dans le milieu d'origine et dans le milieu homogénéisé, nous avons évalué la faible sensibilité du champ d'ondes aux hétérogénéités à petite échelle qui ont été lissées par l'homogénéisation du milieu.

Ensuite, une application de FWI synthétique est réalisée pour déduire les propriétés élastiques de ce modèle de sous-sol et étudier sa

capacité à récupérer des informations sur les structures géologiques du sous-sol. Les résultats montrent que même dans un cadre FWI idéalisé — avec des données sans bruit et une bonne illumination — l'étendue exacte et le déplacement des failles individuelles sont incertains et, en particulier dans les zones où la géologie est plus complexe, plusieurs configurations de failles alternatives sont possibles. Une approche visant à évaluer l'incertitude dans la solution FWI est alors proposée. En particulier, l'incertitude est évaluée en calculant une matrice de covariance a posteriori approximative basée sur une approximation de rang faible de la matrice Hessienne inverse du misfit des données.
Enfin, l'opérateur d'homogénéisation est appliqué au milieu élastique reconstruit par FWI afin d'obtenir la solution FWI homogénéisée et d'examiner la quantité d'anisotropie effective, contenant l'information sur les structures de petite échelle visibles dans les données de forme d'onde, qui a été récupérée.

Downscaling pour la quantification de l'incertitude associée aux failles

Objectif et paramétrage

Le Chapitre 4 présente une application dans laquelle le downscaling vise à quantifier l'incertitude structurale. Plus précisément, à partir des propriétés élastiques effectives lissées du sous-sol, dérivées via la HFWI et représentant les données observées, le problème inverse de downscaling a pour objectif d'inférer les paramètres du modèle liés à la faille, décrivant sa géométrie et son rejet. Un formalisme bayésien est adopté pour traiter le caractère mal posé de ce problème inverse en incorporant des connaissances géologiques a priori.
Afin de garantir la compatibilité structurale des réalisations du modèle, le déplacement lié à la faille est généré avec une approche de modélisation cinématique (Wellmann et al., 2016, Godefroy et al., 2018). L'avantage majeur de la modélisation cinématique des failles est sa simplicité de mise en œuvre et son faible coût computationnel par rapport à la modélisation dynamique ou géomécanique complète, puisque la modélisation est purement géométrique et ne tient pas compte de la mécanique complète contrainte-déformation. En particulier, le modèle cinématique de faille se caractérise par un champ de déplacement discontinu défini dans une forme elliptique centrée sur la faille. La position de la faille dans le domaine (définie par les coordonnées du centre de l'ellipse), son pendage (donné par

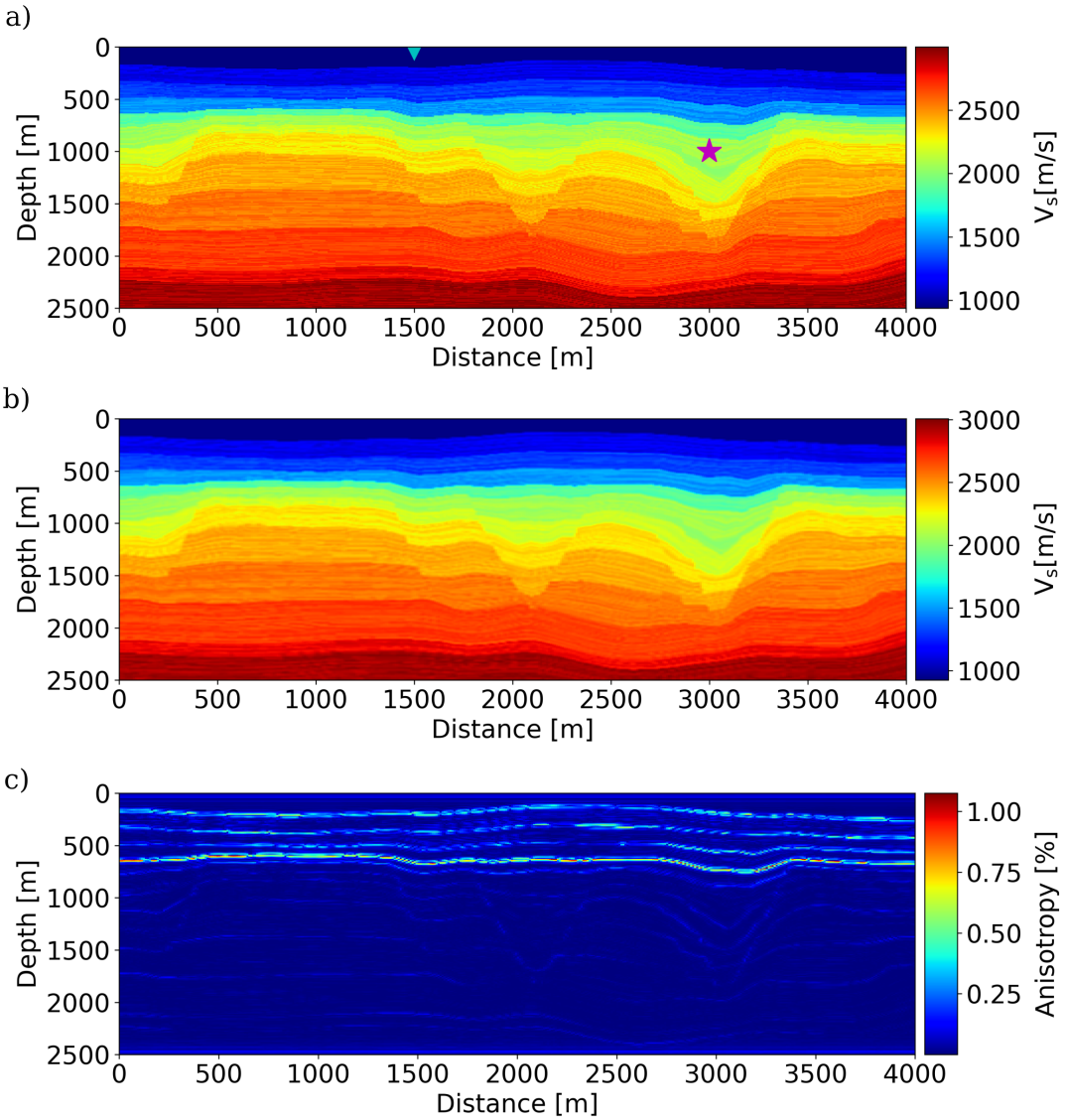

FIGURE 6 : a) Vitesse des ondes S du milieu élastique d'origine à fine échelle. Pour la simulation des ondes sismiques, un récepteur (triangle vert) est positionné à la surface, tandis que la source sismique (étoile bleue) est située à l'intérieur du milieu. b) Vitesse effective des ondes S du milieu homogénéisé. c) Anisotropie effective du milieu homogénéisé.

l'angle de rotation de l'ellipse), l'étendue spatiale du déplacement induit par la faille (définie par les demi-axes de l'ellipse) ainsi qu'un rejet maximal, supposé se produire au centre de la faille, sont traités comme des inconnues dans le problème inverse de downscaling. Ces variables liées à la faille sont échantillonnées afin d'obtenir un ensemble de configurations de faille possibles, cohérentes avec les résultats de la HFWI et les informations a priori.

Application

La méthodologie est appliquée à une région spécifique d'un modèle de sous-sol de référence à couches fines, où la résolution sismique est insuffisante pour interpréter avec précision la structure de la faille. Le downscaling est réalisé dans un cadre bayésien stochastique, dans lequel un schéma d'échantillonnage MCMC est utilisé pour caractériser la distribution a posteriori des paramètres du modèle inconnus décrivant la structure de la faille.
À travers deux tests distinctes, nous étudions les avantages d'utiliser comme observations le tenseur élastique effectif complet — qui encode l'information sur l'anisotropie — de la solution HFWI, plutôt que de se baser uniquement sur sa première composante. Dans le premier exemple, les données sont générées de manière synthétique en homogénéisant le milieu élastique de référence, auquel du bruit a été ajouté afin de rendre l'inversion plus difficile. Dans ce cas, l'algorithme s'est montré très efficace pour retrouver la structure exacte de la faille, grâce au contenu informatif des données, en particulier l'information sur l'anisotropie effective fournie par les composantes hors-diagonales du tenseur élastique complet. Dans un second test, nous avons testé le même algorithme avec des données observées moins contraignantes : la première composante du tenseur élastique effectif obtenu en homogénéisant les propriétés élastiques dérivées de la FWI. Dans ce cas, même avec une incertitude plus importante, le modèle moyen reflète la structure cible de la faille (Fig. 7), démontrant l'efficacité de la méthodologie. Les hypothèses et simplifications faites lors des tests synthétiques ainsi que les défis supplémentaires à relever dans le cas d'une application sur données réelles sont également discutés dans ce chapitre.

Downscaling pour la quantification des incertitudes stratigraphiques et associées aux failles

Objectif et paramétrage

Le Chapitre 5 étend le travail présenté dans le chapitre précédent à une application visant à réduire l'incertitude sur les structures de faille et sur les structures stratigraphiques de petite échelle. À cette fin, la méthodologie combine la modélisation stochastique des failles avec des simulations géostatistiques pour générer plusieurs réalisations de modèles de sous-sol faillés intégrant des variations de vitesse à petite échelle.

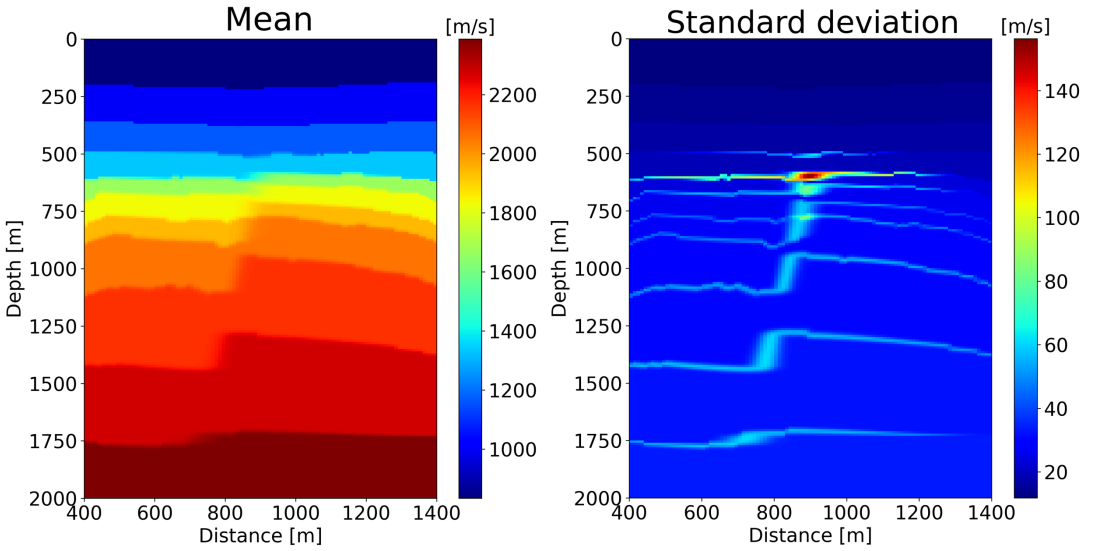

FIGURE 7 : Statistiques a posteriori de l'inversion par réduction d'échelle avec observations FWI homogénéisées : modèle moyen et écart type dans lequel la faille est dérivée des échantillons a posteriori après la phase de burn-in des 4 chaînes MCMC.

Les simulations géostatistiques sont couramment utilisées dans les problèmes inverses des sciences de la Terre pour simuler des modèles de corrélation spatiale cohérents avec les observations géologiques a priori. En effet, ces techniques utilisent souvent les données pour contraindre les réalisations, garantissant ainsi que la distribution spatiale générée des propriétés modélisées respecte les observations. Dans notre problème inverse de downscaling, les simulations géostatistiques sont conditionnées à un ensemble de points pilotes où les valeurs de vitesse sont échantillonnées et inférées. De cette manière, l'échantillonnage MCMC explore l'espace des paramètres en ajustant les valeurs aux points pilotes, ce qui influence à son tour le champ de vitesses simulé dans l'ensemble du domaine. Cette paramétrisation est choisie afin de réduire la dimensionnalité de l'espace des paramètres du modèle, tout en respectant la structure de corrélation spatiale. Dans ce cadre, la densité de la grille de points pilotes est utilisée pour réguler le compromis entre précision du modèle de vitesse et faisabilité numérique.

Application

La méthodologie est testée sur une application synthétique, dans laquelle le milieu HFWI est généré de manière synthétique et un

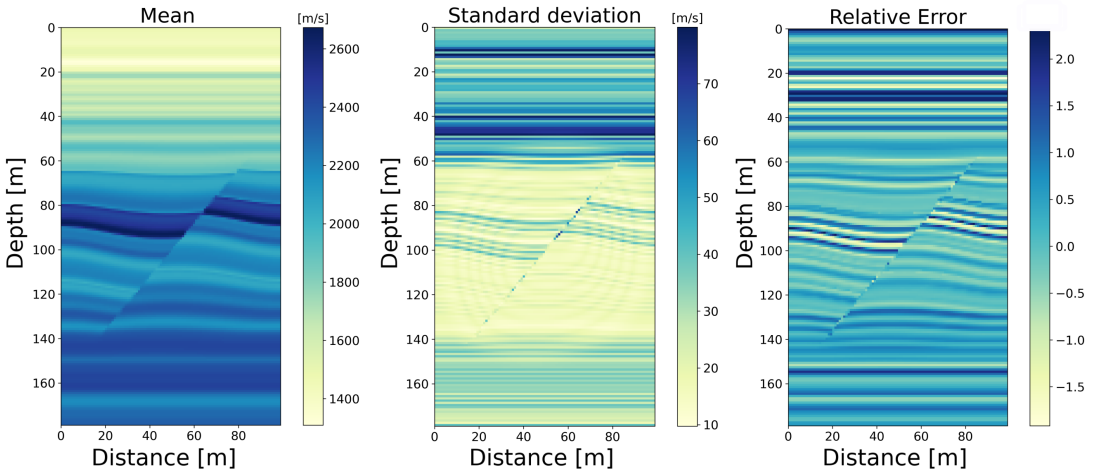

FIGURE 8 : Statistiques a posteriori après burn-in : moyenne, écart type et erreur relative du milieu à petite échelle obtenu à partir d'échantillons a posteriori de 4 chaînes MCMC. La structure de la faille a été récupérée avec succès et bien contrainte, tandis qu'une incertitude plus élevée est observée pour certaines couches de vitesse à petite échelle.

modèle impliquant une seule faille décalant une stratigraphie à fines couches est utilisé pour représenter le modèle du sous-sol de référence.
Pour cette application, nous supposons que la vitesse dans le modèle du sous-sol peut être décomposée en un champ de vitesse macro-couché (considéré comme connu) et un champ de vitesse résiduel utilisé pour reproduire les structures à fines couches à l'intérieur de chaque couche majeure. Plus précisément, ce champ résiduel est simulé en générant des échantillons à partir d'un champ aléatoire gaussien anisotrope avec une simulation géostatistique par bandes tournantes (turning bands). Dans l'inversion de mise à l'échelle descendante, à la fois les paramètres liés à la faille et le champ de vitesse résiduel aux points pilotes sélectionnés sont échantillonnés via MCMC avec un algorithme M-H adaptatif simple.
Les résultats démontrent la capacité de la méthodologie à réduire l'incertitude sur les caractéristiques inconnues à petite échelle (stratigraphie à fines couches et structure de faille), malgré le bruit spatialement corrélé ajouté aux données et la large distribution a priori uniforme. Les modèles de moyenne a posteriori, d'écart-type et d'erreur relative sont présentés dans la Figure 8. La bonne performance est attribuée en partie à des simplifications spécifiques qui ont été faites dans l'application synthétique ; celles-ci, ainsi que d'autres considérations pertinentes, sont discutées dans ce chapitre.

Chapter 1

General introduction

Since the late 20th century, seismic tomography has contributed fundamentally to the understanding of the Earth heterogeneous and anisotropic structures at various scales, from resource exploration to global-scale studies. In global seismology, for instance, tomographic models provide a fundamental framework for interpreting large tectonic structures and underlying geodynamic processes, whereas in exploration geophysics, their interpretation guides the development of exploitation strategies and the assessment of financial, societal, and ecological risks.

In exploration geophysics, seismic wavefields are commonly employed to obtain a reflectivity model of the subsurface, referred to as seismic image. Conventionally, subsurface structural models are built from attributes derived from these images. Thanks to ever-growing datasets and computational power in the last two decades, many progresses have been made to increase the resolution of seismic images, and consequently improve the accuracy of structural interpretations. Notably, seismic image resolution has improved as a result of accurate velocity models generated via full waveform inversion (FWI). This technique has been applied extensively in recent years at regional scales (e.g., Tape et al., 2010, French et al., 2013, Colli et al., 2013) and even global scale (Bozdağ et al., 2016), revealing more detailed and reliable structural features than previous studies. However, due to the finite-frequency nature of seismic waves used to probe the Earth's interior, the resolved Earth model is more likely a smooth representation of the true one, or an 'effective' Earth (e.g., Bodin et al., 2015). Indeed, as heterogeneities at scales smaller than the shortest wavelength of seismic waves cannot be resolved, this image is a smoothed view of the real complex structures seen by long-period seismic waves. Inability or uncertainties in the detection of small-scale features (such as faults, fractures, thin beds or subtle geological changes) can lead to oversimplified subsurface structural models and unreliable model-based forecasts. This challenging problem is tackled in this thesis by developing an approach aimed at reducing uncertainty on small-scale geological structures from tomographic models.

In this introductory chapter, I first provide an overview of the conventional seis-

mic imaging and geologic structural interpretation workflow and I highlight the limitations of current techniques. Then, I introduce the methodological approach and the work developed in this research, explaining the contribution that it brings to this field. Finally, I summarize the outline of the thesis.

1.1 Seismic imaging and structural geologic interpretations

For decades, the conventional seismic imaging and structural interpretation workflow has relied mainly on the following stages (e.g., Lines & Newrick, 2004):

- Seismic data acquisition: seismic data are acquired on the field by conducting a seismic survey in which seismic waves are generated by controlled active sources and propagated into the ground (Fig. 1.1). During their travel in the subsurface, waves encounter heterogeneities, where they 'refract', 'reflect', or 'convert'. The reflected energy travels back to the surface, where it is eventually measured by an array of sensors and recorded. The objective is to create from the recordings (Fig. 1.2b), referred to as seismograms, an image of these heterogeneities, which are mainly the geological features present in the subsurface.

- Velocity model building (VMB): the estimation of the velocity field of the Earth in which waves propagate is the key element to create an image of the subsurface and it is usually the most challenging step of the seismic imaging workflow. Indeed, reflection seismic data provide only an indirect measure of the subsurface properties, and an infinite-dimensional manifold of solutions exists (Backus & Gilbert, 1967). From a mathematical point of view, estimating the imaging velocity from these data is an ill-posed inverse problem and it is usually recast as a numerical optimization problem with suitable regularization methods (e.g. Tikhonov, 1963, Aster et al., 2018).
  Over the past few decades, full waveform inversion (FWI) has emerged as the transformative technology for velocity model building. From an initial guess of the velocity model (Fig. 1.2c), FWI is solved iteratively to obtain a velocity model of the subsurface in which the synthetic seismic data best fit the observed data (Fig. 1.2d).

- Seismic migration: the estimated velocity model and the recorded data are used as inputs to an algorithm called migration, which aims at producing high-resolution reflectivity images (Fig. 1.2e). As seismic migration is accomplished by extrapolating source and receiver wavefields into the subsurface using a given velocity model and estimating the reflectivity

model at each subsurface grid point by applying an imaging condition (e.g., cross-correlation, deconvolution) to the incident and reflected wavefields (Claerbout, 1971). For complex geological settings, numerical methods are generally employed, such as reverse-time migration (RTM) (Baysal et al., 1983). Because conventional migration algorithms can only handle single-scattering energy, the input data to migration must be prepared to meet such a requirement. As a consequence, a preprocessing step is required before migration in order to retain only primary reflections by attenuating the rest of the energy from the full-wavefield data through techniques such as deghosting and demultiple (Verschuur et al., 1992, Wang et al., 2013b).

- Structural interpretation: traditionally a team of experts manually interprets the geological features of the subsurface by examining the migration image and mapping structural features (horizons, anticlines, synclines, thrusts, salt bodies, fault networks). Information extracted from seismic images is combined with a prior geological knowledge to build a complete structural model, i.e. a set of consistent structural surfaces representing the geologic model, with a variety of methods (for an overview, see Caumon et al., 2009, Caumon, 2010, Jessell et al., 2014). Over the past few decades, considerable research has been devoted to the development of automatic interpretation methods as an alternative to manual interpretation by domain experts, which is often labour-intensive and subjective. Recent works on automated seismic interpretation, mainly through the use of neural networks (NNs), have shown increasing success in interpreting high-quality synthetic seismic data (e.g., Huang et al., 2017, Dramsch & Lüthje, 2018, Wu et al., 2018). However, while NNs can infer abstract features from training data, they lack the capacity to account for geological reasoning. Thus they are likely to require additional information about the structures in areas where interpretation suffers from low image quality.

1.1.1 Limitations for high-resolution imaging

Recent advances in seismic imaging workflows have led FWI to output the subsurface image or reflectivity directly (e.g., Zhang et al., 2020, Huang et al., 2021, Wei et al., 2023). The modified FWI workflow models and uses the full-wavefield data, including primaries and multiples, reflection and transmission waves, to iteratively invert for the reflectivity together with velocity (Zhang et al., 2020), thus potentially eliminating the need to go through time-consuming and error-prone preprocessing and migration steps. Results have shown that the reflectivity image directly output from FWI can indeed provide additional structural information with higher S/N compared to conventional migrated images (Huang et al.,

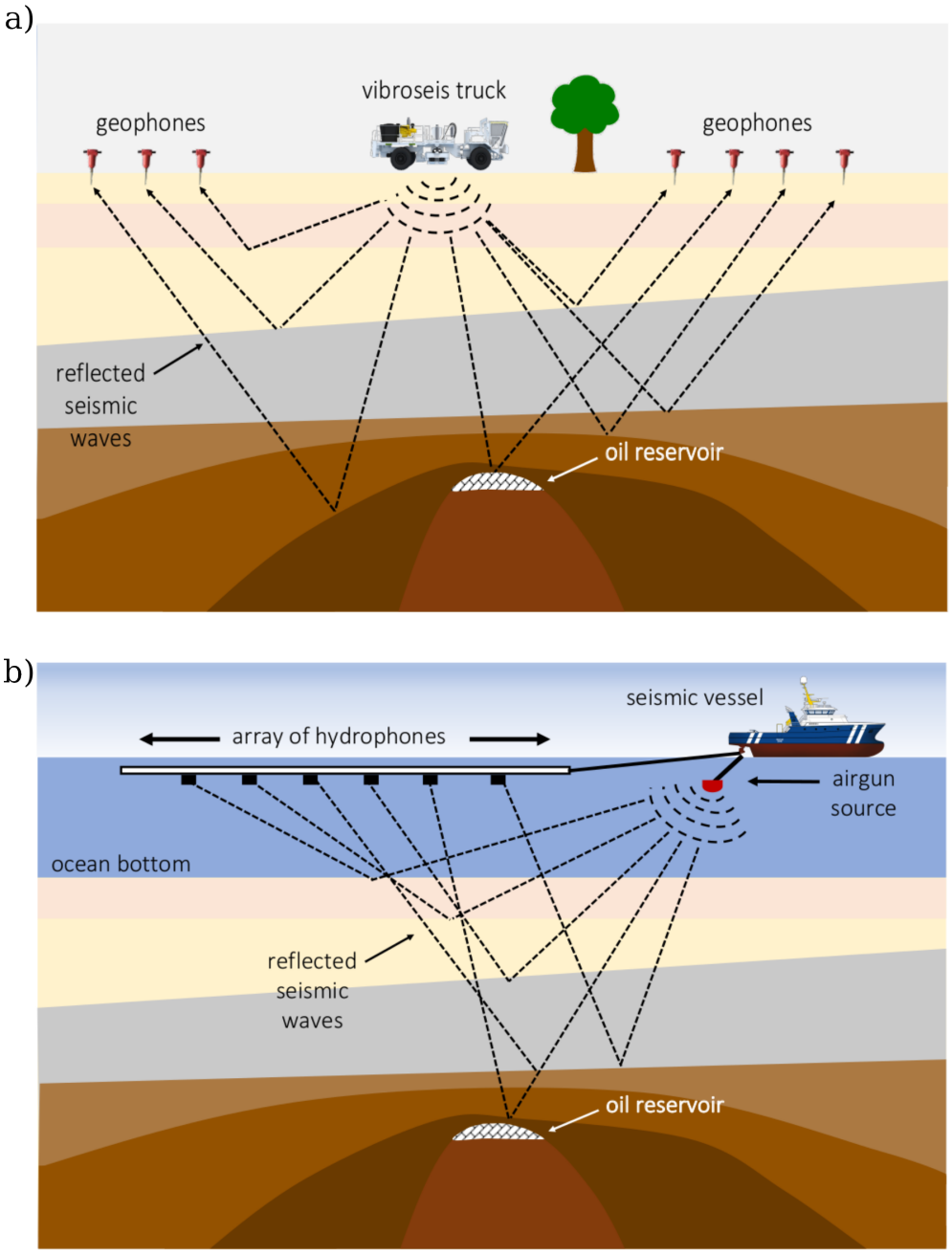

Figure 1.1: Schematic diagrams illustrating onshore (a) and offshore (b) seismic acquisition surveys. Black arrows represent seismic waves propagating through the subsurface. From Barnier (2022).

2021). Despite these advances, the ability to obtain an accurate high-resolution FWI image is strongly limited by several factors (Wei et al., 2023). In primis, the quality of the available seismic data. A good input data set should provide sufficient constraints for the velocity model update from low to high wavenumbers. This means we would need data acquired with long-enough offsets, full-azimuth

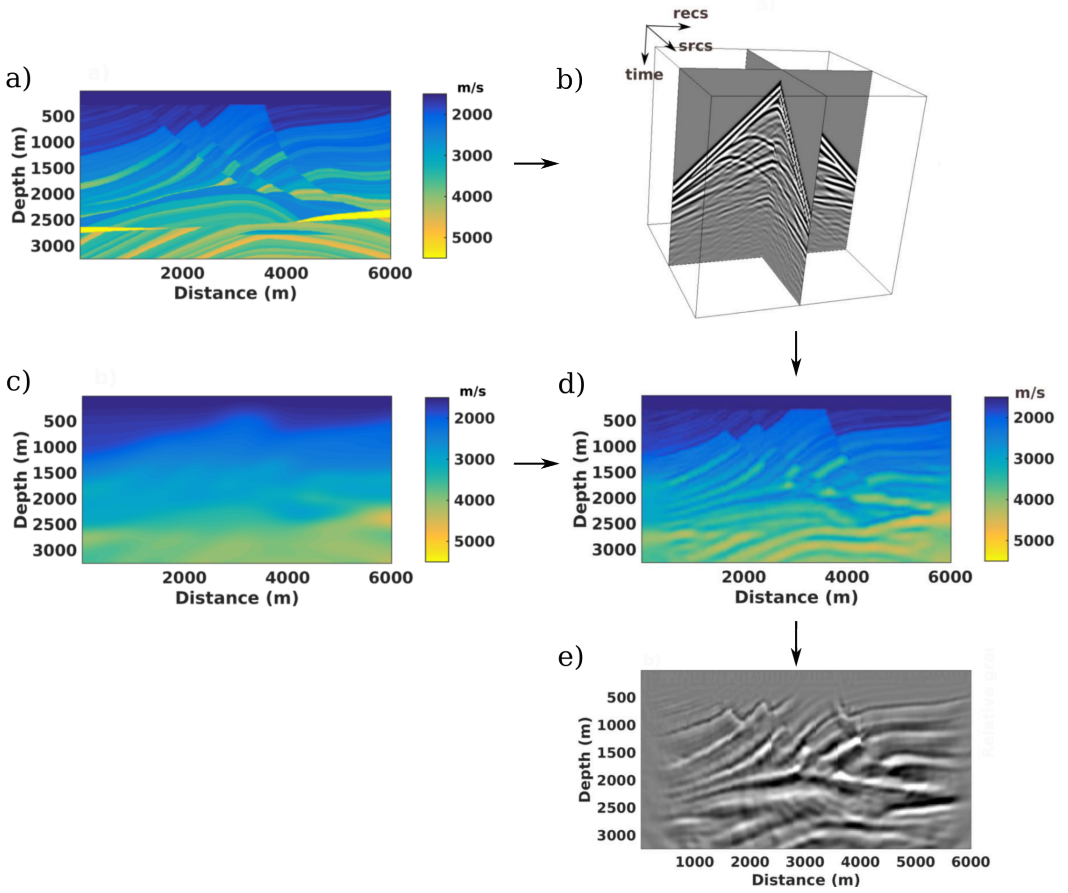

Figure 1.2: Typical seismic imaging workflow. a) Portion of the Marmousi velocity model (Versteeg, 1994) used for numerical experiments representing the reference (true) velocity model, which is unknown in practice. b) "Observed" data generated from the reference model. c) Initial velocity model obtained by smoothing the reference model. d) Velocity model estimated from the observed data and the initial velocity by waveform inversion. e) Seismic image of the reference model obtained by migrating the observed data using the velocity model in d). After Irakarama et al. (2019).

coverage, good low-frequency S/N and dense spatial sampling. In addition, the quality of the estimated solution is sensitive to the accuracy of the physical model used in FWI. Indeed, to be able to accurately describe the wave propagation in the subsurface and precisely explain the recorded data, we should account for elasticity and attenuation (viscoelastic effects), as well as anisotropy. As a consequence, the computational cost will inevitably increase. In practice, to reduce the costs, the frequency content is limited (limiting the maximum frequency used in the inversion) and simplifying assumptions on the complex physics are made. A further — and perhaps the most challenging — issue comes from the fact that FWI problem is inherently nonlinear and the inversion results depend on an initial velocity model. This means that, if the initial guess is too far from the true solution, most optimization schemes tend to converge to local minima, recovering

an inaccurate velocity model.

1.1.2 Uncertainty in geological interpretations

Given the indirect, noisy and non-unique nature of reflection seismic data processing into images, geological interpretation is inherently uncertain: multiple geologically possible structural models can be interpreted from the same seismic image (Bond et al., 2007, Bond, 2015, Alcalde et al., 2017, Irakarama et al., 2019) and structural interpretation results often depends on the expertise of the interpreter (Fig. 1.3). However, especially in complex geological settings, multiple interpretations can exhibit significant variations with important implications. For example, variations in fault displacement and fault pattern may strongly affect fluid flow behavior, and small perturbation of fault position can affect the production behavior in case some wells are located close to a fault.

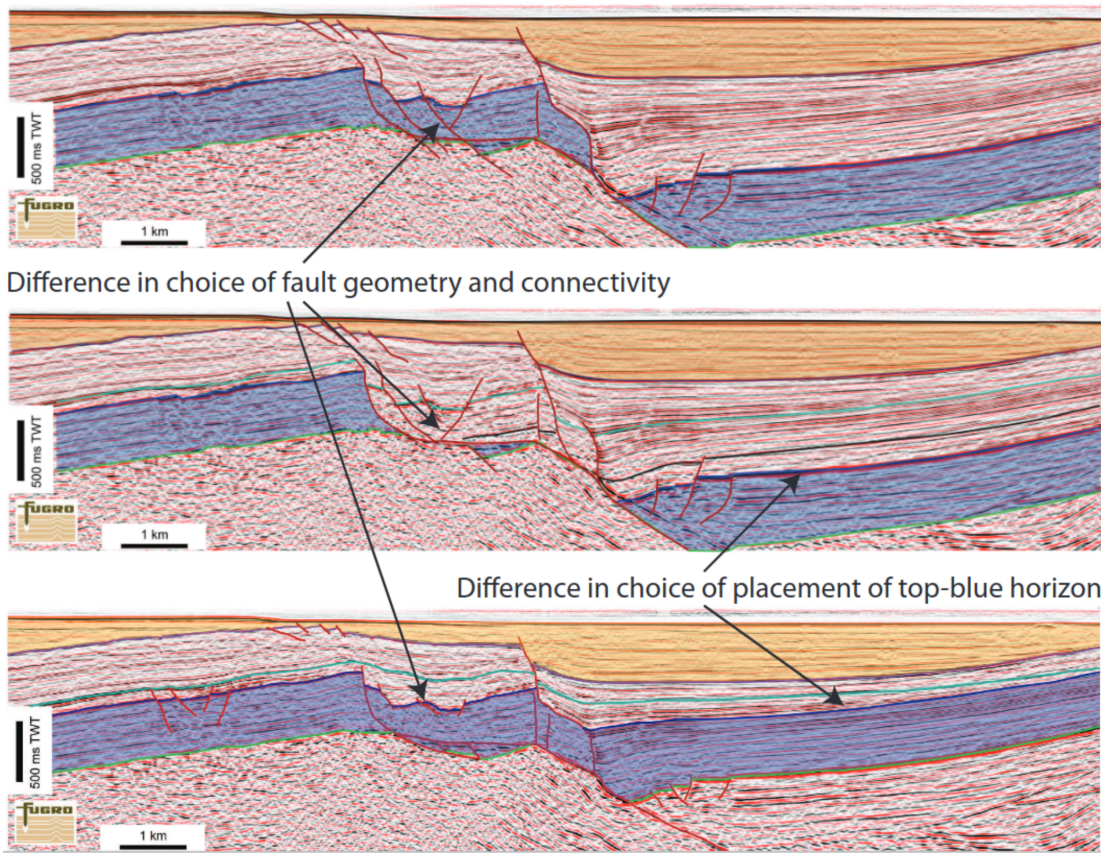

Figure 1.3: Different interpretations of the same seismic image reflecting uncertainty in geological interpretation, such as subjectivity in fault placement, or existence. Modified from Bond (2015).

This problem is well recognized and, up to now, different methodologies to capture the uncertainty in the geological structure have been developed, and some have been included as an integrated part of the geologic modeling workflow (Leahy &

Skorstad, 2013, Aarnes et al., 2014, Howley & Meyer, 2015, Bárbara et al., 2019). Generating equiprobable multiple structural models from a single data set reproducing the best-estimate interpretation is one of the earliest methods proposed to evaluate structural uncertainties in geomodeling (Lecour et al., 2001, Hollund et al., 2002, Thore et al., 2002). Several authors have used this method to sample prior structural uncertainty by generating a set of possible structural models honouring geological/geophysical constraints, subsequently reducing uncertainty through inverse approaches informed by subsurface data (Suzuki et al., 2008, Jessell et al., 2010, Wellmann et al., 2010, 2014). Bayesian approaches to stochastic structural geomodeling (Caers, 2011, Cherpeau et al., 2012, de la Varga & Wellmann, 2016, Wang et al., 2017) have similarly been adopted. These approaches provide informed prior parametrizations, which can be further constrained using case-specific geological likelihood functions and supplementary data.

Beyond these methods, complementary strategies have emerged to directly link FWI models with geological interpretations. Indeed, FWI focuses purely on matching wave physics, not necessarily on producing geologically realistic results. Several techniques have thus been developed for reducing the inverse model's geological interpretation ambiguity through the introduction of constraints during the inversion. For instance, some studies propose to parametrize the FWI problem using a level-set method to represent the interface of complex geometries, such as the boundary of salt bodies (e.g., Kadu et al., 2017, Dahlke et al., 2020, Muir & Tsai, 2020, Muir et al., 2022). With this approach, the inversion updates the shape and position of geological bodies, faults, or horizons (represented by the level set function), while the material properties inside/outside those bodies may be treated as homogeneous or separately parametrized. This ensures geologically realistic parametrizations that represent sharp interfaces or closed structures.

More recently, the use of deep learning to incorporate prior information on geological structures or to directly learn the map between data and model domains has been employed (e.g., Li et al., 2021, Zhang & Alkhalifah, 2022, Wang & Alkhalifah, 2024, Zhang et al., 2025). Deep learning methods have the ability to capture complex geological patterns. Indeed, unlike conventional methods, which often depend on linear models and physical assumptions that may not fully capture the complexity of real-world data, deep learning models can directly learn relevant features from data leading to improvements in the efficiency of seismic interpretation and inversion tasks (Fernandes et al., 2023). Nevertheless, these methods also face several challenges, including poor generalizability and the tendency to overfit by producing geological features unsupported by the data; and often they lack of uncertainty estimations, making it difficult to assess the solution confidence. Alternatively, various approaches of Bayesian probabilistic FWI have been introduced to incorporate geologic prior knowledge and to assess uncertainty on the solution (e.g., Visser et al., 2019, Kolbjørnsen et al., 2020, Zhao & Curtis, 2024, Singh et al., 2018, Mosser et al., 2020). Fully nonlinear Bayesian FWI, which avoids Gaussian assumptions on the prior and posterior distributions, can

be solved using sampling techniques such as Markov Chain Monte Carlo (MCMC) (Neal, 1993, Gilks et al., 1995, Brooks et al., 2011). For these techniques, the main limitation is the increased computational demand associated with running a large amount of forward models to evaluate the numerous samples needed to describe the solution space. Reducing computational cost often involves reducing the number of model parameters, leading to a loss of resolution.

Given the limitations and complexity of current workflows, we emphasize the need for new methodologies that deliver accurate geological information at small scales and ensure feasibility of uncertainty assessments in practical applications while maintaining reasonable computational costs. This motivates the development of the methodology proposed in this thesis.

1.2 FWI and smooth equivalent models

Because seismic inversions are performed in a limited frequency band, due to physical and computational limitations, FWI provides a smooth representation of the real complex structures (assuming that the global minimum of the objective function is achieved). The FWI-derived model is similar to a smooth model obtained from the application of the homogenization operator to a smaller-scale heterogeneous model (Bodin et al., 2015, Capdeville & Métivier, 2018, Hedjazian et al., 2021).
In the context of wave propagation, homogenization techniques have been applied to replace small-scale with smooth effective models (e.g., Capdeville et al., 2010b, Cupillard et al., 2019) and resulted in more efficient wave modeling with preserved accuracy (Fig. 1.4). Once applied, the homogenization operator smooths out the heterogeneities of scale smaller than a given minimum wavelength. The result is a smooth equivalent medium (the medium 'seen' by the wavefield), such that the limited frequency waveforms computed in the true and homogenized medium are the same up to a controlled error (Capdeville & Marigo, 2007).
In the FWI context, the homogenization has already been used by several authors (Capdeville et al., 2013, Fichtner et al., 2013b, Afanasiev et al., 2016, Capdeville & Métivier, 2018). In Fichtner et al. (2013b), the non-periodic homogenization is used to construct smooth long-wavelength equivalent models to apply a full waveform inversion method that integrates data across a wide range of spatial scales and frequencies. Advantages include ability to resolve structures at various scales simultaneously while still employing information from the entire seismogram within reasonable computational bounds. In Afanasiev et al. (2016), the homogenization is applied to constrain FWI gradient updates in the layered medium case, ensuring that fine-scale model adjustments preserve the long-wavelength characteristics of the seismic model. In Capdeville & Métivier (2018) the homogenization is applied to help interpret the results of full waveform inversion. This

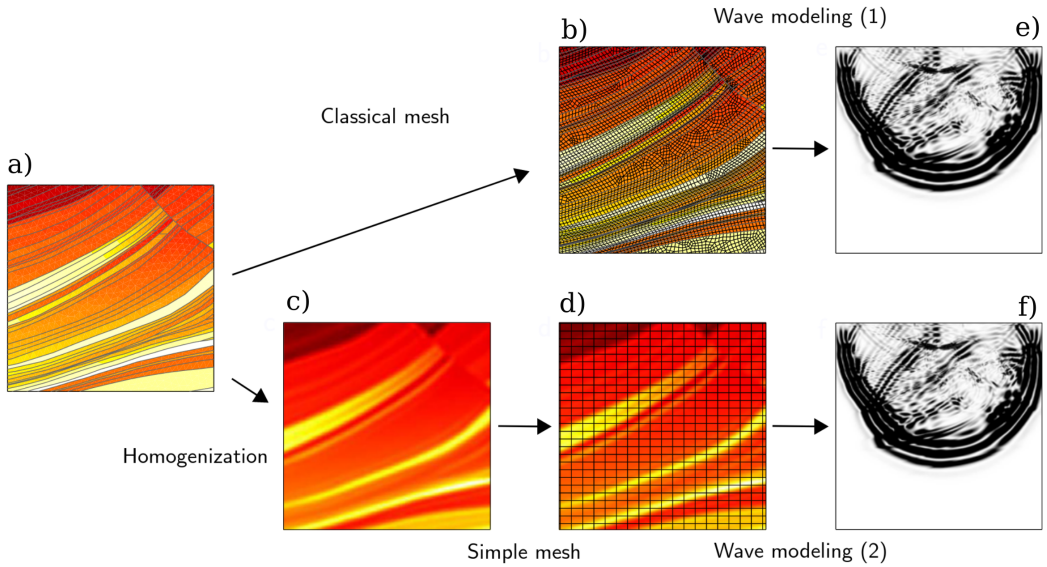

Figure 1.4: Homogenization principle illustrated on a portion of the Marmousi model. In wavefield simulations, waves are propagated up to a maximum frequency in a complex elastic model (a). The first and classical solution is to mesh all discontinuities of (a) to obtain the complex mesh (b) and then to perform the waveform modeling (1) to obtain the wavefield (e). The second option is to use homogenization to compute the effective elastic model (c) from (a) valid up to the maximum frequency. (c) is fully anisotropic but smooth (without being spatially uniform), which leads to a simple mesh (d). Finally the waveform modeling (2) is performed with the same numerical solver as the one used for waveform modeling (1) to obtain the wavefield (f), but at a lower numerical cost than (e). From Capdeville et al. (2015a).

study extends results from the stratified media case of Capdeville et al. (2013), and it numerically shows that FWI based on limited frequency band data can retrieve, at best (provided that data is not suffering from noise or poor coverage), the smooth homogenized version of the reference model. Hence, authors suggest to invert for a smooth effective medium through the homogenized FWI (HFWI), where the FWI solution is homogenized after each iteration or only after the last iteration. Results also show that incorporating homogenization constraints can improve the robustness and convergence of FWI by effectively guiding the solution space.

However, the smooth effective models produced by HFWI leave possible ambiguities for geological interpretations. For this reason, Hedjazian et al. (2021) propose to solve the seismic imaging problem as a two-stage inversion: first perform a deterministic waveform inversion where the aim is to recover a macroscale effective elastic medium, based on HFWI; then carry out a second inverse problem, the downscaling or inverse homogenization, to capture geologically realistic small-scale structures. The authors evaluated the methodology on a near-surface

problem involving a 2-D homogeneous medium with elliptic cavities. Because the downscaling problem is non-unique, the inversion is recast as a Bayesian inference problem to find the set of all possible models with small-scale cavities sharing the same long wavelength effective properties using some a *priori* information. Along the same lines, Santos et al. (2024) apply inverse homogenization to downscale tomographic images of marble cake mantle models by using a generative adversarial network (GAN)-based sampler to generate prior models.
Our work follows the ideas of Capdeville & Métivier (2018) and Hedjazian et al. (2021), and focuses on the downscaling or inverse homogenization problem using a dedicated parametrization for faulted layers.

1.3 Downscaling FWI to high resolution structural models with inverse homogenization

In this thesis we consider seismic inversion as a two-stage inversion problem and we apply inverse homogenization to locally downscale a smooth effective model obtained through HFWI, as in Hedjazian et al. (2021). The objective is to infer and quantify uncertainty on geological structures at the small-scale, i.e., at scales below the FWI resolution limit.
The inverse homogenization is applied to the two-dimensional case of a fault characterization problem in a finely stratified medium in order to properly detect fault geometric features and estimate small-scale elastic property variations. Indeed, while identifying large-scale faults is relatively simple due to their associated seismic horizon offsets resulting from larger throw, faults with throws that fall at or below vertical seismic resolution are challenging to identify, but yet critical for controlling permeability and fluid flow patterns in the sub-surface (e.g. Walsh et al., 1991, Mitchell & Faulkner, 2012).
In the Bayesian formulation of the inverse problem, prior knowledge about model parameters (i.e., structural and associated small-scale elastic parameters) is updated with information from the observed data (i.e., homogenized FWI result) to calculate a posterior probability distribution over possible fine-scale structural models (Fig. 1.5). The fundamental contribution is thus the development of a downscaling Bayesian framework that incorporates stochastic structural modeling, wherein fault displacement is produced by using a direct kinematic modeling approach (Godefroy et al., 2018) and fine-scale stratigraphy is modeled with a geostatistical approach (Goovaerts, 1997, Pyrcz & Deutsch, 2014). The methodology, validated on 2-D synthetic experiments, shows several advantages:

- uses a cheap forward model (the homogenization operator) which makes computationally the use of stochastic inversion techniques possible;

- explicitly encodes geologic information into the inverse problem;

- can be applied on subsets of large-scale FWI models, providing a target-oriented uncertainty characterization framework and consequently allowing for model dimension reduction.

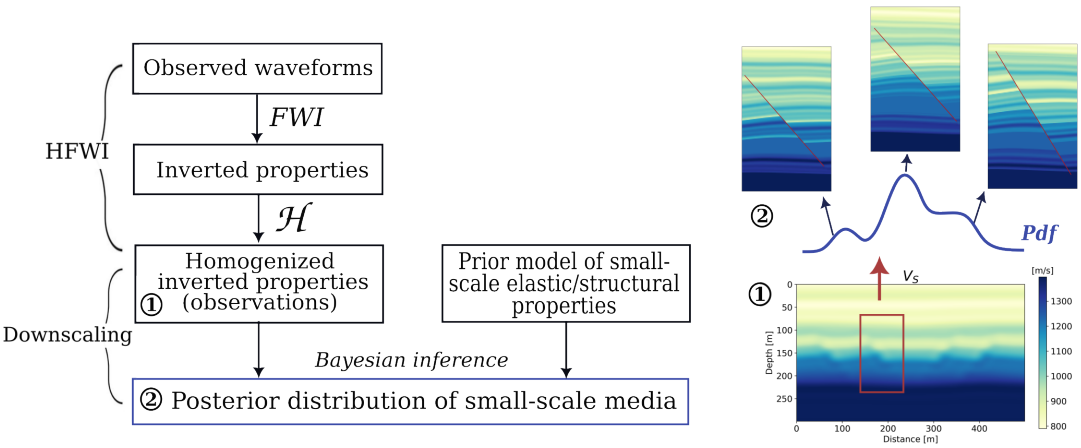

Figure 1.5: Overview of the methodology. Downscaling is cast into a Bayesian framework where information coming from the observed data, a smooth effective medium derived from HFWI, is combined with prior knowledge on the small-scale elastic and structural properties to retrieve the posterior distribution of all possible small-scale models that fit the data set.

1.4 Thesis outline

The thesis is structured as follows:

- Chapter 2 examines the fundamentals of the theory of inverse problems, establishing the conceptual and methodological basis for the subsequent application of inversion techniques in this thesis. The chapter introduces the concepts of nonlinear and ill-posed problems and develops the two principal formulations of inverse problems: the deterministic and the probabilistic Bayesian approaches, highlighting the connection between them and how these approaches account for the uncertainty in the solution. Then, an overview of inversion algorithms employed within each formulation is provided. For the deterministic approach, local optimization methods such as gradient-based algorithms are discussed. Particular attention is given to the BFGS optimization scheme, as it will be employed in the numerical examples of FWI in Chapter 3. For the Bayesian formulation, stochastic sampling techniques, particularly Markov Chain Monte Carlo (MCMC) methods, are introduced as tools for exploring posterior distributions and quantifying uncertainty. This algorithm is used to solve the downscaling inverse problem presented in Chapters 4 and 5.

- Chapter 3 starts with a description of the basic theoretical foundations of the non-periodic homogenization for elastic waves, which represents the forward problem in the downscaling inversion. A numerical example is presented as a showcase of the effect of the homogenization operator (and the choice of its inherent parameters) on a fine-scale elastic medium. Next, the theoretical framework of local optimization-based FWI and HFWI, whose solution plays the role of the observed data in the downscaling inversion, is introduced. Since the Bayesian formulation of the downscaling problem calls for estimating the data uncertainty to properly measure the mismatch between observations and predictions, a strategy to evaluate uncertainty on the FWI inverted model is investigated. Through a synthetic example of elastic FWI we investigate numerically the validity of the approach.

- Chapter 4 and Chapter 5 are devoted to the core topic of this work: the downscaling formulated as an inverse homogenization problem. Chapter 4 presents an application in which downscaling aims at quantifying structural uncertainty. Specifically, the methodology has been applied to a specific region in a reference fine-layered earth where seismic resolution was insufficient to accurately interpret the fault structure. Downscaling is carried out within a stochastic Bayesian framework with a fault-related parametrization in order to directly incorporate geologic uncertainty into the inversion process. Chapter 5 extends these results to an application whose objective is to resolve uncertainty on both structural (fault-related) and stratigraphic (velocity) small-scale structures. For this purpose, the model is parametrized as a structural-velocity model where the probabilistic framework of the downscaling inverse problem integrates to stochastic fault modeling geostatistical techniques for simulating realizations of the velocity field at small-scales. The benefits of the proposed method are illustrated on synthetic tests, therefore limitations and challenges of applying the approach to real data are discussed in detail.

- Chapter 6 provides a synthesis of the work presented in this thesis and an outlook on future applications and research directions that may be pursued to extend this work.

Chapter 2

Inverse problem theory

Most of our knowledge of the Earth's interior comes from analyzing data collected at the surface. Extracting information about the subsurface (e.g., rock density, velocity of wave propagation, electrical conductivity or the location of inhomogeneities in inaccessible areas) from a set of measurements (e.g., seismic, magnetic, gravity or borehole data) relies on an inverse problem. The fundamental framework for geophysical data inversion was laid down by Backus & Gilbert (1967, 1968, 1970) and Wiggins (1972), and first major achievements in this field came from seismic tomography, in which long-period seismic waves were used to image portions of the Earth (Aki & Lee, 1976, Aki et al., 1977, Dziewonski et al., 1977). In recent years, with modern computing power and increasing amounts of high-quality data being recorded, stored, and made available, addressing inverse problems, even for large-scale datasets, has evolved into a standard practice and has become indispensable for geophysical investigations.

Inverse problems arise naturally in many physical applications where the aim is to reconstruct a model $\mathbf{m}$ from a set of indirect measurements $\mathbf{d}$ (Fig. 2.1). The first step consists in solving the forward (or direct) problem, which involves using the model $\mathbf{m}$ to generate the observables $\mathbf{d}$ through a forward operator containing the physical laws known to govern the process (Tarantola, 2005).
The solution strategies to solve inverse problems are dependent and specific to the characteristics of the problem (Tarantola & Valette, 1982, Snieder & Trampert, 2003). Broadly speaking, inverse problems fall into two fundamental classes: linear and nonlinear. A linear inverse problem arises when the mathematical relationship between observables and unknowns is linear, or assumed to be linear. An additional distinction can be drawn between continuous and discrete inverse problems. In continuous inverse problems, the properties of the medium are defined as continuous functions over space (and possibly time). In practice, by fragmenting the medium, discrete and finite sequences of model parameters are generally inferred.
Whether we are dealing with continuous or discrete, linear or nonlinear inverse problems, there are three important questions arising in its solution (Hadamard,

1902):

1. Does it exist?
2. Is it unique?
3. Is it stable?

A problem that has either no solution in the desired class, or many (two or more) solutions, or an unstable solution (i.e., arbitrarily small errors in the measurement data may lead to indefinitely large errors in the solutions) is defined as an "ill-posed problem".
Geophysical inverse problems are mostly discrete, nonlinear and ill-posed. The ill-posedness arises because observations are sparse, noisy, and indirect, leading to non-unique and possibly unstable solutions. Indeed, since only a finite amount of data is available to reconstruct a model with infinitely many degrees of freedom, the inverse problem is inherently non-unique: multiple models may fit the data equally well. Therefore, the model obtained through inversion does not necessarily coincide with the true model that one seeks. Furthermore, real data (and often physical theories) are contaminated with errors and the estimated model is therefore affected by these errors as well.

Traditionally, nonlinear inverse problems in discrete form are solved via local optimization methods using regularization techniques (Fernández-Martínez et al., 2013). These techniques are typically aimed at producing a reasonable estimate of the quantities of interest based on the data available. However, apart from estimating a model that is consistent with the data, it is also essential to somehow quantify the error between the estimated model and the true model, referred to as the appraisal problem (Snieder, 2001). Ideally, one would like to characterize the full range of acceptable solutions. In contrast to local optimization techniques, statistical inversion yields a probability distribution from which various estimates can be derived and their reliability assessed. In this framework, the inverse problem is reformulated as a problem of Bayesian inference where all quantities are modeled as random variables. Bayesian inference methods have been famous in the field of geophysics since the pioneering works of Tarantola & Valette (1982), Mosegaard & Tarantola (1995), Sambridge & Mosegaard (2002), and have become common practice for uncertainty quantification.

In the following sections we present background on the deterministic and Bayesian approaches to solve nonlinear inverse problems, highlighting the connection between them and analyzing how uncertainty is addressed in the two cases. We also introduce some popular algorithmic approaches which are used when adopting deterministic and Bayesian inversions. These include respectively BFGS methods, widely used in FWI as efficient local optimization strategies, as will be seen in

Chapter 3; Markov chain Monte Carlo (MCMC) methods, with specific attention devoted to the Metropolis-Hastings algorithm, which is employed for the solution of the downscaling inverse problem discussed in Chapters 4 and 5.

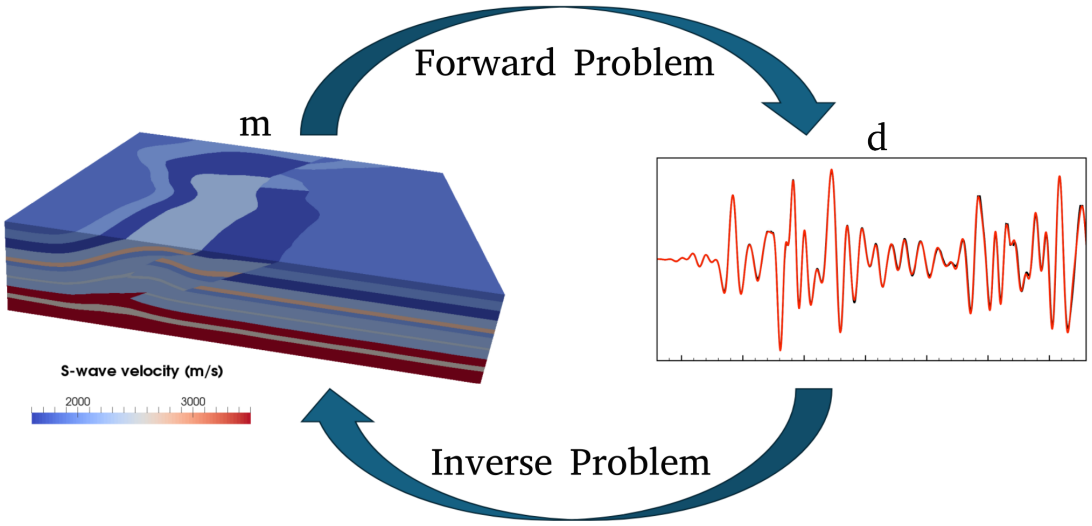

Figure 2.1: Forward and Inverse problem. The forward problem maps parameters of the material $\mathbf{m}$ into observable measures $\mathbf{d}$. The inverse problem aims to estimate unknown material parameters $\mathbf{m}$ from available observations $\mathbf{d}$. In this figure the inverse problem seeks to identify the material properties of a subsurface model based on observed seismic data. Modified from Cupillard et al. (2019).

2.1 Ill-posedness of inverse problems

We start by considering the problem of finding $\mathbf{m} \in \mathbb{R}^m$ that satisfies the equation

$$\mathbf{d} = \mathbf{G}\,\mathbf{m}, \tag{2.1}$$

where $\mathbf{d} \in \mathbb{R}^n$ is given. The physical phenomenon that relates the unknowns $\mathbf{m}$ and the measurements $\mathbf{d}$ is modeled by a matrix $\mathbf{G} \in \mathbb{R}^{n\times m}$. This type of problem may also be commonly referred to as a parameter estimation linear inverse problem. Because data in real life is contaminated by noise we can rewrite (2.1) as

$$\mathbf{d}_\eta = \mathbf{G}\,\mathbf{m} + \boldsymbol{\eta}, \tag{2.2}$$

where $\boldsymbol{\eta} \in \mathbb{R}^n$ represents the observational noise.
While direct or forward problems are in general assumed to be well-posed, inverse problems are generally ill-posed and thus we are faced with issues where we cannot guarantee the existence of the solution, its uniqueness or a stability in the solution

process. The concept of ill-posed problems was first proposed by Hadamard (1902). According to Hadamard, inverse problems are ill-posed if they break at least one of the following conditions:

1. There exists at least one solution. (Existence)
   Assume that $m < n$ and $\mathbf{G} : \mathbb{R}^m \to \mathcal{R}(\mathbf{G}) \subsetneq \mathbb{R}^n$, where the range of $\mathbf{G}$, $\mathcal{R}(\mathbf{G})$, is a proper subset of $\mathbb{R}^n$. Then, we assume that $\mathbf{G}$ has a unique inverse $\mathbf{G}^{-1} : \mathcal{R}(\mathbf{G}) \to \mathbb{R}^n$. Due to the presence of noise in the measurements $\mathbf{d}_\eta \notin \mathcal{R}(\mathbf{G})$. This means that simply inverting $\mathbf{G}$ with the data given in (2.2) is not possible.

2. There is at most one solution. (Uniqueness)
   Assume that $m > n$ and $\mathbf{G} : \mathbb{R}^m \to \mathbb{R}^n$. In this case the system is underdetermined: we have more unknowns than equations which means that there are several possible solutions.

3. The solution depends continuously on data. (Stability)
   Assume next the case $m = n$, in which there exists $\mathbf{G}^{-1} : \mathbb{R}^n \to \mathbb{R}^n$. The condition number $\kappa = \lambda_1/\lambda_n$, where λ_1 and λ_n are the biggest and smallest eigenvalues of $\mathbf{G}$, may be very large. Such a matrix is said to be ill-conditioned and is almost singular. In this case the problem is sensitive even to smallest errors in the measurement. Hence the reconstruction $\tilde{\mathbf{m}} = \mathbf{G}^{-1}\mathbf{d}_\eta = \mathbf{m} + \mathbf{G}^{-1}\boldsymbol{\eta}$ does not produce a meaningful solution but will be dominated by $\mathbf{G}^{-1}\boldsymbol{\eta}$.

Because many inverse problems fail one or more of these conditions — often the third (stability) — the key question of inverse problem theory resides in finding stable solutions.

Instability in the solution of the inverse problem is amplified when the underlying problem is nonlinear:

$$\mathbf{d}_\eta = \mathbf{f}(\mathbf{m}) + \boldsymbol{\eta}, \tag{2.3}$$

where $\mathbf{f}(\cdot)$ refers to a nonlinear forward operator or function mapping the model space to the data space.
In many practical problems, nonlinear inversion is treated as an optimization problem. In this case, a stable solution is found by minimizing the misfit between observed and predicted data in an appropriate norm, while also minimizing a regularization term that penalizes unwanted features of the parameters. Another way of tackling ill-posed nonlinear inverse problems is Bayesian inversion, which provides a posterior probability density of the model parameters of interest, thus encoding the degree of confidence in their estimate. These two different (but complementary) approaches are covered in the remainder of this chapter.

2.2 The deterministic framework to inverse problems

The deterministic inverse problem seeks to find $\tilde{\mathbf{m}}$ that minimizes the misfit between observed and predicted data (or misfit function) through an iterative approach. In the most common case, $\tilde{\mathbf{m}}$ is found by minimizing the least-squares misfit function

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - \mathbf{f}(\mathbf{m})\|_2^2. \tag{2.4}$$

Many other misfit measures exist, most notably the L^1 measure of misfit, which is known to be robust in the presence of outliers (Claerbout & Muir, 1973, Jeong et al., 2013).

To address the ill-posedness of the inverse problem and guarantee a unique and stable solution an appropriate regularization term is often introduced (Engl & Ramlau, 2015, Hofmann, 2013). The main role of the regularization is to select, among all admissible solutions, those that depend continuously on the observed data while also satisfying additional desired properties imposed by the chosen regularization. As discussed in Section 2.3.1, the regularization term in the deterministic framework can be interpreted as encoding prior knowledge, akin to the prior distribution in the Bayesian framework. One particular choice is the quadratic regularization term that penalizes the distance to a reference model $\mathbf{m}_{ref}$, the so-called zero order Tikhonov regularization. This extends (2.4) to

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d} - \mathbf{f}(\mathbf{m})\|_2^2 + \frac{\lambda}{2}\|\mathbf{m} - \mathbf{m}_{ref}\|_2^2, \tag{2.5}$$

where λ is a regularization parameter controlling the trade-off parameter between data fit and model stabilization. Another choice of regularization is called total variation (TV) regularization, popular in imaging and geophysics (e.g. Rudin et al., 1992, Gholami, 2015), where one chooses the regularization term as $\frac{\lambda}{2}\|\nabla\mathbf{m}\|_1$ to penalize the magnitude of the gradient of the model, thus encouraging solutions with sharp boundaries.

In the deterministic framework, the nonlinear problem is solved by successive linearizations using iterative optimization algorithms, some of which are introduced in the following section. We refer to Fletcher (2000) and Nocedal & Wright (2006) for a detailed description of optimization algorithms.

2.2.1 Local optimization algorithms

Local optimization methods begin with an initial guess of the model $\mathbf{m}_0$ and generate a sequence of improved estimates $\mathbf{m}_k$ using

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k, \tag{2.6}$$

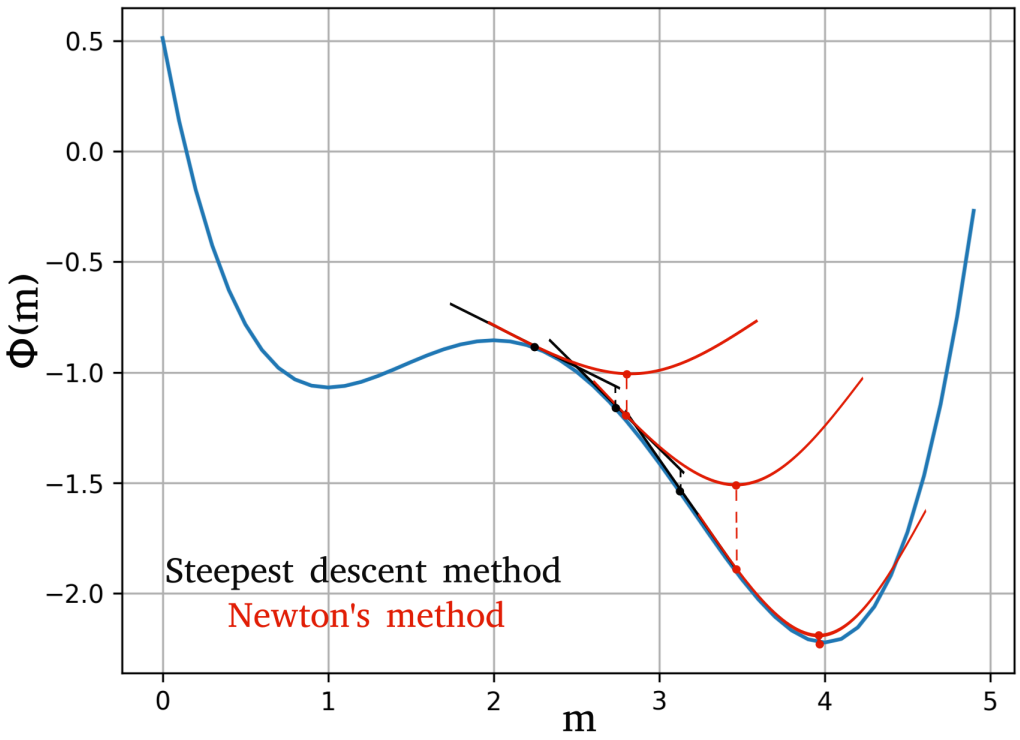

Figure 2.2: A gradient descent step (black) and a Newton step (red) on the same misfit function (blue). The gradient step moves the point downwards along the linear approximation of the function. The Newton step moves the point to the minimum of the parabola, which is used to locally approximate the misfit function, leading to a faster convergence rate.

until it terminates, hopefully at the "best-fit" solution $\tilde{\mathbf{m}}$ (Fig. 2.3).
Algorithms differ in the strategy adopted to compute a search direction $\Delta\mathbf{m}_k$ and an appropriate distance, namely the step length α_k, to move from one estimate $\mathbf{m}_k$ to the next $\mathbf{m}_{k+1}$. According to Nocedal & Wright (2006), regardless of the specifics, good algorithms should possess the following properties:

- Robustness. They should perform well on a wide variety of problems in their class, for all reasonable values of the starting point.
- Efficiency. They should not require excessive computer time or storage.
- Accuracy. They should be able to identify a solution with precision, without being overly sensitive to errors in the data or to the arithmetic rounding errors that occur when the algorithm is implemented on a computer.

However, these three properties are generally in conflict. As such, trade-offs between convergence rate and storage requirements, or between robustness and speed, have to be accounted in the choice of the algorithm.
Most of the algorithms adopt a line search strategy for deciding how to move

from one estimate to the next. Line search starts by choosing $\Delta\mathbf{m}_k$ and then identifying the step length α_k, to move from the current estimate $\mathbf{m}_k$ to a new estimate $\mathbf{m}_{k+1}$ with a lower misfit function value. The step length α can be found by approximately solving the following one dimensional minimization problem

$$\min_{\alpha>0} \phi(\mathbf{m}_k + \alpha\Delta\mathbf{m}_k). \tag{2.7}$$

In practice, the line search algorithm generates a limited number of trial step lengths until it finds one that loosely approximates the minimum of (2.7). In the following, we will focus on the choice of the search direction $\Delta\mathbf{m}_k$ in line search methods.
For instance, Newton methods derive $\Delta\mathbf{m}_k$ from a second-order Taylor series approximation to $\phi(\mathbf{m}_k + \Delta\mathbf{m}_k)$:

$$\phi(\mathbf{m}_k + \Delta\mathbf{m}_k) \approx \phi(\mathbf{m}_k) + \mathbf{g}_k^T(\Delta\mathbf{m}_k) + \frac{1}{2}\Delta\mathbf{m}_k^T\mathbf{H}_k\Delta\mathbf{m}_k, \tag{2.8}$$

where the gradient vector is $\mathbf{g}_k = \nabla\phi(\mathbf{m}_k)$ and the Hessian matrix is $\mathbf{H}_k = \nabla^2\phi(\mathbf{m}_k)$. Taking the derivative of this approximation with respect to $\Delta\mathbf{m}$ yields

$$\nabla\phi(\mathbf{m}_k + \Delta\mathbf{m}_k) \approx \mathbf{g}_k + \mathbf{H}_k\Delta\mathbf{m}_k, \tag{2.9}$$

and knowing that the minimum in the vicinity of the point $\mathbf{m}_k$ is reached when the first derivative (i.e., the left-hand side of equation 2.9) is zero, the expression of the perturbation model is given by

$$\Delta\mathbf{m}_k \approx -\mathbf{H}_k^{-1}\mathbf{g}_k, \tag{2.10}$$

where the sign minus in the equation indicates that the perturbation model is searched in the opposite direction of the gradient $\mathbf{g}_k$ of the misfit function ϕ at point $\mathbf{m}_k$, i.e., along the steepest descent direction. The main drawback of these methods is the explicit computation of the matrix of second derivatives which can sometimes be an error prone and expensive process (Nocedal & Wright, 2006). The Hessian can be thought of as providing information on the local curvature of the complex topography of the m-dimensional misfit function surface. Minimizing (2.4) is thus equivalent to approximating the misfit function $\phi(\mathbf{m})$ with a local quadratic surface and moving to its minimum (Fig. 2.2). Newton direction is reliable as long as the difference between the Taylor expansion (2.8) and $\phi(\mathbf{m}_k + \Delta\mathbf{m})$ is not too large, that is, as long as the misfit function can be locally approximated by a quadratic function; and it can be used in a line search method, typically with a unit step $\alpha = 1$, when $\mathbf{H}_k$ is positive definite. In such case, $\mathbf{g}_k^T\Delta\mathbf{m}_k < 0$ (unless the gradient is zero), so $\Delta\mathbf{m}_k$ is a descent direction.
A common strategy is that of modifying the search direction to make it satisfy the descent condition while retaining the benefit of the second-order information

contained in $\mathbf{H}_k$. For instance, the Gauss–Newton ignores part of the Hessian so that the matrix is symmetric positive definite by construction. Other approaches, named quasi-Newton algorithms, start with an initial approximation to the Hessian and make this approximation more accurate at each iteration by using new gradient information. These methods avoid the difficulties related to the computation of the Hessian and yet still attain a superlinear rate of convergence (Nocedal & Wright, 2006). A very popular approximation is the one provided by Broyden, Fletcher, Goldfarb and Shanno (Broyden, 1970, Fletcher, 1970, Goldfarb, 1970, Shanno, 1970) known as the BFGS approximation, which is examined in more detail in the following section. Besides the Gauss or quasi-Newton methods, other algorithms, simpler to implement, can be employed for the numerical optimization such as the nonlinear conjugate gradient method (Hestenes et al., 1952, Fletcher & Reeves, 1964, Shewchuk et al., 1994). According to this method, each search direction is a linear combination of the steepest descent direction and the previous direction $\Delta\mathbf{m}_{k-1}$,

$$\Delta\mathbf{m}_k = -\mathbf{g}_k + \beta_k \Delta\mathbf{m}_{k-1}, \tag{2.11}$$

where β_k is a scalar ensuring that $\Delta\mathbf{m}_k$ and $\Delta\mathbf{m}_{k-1}$ are conjugate. There are many variations of the conjugate gradient method that provide different expressions for β. A widely known variant is the one proposed by Polak & Ribiere (1969) where

$$\beta_k = \frac{\mathbf{g}_k^T(\mathbf{g}_k - \mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T\mathbf{g}_{k-1}}. \tag{2.12}$$

The simplest approximation is the one provided by the steepest descent method (Cauchy et al., 1847, Raydan, 1997) in which the Hessian matrix is simply replaced by the identity matrix. In this way, the chosen search direction is equivalent to the steepest-descent direction (Fig. 2.2)

$$\Delta\mathbf{m}_k = -\mathbf{g}_k. \tag{2.13}$$

Despite the approach's popularity due to its ease of implementation, its main drawback is the slow convergence rate (first order convergence only) which worsens as the ill-conditioning of the inverse problem increases.

2.2.1.1 BFGS methods

In quasi-Newton methods, the new estimate $\mathbf{m}_{k+1}$ is searched along the direction

$$\Delta\mathbf{m}_k = -\mathbf{A}_k^{-1}\mathbf{g}_k. \tag{2.14}$$

The only difference with the Newton method is that an approximate Hessian $\mathbf{A}_k$ is used. Quasi-Newton methods typically start with an initial positive-definite matrix approximation $\mathbf{A}_0$, which is updated after each step to account for the

Algorithm 1 L-BFGS two-loop recursion

$\mathbf{q} \leftarrow \mathbf{g}_k$;
for $i = k-1, k-2, ..., k-m$ **do**
  $\alpha_i \leftarrow \rho_i \mathbf{s}_i^T \mathbf{q}$;
  $\mathbf{q} \leftarrow \mathbf{q} - \alpha_i \mathbf{y}_i$;
end for
$\mathbf{r} \leftarrow \mathbf{B}_k^0 \mathbf{q}$;
for $i = k-m, k-m+1, ..., k-1$ **do**
  $\beta \leftarrow \rho_i \mathbf{y}_i^T \mathbf{r}$;
  $\mathbf{r} \leftarrow \mathbf{r} + \mathbf{s}_i(\alpha_i - \beta)$
end for
stop with result $\mathbf{r} = \mathbf{B}_k \mathbf{g}_k$.

additional knowledge gained during the step. These updates use the fact that changes in the gradient $\mathbf{g}$ provide information about the second derivative of $\phi(\mathbf{m})$ along the search direction. In particular, the new approximation $\mathbf{A}_{k+1}$ has to satisfy two conditions: the secant and the curvature conditions (Nocedal & Wright, 2006).
The first condition can be written as

$$\mathbf{A}_{k+1} \alpha_k \Delta \mathbf{m}_k = \mathbf{g}_{k+1} - \mathbf{g}_k, \tag{2.15}$$

and by defining the vectors

$$\mathbf{s}_k = \mathbf{m}_{k+1} - \mathbf{m}_k = \alpha_k \Delta \mathbf{m}_k, \quad \mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k, \tag{2.16}$$

equation (2.15), known as the secant equation, becomes

$$\mathbf{A}_{k+1} \mathbf{s}_k = \mathbf{y}_k. \tag{2.17}$$

The curvature condition is satisfied if $\mathbf{s}_k^T \mathbf{y}_k > 0$. While for strong convex functions this condition is always true, for non-convex functions it is not always satisfied. However, it is guaranteed to hold if we impose further requirements — the Wolfe or strong Wolfe conditions — on the line search. See Nocedal & Wright (2006) for demonstration.
The BFGS algorithm updates at each iteration the inverse Hessian approximation $\mathbf{A}_k^{-1} = \mathbf{B}_k$ instead of $\mathbf{B}_k$ itself, resulting in a more practical, and generally preferred, implementation since it allows the search direction to be calculated by means of a simple matrix–vector multiplication

$$\Delta \mathbf{m}_k = -\mathbf{B}_k \mathbf{g}_k. \tag{2.18}$$

The updated approximation $\mathbf{B}_{k+1}$ is required to be symmetric and positive definite, and must satisfy the secant equation (2.17) as well. The BFGS updating

formula to derive $\mathbf{B}_{k+1}$ is given by

$$\mathbf{B}_{k+1} = (\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^T)\mathbf{B}_k(\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T) + \rho_k \mathbf{s}_k \mathbf{s}_k^T, \tag{2.19}$$

with

$$\rho_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}, \tag{2.20}$$

and $\mathbf{I}$ denoting the identity matrix. It can be shown that the update (2.19) generates positive definite approximations whenever the initial approximation $\mathbf{B}_0$ is positive definite and $\mathbf{s}_k^T \mathbf{y}_k > 0$.
In practice, a variant of the BFGS strategy, named limited-memory BFGS (L-BFGS), is used to solve large problems. The fundamental idea is that, instead of computing and storing the approximate inverse Hessian at every iteration, it directly computes the product $\mathbf{B}_k \mathbf{g}_k$ by applying successive rank 2 updates of the initial approximation $\mathbf{B}_k^0$. In particular, the algorithm stores a limited number m, also called the memory value, of vector pairs $\{(\mathbf{s}_i, \mathbf{y}_i)\}_{i=k-m,\ldots,k-1}$ such that when the step number k exceeds m, the oldest vector pair in the set of $\{(\mathbf{s}_i, \mathbf{y}_i)\}$ is replaced by the new pair $(\mathbf{s}_k, \mathbf{y}_k)$ obtained from the current step. The rank 2 updates are then computed in a two-loop recursion scheme consisting in a sequence of inner products and vector summations involving $\nabla\phi_k$ and the vector pairs $\{(\mathbf{s}_i, \mathbf{y}_i)\}$. The scheme is shown in Algorithm 1 from Nocedal & Wright (2006).

2.3 The Bayesian framework to statistical inverse problems

From the perspective of statistical inversion, the inverse problem is recast within a Bayesian inference scheme. In this framework, the objective is to infer and assess the uncertainty about unknown model parameters by incorporating all the available knowledge of the measurement process as well as information of the unknowns that are available prior to the measurement. The inverse problem is thus formulated from three ingredients: the prior $\pi_{prior}(\mathbf{m})$, which includes all the knowledge coming from prior information one has about the model parameters $\mathbf{m}$ before any measurement is taken; the likelihood $\pi_{like}(\mathbf{d}|\mathbf{m})$, which describes the conditional probability that $\mathbf{m}$ will result in the measurement $\mathbf{d}$; and the posterior $\pi_{post}(\mathbf{m}|\mathbf{d})$:

$$\pi_{post}(\mathbf{m}|\mathbf{d}) \propto \pi_{like}(\mathbf{d}|\mathbf{m})\pi_{prior}(\mathbf{m}), \tag{2.21}$$

which provides the statistical solution of the inverse problem as a probability density of the model parameters $\mathbf{m}$. In the specific case of linear inverse problem (2.1) with additive noise modeled as Gaussian with zero mean and a covariance matrix $\mathbf{C}_d \in \mathbb{R}^{n\times n}$, and Gaussian prior density with $\mathbf{m}_{prior}$ as the mean and

$\mathbf{C}_m \in \mathbb{R}^{m\times m}$ as the covariance matrix; then the posterior probability density of the model parameters (2.21) is given explicitly (within a normalizing constant) by

$$\pi_{post}(\mathbf{m}|\mathbf{d}) \propto \exp\Big(-\frac{1}{2}\|\mathbf{G}\mathbf{m}-\mathbf{d}\|^2_{\mathbf{C}_d} - \frac{1}{2}\|\mathbf{m}-\mathbf{m}_{prior}\|^2_{\mathbf{C}_m}\Big), \tag{2.22}$$

where the notation $\|\cdot\|^2_{\mathbf{C}}$ denotes the squared norm weighted by the data $\mathbf{C}_d$ and prior $\mathbf{C}_m$ covariance matrices. If the inverse problem is nonlinear, despite the choice of Gaussian prior and noise probability distributions, the posterior probability distribution of the model parameters is generally not Gaussian, due to the nonlinearity of the forward operator $\mathbf{f}(\mathbf{m})$. The more nonlinear $\mathbf{f}(\mathbf{m})$ is, the further is $\pi_{post}(\mathbf{m}|\mathbf{d})$ from a Gaussian function. In this case, in order to find the posterior distribution in terms of statistical measures, such as mean and covariance, one can approximate the posterior by linearizing $\mathbf{f}(\mathbf{m})$ at the maximum a posteriori point ($\mathbf{m}_{MAP}$)

$$\mathbf{f}(\mathbf{m}) \approx \mathbf{f}(\mathbf{m}_{MAP}) + \mathbf{J}(\mathbf{m}-\mathbf{m}_{MAP}), \tag{2.23}$$

where $\mathbf{m}_{MAP}$ is a vector containing the parameters that maximize the posterior

$$\mathbf{m}_{MAP} := \arg\max_{\mathbf{m}} \pi_{post}(\mathbf{m}|\mathbf{d}), \tag{2.24}$$

and $\mathbf{J} \in \mathbb{R}^{m\times n}$ is the Jacobian matrix of $\mathbf{f}(\mathbf{m})$ evaluated at $\mathbf{m}_{MAP}$. In doing so, we get the parameters of the Gaussian distribution $\pi_{post}(\mathbf{m}|\mathbf{d})$ that locally approximate $\mathcal{N}(\mathbf{m}_{MAP}, \mathbf{C}_M)$ with the posterior covariance matrix $\mathbf{C}_M \in \mathbb{R}^{n\times n}$ given by

$$\mathbf{C}_M \simeq (\mathbf{J}^T\mathbf{C}_d^{-1}\mathbf{J} + \mathbf{C}_m^{-1})^{-1}. \tag{2.25}$$

In summary, under the assumptions of additive Gaussian noise, Gaussian prior, and linearized forward model, the solution of the Bayesian inverse problem is reduced to the characterization of a Gaussian posterior distribution. An alternative to entirely avoid linearizing approximations is to use sampling techniques, which aim to approximate or sample the full posterior, thus capturing nonlinearities and multimodal behavior. Some of these approaches are covered in Section 2.3.2.

2.3.1 Relation between the Hessian and the posterior covariance

The connection between the misfit function from deterministic optimization and the posterior probability density in the statistical setting is made explicit when we seek the maximum a posteriori estimate $\mathbf{m}_{MAP}$. Approximating the mean of the posterior distribution by finding the MAP point is equivalent to solving a regularized deterministic inverse problem, where $\mathbf{C}_m$ plays the role of the regularization operator, and $\mathbf{C}_d$ is a weighting for the data misfit term (Martin et al.,

2012a, Bui-Thanh et al., 2012):

$$\mathbf{m}_{MAP} \approx \arg\min_{\mathbf{m}} \phi(\mathbf{m}), \tag{2.26}$$

with

$$\phi(\mathbf{m}) = \frac{1}{2}\|\mathbf{f}(\mathbf{m}) - \mathbf{d}\|^2_{\mathbf{C}_d} + \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{prior}\|^2_{\mathbf{C}_m}. \tag{2.27}$$

This approximation is exact when the map from parameters to observables is linear (2.1), and we can write the posterior density as

$$\pi_{post}(\mathbf{m}|\mathbf{d}) \propto \exp[-\phi(\mathbf{m})]. \tag{2.28}$$

In this case, the posterior covariance matrix $\mathbf{C}_M$ is equal to the inverse of the Hessian $\mathbf{H}^{-1}$ of $\phi(\mathbf{m})$ at $\mathbf{m}_{MAP}$.
Following Martin et al. (2012a), we illustrate this in the case of a linear inverse problem. Accordingly, we can expand the deterministic cost function (2.27), or negative log posterior, as

$$\phi(\mathbf{m}) = \frac{1}{2}(\mathbf{G}\mathbf{m} - \mathbf{d})^T\mathbf{C}_d^{-1}(\mathbf{G}\mathbf{m} - \mathbf{d}) + (\mathbf{m} - \mathbf{m}_{prior})^T\mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_{prior}). \tag{2.29}$$

Then, the gradient $\mathbf{g}$ and the Hessian $\mathbf{H}$ are given by

$$\mathbf{g} = \nabla\phi(\mathbf{m}) = \mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{G}\mathbf{m} - \mathbf{d}) + \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_{prior}), \tag{2.30}$$

and

$$\mathbf{H} = \nabla^2\phi(\mathbf{m}) = \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}. \tag{2.31}$$

Hence, the negative log posterior can be rewritten as

$$\phi(\mathbf{m}) = \frac{1}{2}(\mathbf{m} - \tilde{\mathbf{m}})^T\mathbf{H}(\mathbf{m} - \tilde{\mathbf{m}}) + const. \tag{2.32}$$

By requiring the gradient $\mathbf{g}$ to be zero at the minimizer $\tilde{\mathbf{m}}$, we obtain that

$$\tilde{\mathbf{m}} = \mathbf{H}^{-1}(\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{d} + \mathbf{C}_m^{-1}\mathbf{m}_{prior}). \tag{2.33}$$

This implies that the posterior probability density is a Gaussian $\mathcal{N}(\tilde{\mathbf{m}}, \mathbf{H}^{-1})$ and the maximum a posteriori point $\mathbf{m}_{MAP} = \tilde{\mathbf{m}}$.
However, as we know, in the general case (non-Gaussian prior, nonlinear problem) the posterior is not necessarily Gaussian, so the inverse of the Hessian is only an approximation of the covariance matrix $\mathbf{C}_M$ of the posterior.
This matrix in Bayesian inversion provides a measure of the uncertainty about the optimal estimate $\tilde{\mathbf{m}}$ given the prior and data uncertainty. As such, many of the studies concerning uncertainty quantification in deterministic methods are based on the approximation of the inverse Hessian matrix at $\tilde{\mathbf{m}}$ as the covariance matrix of the posterior density within the Bayesian inference framework (e.g.,

Martin et al., 2012a). However, the linearization assumption means that this method will only be effective if the misfit function exhibits a single minimum and its surrounding architecture is approximately quadratic, while its real nonlinear topography often consists of one or more valleys of low misfits (Fig. 2.3), eventually connected by saddle points (Fernández-Martínez et al., 2012). Thus, at best we can sample uncertainty information in the vicinity of the solution given by local optimization. The application of regularization may help to conform the misfit function to this shape. As such, any estimate of model uncertainty depends not only on a good knowledge of data noise, but also on the imposed regularization being consistent with prior information (Rawlinson et al., 2014). A further drawback of these measures of uncertainty is that the inversion of potential large matrices is required. Many studies address this problem by using low-rank approximations or matrix probing techniques where information about specific properties of a matrix can be estimated through the application of the matrix to random vectors (Halko et al., 2011). In Chapter 3 we investigate these aspects in relation to local uncertainty estimations within the context of the FWI inverse problem.

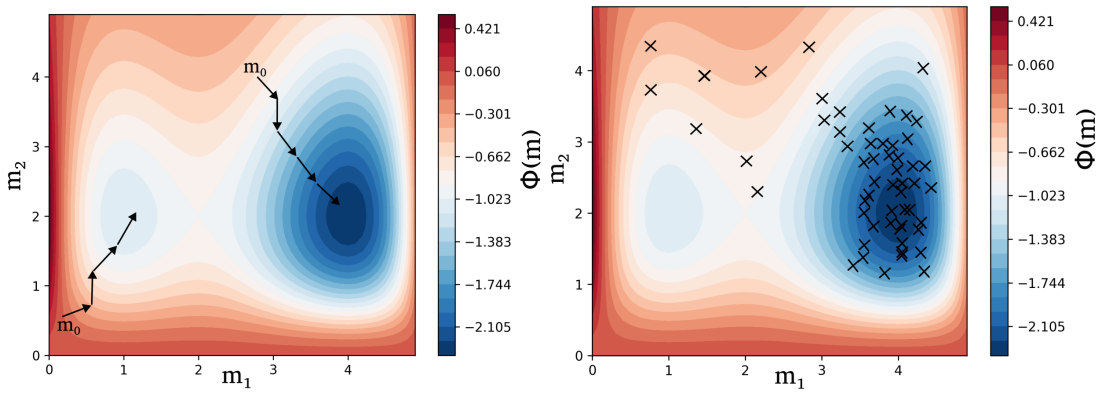

Figure 2.3: Schematic comparison between Steepest descent (local optimization) and Monte Carlo sampling for a simple two-parameter model space. Steepest descent (right plot) traces a single trajectory from an initial solution $\mathbf{m}_0$ toward the nearest minimum. Thus, depending on the initial model, it may converge to a nearby local minimum rather than the global one. Monte Carlo approach (left plot) explores the model space by randomly generating candidate models according to a given proposal distribution. This stochastic exploration allows the algorithm to probe multiple regions of the model space and avoid being trapped in local minima.

2.3.2 Statistical inversion algorithms

Statistical inversion approaches address inverse problems by characterizing the posterior probability distribution. Because exact solutions are rarely available (especially in nonlinear and high-dimensional cases), various computational tools and approximations are required to extract information from the posterior distribution. The choice of the most suitable method depends on the requirements of the model. In some cases, it may be desirable to obtain a single model outcome that can be assumed to best match prior parameter distributions and the likelihood functions. In such a case, estimating the maximum a posteriori (MAP) may be sufficient.

Commonly, rather than assessing the probability density at specific points, we are concerned with delineating the range of admissible models, thus quantifying parameter uncertainty. One approach to this problem is to approximate the posterior probability distribution by employing stochastic sampling strategies referred to as Monte Carlo methods (Fig. 2.3). These methods rely on two fundamental principles: (1) generating random samples from a target distribution and (2) accepting or rejecting those samples based on a weight determined by the scaled candidate density. Different sampling algorithms arise from variations in how these two principles are implemented, such as Markov chain Monte Carlo (MCMC) methods (Metropolis-Hastings sampling (Metropolis et al., 1953), Gibbs sampler (Geman & Geman, 1984), Hamiltonian MCMC (Duane et al., 1987) and reversible jump MCMC (Green, 1995)), or genetic algorithms (Mitchell, 1998), to name a few. However, whilst these methods can be very effective when tuned carefully to the particular problem at hand, they are undeniably costly and prohibitive for inverse problems with expensive forward problems and high dimensional parameter spaces.

Several alternatives to stochastic sampling-based methods exist. For instance, ensemble-based methods, particularly the ensemble Kalman filter (EnKF) are becoming increasingly popular (Burgers et al., 1998, Evensen & Van Leeuwen, 2000, Evensen, 2009). This approach, initially introduced as an extension to the linear Kalman filter (Kalman, 1960) for state estimation in time-varying dynamical systems, is widely used for model parameter estimation. The EnKF, formulated on the basis of Gaussian assumptions, uses an ensemble of realizations, propagated through the forward model and updated with observations, to represent the uncertainty in the parameters. The ensemble is iteratively updated using the Kalman gain, a weighting matrix calculated from the mean and covariance of prior pdf and cross-covariance between parameters and observations. This ensemble approach allows to circumvent the explicit computation of a typically huge posterior covariance matrix by replacing it with an empirical covariance estimated directly from the ensemble members. However, EnKF may require a large number of realizations to correctly map the solution space of the unknown model parameters, ultimately resulting in high computational cost (Patel & Jain,

2019). In addition, the Gaussian assumption at the base of the methodology is hard to justify in general (Stuart, 2010) and can lead to inaccuracies in highly nonlinear and non-Gaussian systems (Zhou et al., 2014). In this case, other approaches, such as hybrid particle-filter-EnKFs (Hoteit et al., 2012, Stordal et al., 2011), or Gaussian mixture (Dovera & Della Rossa, 2011), can be applied (see, e.g., Bocquet et al., 2010, for a review).
When the posterior is intractable or expensive to directly sample from, a fast deterministic alternative is offered by Variational Inference (Scherzer et al., 2009, Stuart, 2010). These methods turn Bayesian inference into an optimization problem and seek to approximate the posterior distribution by minimizing a functional over a suitable class of distributions or functions.
In the following section we will focus on MCMC methods, and particularly on the well known Metropolis–Hastings (MH) algorithm.

2.3.2.1 Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) methods aim to sample from an unknown target distribution $\pi(\mathbf{m})$ by "moving" through the parameter space. In this stochastic process, at each iteration k, a sample $\mathbf{m}_{k+1}$ is drawn from a transition probability $P(\mathbf{m}_{k+1}, \mathbf{m}_k)$ where $\mathbf{m}_k$ is the current sample. The transition probability (or kernel) depends on a proposal density $q(\mathbf{m}_{k+1} \mid \mathbf{m}_k)$ and possibly an acceptance rule, which determines whether the proposal $\mathbf{m}_{k+1}$ is accepted. For an order 1 Markov chain, the probability of moving to $\mathbf{m}_{k+1}$ depends only on the sample at the previous iteration $\mathbf{m}_k$ (and not on any of the earlier iterations). In this sense it has the shortest possible 'memory', and this is an advantage from the point of view of simplicity and required memory for the algorithm (Mosegaard & Sambridge, 2002).
There are different ways to implement MCMC, but all MCMC methods share the goal of constructing a Markov chain whose transition kernel has the desired target distribution as its (unique) stationary distribution. The main differences across MCMC algorithms lie in how the transition kernel is designed, i.e. how they propose new samples and explore the target distribution. We address a particular class of MCMC methods known as Metropolis–Hastings (MH) methods. Before going into specifics of this method, we first discuss some important properties of Markov chains. Note that this is a brief, and therefore, incomplete introduction about Markov chains, and we refer to Kaipio & Somersalo (2005), on which the following section is based, for a thorough introduction.

2.3.2.2 Properties of Markov chains

As the samples in MCMC form a Markov chain, we must ensure that the simulated sequence is a Markov Chain with a unique stationary distribution and that this distribution converges to the target distribution we want to sample. To explain

when a Markov chain fulfills these conditions, we start by introducing some basic notions from probability theory and discussing fundamental concepts concerning the transition kernels. Let π denote a probability measure over $\mathbb{R}^m$ and let's define a discrete time stochastic process $\{\mathbf{M}_k\}_{k=1}^{\infty}$ as an ordered set of random variables $\mathbf{M}_k \in \mathbb{R}^m$.
Given the transition kernel P, a time-homogeneous Markov chain is a stochastic process $\{\mathbf{M}_k\}_{k=1}^{\infty}$ with the following property

$$\begin{aligned}\pi(\mathbf{M}_{k+1} = \mathbf{m}_{k+1} \mid \mathbf{m}_1, \ldots, \mathbf{m}_k) &= \pi(\mathbf{M}_{k+1} = \mathbf{m}_{k+1} \mid \mathbf{m}_k) \\ &= P(\mathbf{m}_k, B_{k+1}),\end{aligned} \tag{2.34}$$

where B is a measurable subset of the state space (for example any Borel set $\mathfrak{B}$ in $\mathbb{R}^m$) and the transition kernel $P(\mathbf{m}_k, B_{k+1})$ gives the probability that, given the current state $\mathbf{m}_k$, the next state will lie inside the set B. The first equality implies that the probability for $\mathbf{M}_{k+1} \in B_{i+1}$ conditioned on $\mathbf{M}_1 = \mathbf{m}_1, \ldots, \mathbf{M}_k = \mathbf{m}_k$ equals the probability conditioned on $\mathbf{M}_k = \mathbf{m}_k$ alone. This is referred to as the Markov property, which in other words says that only the current state is used to determine the next one. The second equality means that the rule for how the chain moves from one state to the next is constant over time (i.e., the same at all steps).
To guarantee that samples generated from the Markov chain are distributed according to the target distribution in the long run, it is necessary that the following properties are satisfied:

- The invariance property, which guarantees that if the Markov chain is initially distributed according to π, then the distribution of the chain remains π after a single transition.
  The probability measure π is an invariant measure of $P(\mathbf{m}_k, B_{k+1})$ if $\pi P = \pi$ (i.e., if the distribution of the random variable $\mathbf{M}_k$ before the time step $k \to k+1$ is the same as the variable $\mathbf{M}_{k+1}$ after the step).

- The irreducibility property, which ensures that, regardless of the starting point, the Markov chain generated by the transition kernel P visits with a positive probability any set of positive measure.
  The transition kernel P is irreducible (with respect to π) if, for each $\mathbf{m} \in \mathbb{R}^m$ and $B \in \mathfrak{B}$, with $\pi(B) > 0$, there exists an integer n such that $P^{(n)}(\mathbf{m}, B) > 0$.

- The aperiodicity, which avoids to generate a Markov chain that remains in a periodic loop forever.

Proposition 1. *Let π be a probability measure in $\mathbb{R}^m$ and $\mathbf{M}_k$ a time homogeneous Markov chain with a transition kernel P. Assume further that π is an invariant measure of the transition kernel P, and that P is irreducible and ape-*

riodic. Then for all $\mathbf{m} \in \mathbb{R}^m$,

$$\lim_{N\to\infty} P^{(N)}(\mathbf{m}, B) = \pi(B) \text{ for all } B \in \mathfrak{B}. \tag{2.35}$$

The property (2.35) is the important ergodicity property, which ensures that the Markov chain converges to a unique stationary distribution regardless of the initial state.

2.3.2.3 Metropolis-Hastings algorithm

In the case of the Metropolis–Hastings (MH) algorithm, the aim is to construct a kernel P that proposes trial moves satisfying the detailed balance equation, which is given by

$$\pi(\mathbf{m}_{k+1})P(\mathbf{m}_{k+1}, \mathbf{m}_k) = \pi(\mathbf{m}_k)P(\mathbf{m}_k, \mathbf{m}_{k+1}), \tag{2.36}$$

for all pairs $(\mathbf{m}_k,\ \mathbf{m}_{k+1}) \in \mathbb{R}^m \times \mathbb{R}^m$. Detailed balance is a sufficient condition for invariance, thus it ensures the distribution $\pi(\mathbf{m})$ to be an invariant measure of P.
Let $q : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}_+$ be a given function with the property $\int q(\mathbf{m}_k \mid \mathbf{m}_{k+1})d\mathbf{m}_{k+1} = 1$. We refer to the kernel q as proposal distribution. If q satisfies the detailed balance equation, then we can simply set $P(\mathbf{m}_k, \mathbf{m}_{k+1}) = q(\mathbf{m}_k \mid \mathbf{m}_{k+1})$. Otherwise, the kernel is corrected by a multiplicative factor α to be determined:

$$P(\mathbf{m}_k, \mathbf{m}_{k+1}) = \alpha(\mathbf{m}_k, \mathbf{m}_{k+1})q(\mathbf{m}_k \mid \mathbf{m}_{k+1}). \tag{2.37}$$

This kernel satisfies (2.36) if

$$\alpha(\mathbf{m}_k, \mathbf{m}_{k+1}) = \min\left\{1, \frac{\pi(\mathbf{m}_{k+1})q(\mathbf{m}_{k+1} \mid \mathbf{m}_k)}{\pi(\mathbf{m}_k)q(\mathbf{m}_k \mid \mathbf{m}_{k+1})}\right\}, \tag{2.38}$$

where $\pi(\cdot)$ is the posterior (2.21).
This transition kernel is called the Metropolis-Hastings kernel (or criterion). In practice, once the proposed state $\mathbf{m}_{k+1}$ is generated, the MH algorithm uses (2.38) for choosing to either accept or reject it. It is worth to note that if q is symmetric, that is, if $q(\mathbf{m}_k \mid \mathbf{m}_{k+1}) = q(\mathbf{m}_{k+1} \mid \mathbf{m}_k)$ for all $(\mathbf{m}_k,\ \mathbf{m}_{k+1}) \in \mathbb{R}^m \times \mathbb{R}^m$, then (2.38) simplifies to

$$\alpha(\mathbf{m}_k, \mathbf{m}_{k+1}) = \min\left\{1, \frac{\pi(\mathbf{m}_{k+1})}{\pi(\mathbf{m}_k)}\right\}. \tag{2.39}$$

Hence, we immediately accept moves that go towards higher probability and we may also accept moves that take the state to lower probabilities. Algorithm 2 presents a pseudocode for the MH method.
The relative ease of implementation of the MH algorithm, combined with its

Algorithm 2 Metropolis–Hastings Algorithm

Choose the initial value $\mathbf{m}_0 \in \mathbb{R}^m$
Compute $\pi(\mathbf{m}_0)$
for $k = 1, ..., K$ **do**
  Draw sample $\mathbf{m}' \in \mathbb{R}^m$ from the proposal density $q(\mathbf{m}_{k-1} \mid \mathbf{m}')$
  Compute $\pi(\mathbf{m}')$
  Calculate the acceptance ratio:
  $\alpha(\mathbf{m}_{k-1}, \mathbf{m}') = \min\left(1, \frac{\pi(\mathbf{m}')q(\mathbf{m}' \mid \mathbf{m}_{k-1})}{\pi(\mathbf{m}_{k-1})q(\mathbf{m}_{k-1} \mid \mathbf{m}')}\right)$
  Draw $u \sim \mathcal{U}([0, 1])$
  if $u < \alpha(\mathbf{m}_{k-1}, \mathbf{m}')$ **then**
    Accept (Set $\mathbf{m}_k \leftarrow \mathbf{m}'$)
  else
    Reject (Set $\mathbf{m}_k \leftarrow \mathbf{m}_{k-1}$)
  end if
end for

solid theoretical foundation, has contributed to its widespread use in Bayesian inference. However, the drawback of the MH random walk is that its efficiency strongly depends on the direction and magnitude of the proposal distribution. Generic proposals use $q(\mathbf{m}_{k+1} \mid \mathbf{m}_k) = \mathcal{N}(\mathbf{m}_k \mid \sigma^2\mathbf{I})$ and require a suitable step size σ. In general, large steps are highly exploratory, which decreases the acceptance rates, whereas smaller steps lead to higher acceptance rates but worse convergence and low chain mixture speed. A popular strategy is that of continuously adjusting the step size (i.e., the covariance of the Gaussian proposal distribution) based on the observed acceptance probability of the chain in order to improve its rate of statistical convergence. The most common adaptive single chain methods are the adaptive proposal (AP) (Haario et al., 1999), adaptive Metropolis (AM) (Haario et al., 2001) and delayed rejection adaptive Metropolis (DRAM) algorithm (Haario et al., 2006).

One important aspect to take into consideration for a full understanding of the method is convergence diagnosis. After discarding the initial burn-in phase, which is required to allow the chain to explore the search space sufficiently far from the arbitrary initial state, convergence can be assessed with different diagnostic methods (Brooks & Roberts, 1998). However, in general there is no guaranteed diagnostic tool that ensures convergence of an MCMC algorithm. Particularly with a single chain it is difficult to judge whether convergence has been reached and the full parameter space has been explored adequately in pursuit of the target distribution. In fact, single chain methods suffer similar problems as local optimizers, in the sense that the chain can be stuck around a local optimum of the posterior, or traverse only a portion of the target distribution. For this reason, MCMC applications generally involve running multiple parallel Markov chains starting from different initial states, thus exploring a larger portion of the model space. The use of multiple chains can provide useful information about how well

the chains are mixing (i.e., to what extent the output from individual chains are indistinguishable), and opens up the use of a wide range of statistical measures to test whether convergence to a limiting distribution has been achieved, such as the Gelman Rubin diagnostic (Gelman & Rubin, 1992), which evaluates differences in means and variances across chains to assess convergence.

In the past decades, a variety of different approaches have been proposed to increase the efficiency of MCMC simulation and enhance the original MH algorithm. A very popular strategy consists in introducing a variable T, or temperature, to sample a rescaled (or tempered) distribution $\pi(\mathbf{m} \mid T)$. For example, the Simulated Annealing (SA) algorithm, introduced by Kirkpatrick et al. (1983), makes use of a vector $T = \{T_1, \ldots, T_K\}^T$ of strictly decreasing temperatures, which is constructed either in advance or dynamically, and progressively samples from the conditional distributions $\pi(\mathbf{m} \mid T_k)$ as temperature is adjusted from high to low values. At higher temperatures the distribution becomes "flatter", meaning that it is much easier to escape from local maxima as each is less prominent in the landscape (Sambridge, 2014). This strategy is thus considered as a way to improve the efficiency of probabilistic samplers by allowing chains to move more freely in the model space at higher temperatures, while focusing on a more localized search at lower temperatures. Along a similar line, Parallel Tempering (PT) (Sambridge, 2014) uses multiple tempered chains at different temperature levels. PT enhances convergence by regularly exchanging models between chains at two neighboring temperature levels, a process referred to as an 'exchange swap'. In SA, the vector of decreasing temperatures needs to be carefully chosen in order to be efficient and avoid entrapment in local minima, especially for complex multi-dimensional parameter spaces. In PT, this issue is avoided because the balance of effort between higher temperature exploratory search and lower temperature localized search is kept constant throughout the iterations.

Another very popular improvement to the standard MH algorithm is obtained by making use of covariance information, through the Hessian matrix or the gradient vector, using Hamiltonian Markov chains (HMC) (Duane et al., 1987). HMC is part of a broader class of gradient-based MCMC methods which use the gradient of the log-posterior distribution to guide the trajectory, allowing proposals to follow contours of high probability rather than random moves. This approach is particularly effective in high dimensional spaces and for distributions with strong correlations between parameters, as it reduces the random walk behavior typically seen in traditional MCMC methods and allows for large, guided steps in the parameter space. Another famous example of a gradient-based method is the Langevin Monte Carlo (Grenander & Miller, 1994, Besag, 2001), which combines gradient information with stochastic noise to ensure broad exploration of the parameter space.

Chapter 3

Homogenization and Full Waveform Inversion overview

Introduction

The downscaling inversion proposed in this work aims to resolve small-scale geological heterogeneities causing discontinuities in elastic properties, from smooth effective media. Effective elastic properties are computed with the elastic non-periodic homogenization for seismic wave propagation (Capdeville et al., 2010b, Guillot et al., 2010), representing the forward operator. In this regard, downscaling refers to as an inverse homogenization problem.
The theory of homogenization deals with extracting homogeneous effective parameters from systems with microscopic periodic structures. In seismology, this theory has been extended to non-periodic media, relying on the fact that heterogeneous structures of scales much smaller than the minimum wavelength of seismic waves have very limited influence on waves propagating in the medium. Accordingly, the goal of homogenization is to replace the complex heterogeneous Earth by a simpler effective medium such that the resulting wavefield generated by a seismic source remains approximately (i.e., up to a controlled accuracy) unchanged for a given minimum wavelength.
In this field, the first major contribution comes from Backus (1962), which is dedicated to the upscaling of finely layered media for elastic waves in the low-frequency (long-wavelength) regime. In his work, he derives formulae to quantify the seismic anisotropy observed on wave measurements produced by a finely layered isotropic medium. Since then, several improvements and generalizations of Backus' results were made (e.g., Schoenberg & Sayers, 1995, Gold et al., 2000, Grechka, 2003) and a range of different upscaling methods, each based on a variety of assumptions and approximations, were developed. For example, the wavelet-based homogenization method proposed by Engquist & Souganidis (2008), works on the discretized version of the wave equation. Similarly, a homogenization procedure for wave equation, based on its equidistant discretization and the algebraic

isolation of low-wavenumber components, was introduced by Fichtner & Hanasoge (2017). In the context of tomographic modeling, Jordan (2015) formulated an effective theory relying on the second-order Born approximation for the effective elastic parameters of stochastic media with small-scale heterogeneities.
The method used in this work builds upon the two-scale homogenization approach (Babuška, 1976, Bensoussan et al., 1978, Sánchez-Palencia, 1980). This is an asymptotic mathematical approach involving an explicit scale separation between large and small scales and the introduction of two space variables, one for the large scales and one for the small scales. This methodology has been applied across diverse scientific domains, including seismic wave propagation (e.g., Boutin & Auriault, 1993, Fish & Chen, 2001, 2004, Parnell & Abrahams, 2008, Bacigalupo & Gambarotta, 2014). In this context, within the last decades, the two-scale homogenization approach has been adapted to non-periodic media, with no constraint on the shape and size of the heterogeneities (Capdeville & Marigo, 2007, Capdeville et al., 2010b,c, Guillot et al., 2010, Capdeville et al., 2015b). In contrast to the classical two-scale periodic homogenization, the non-periodic homogenization for general heterogeneous media doesn't make any periodicity assumption on the spatial variations of the original medium. To derive the homogenized properties, the method relies on the numerical resolution of an elastostatic problem and a filtering process, where the filter only depends on the minimum wavelength to be propagated in the medium and a precision factor, so that small and large scales results are separated by the waves and the effective properties actually reveal what the waves 'see'. This approach has been already applied to tackle several multi-scale issues, such as to reduce computational complexity in seismic wave simulations for numerical wavefield simulations (Cupillard & Capdeville, 2018), but also in the inversion context (Capdeville et al., 2013, Afanasiev et al., 2016, Capdeville & Métivier, 2018, Mukumoto et al., 2024), such as for downscaling and interpreting tomographic models (Capdeville & Métivier, 2018, Fichtner et al., 2013a, Wang et al., 2013a, Bodin et al., 2015, Alder et al., 2017, Faccenda et al., 2019, Hedjazian et al., 2021, Santos et al., 2024).
In the light of these works, the inverse homogenization framework is applied here to interpret effective (homogenized) FWI models. As demonstrated by Capdeville & Métivier (2018) with numerical experiments, introducing the homogenization operator in the FWI inverse problem (HFWI) and inverting for smooth effective media, compared to inverting for sharp discontinuities reduces the non-linearity of the problem, and also the dependence of the inversion result to the adopted mesh and parametrization. We rely on these findings to set up a downscaling framework solved within a Bayesian inference scheme where the effective FWI elastic medium (assumed to represent the effective properties of the true Earth) takes the role of the observations to inform, together with some external a *priori* information, all the possible small-scale elastic media compatible with it.
As non-periodic homogenization and HFWI represent two essential building blocks of the downscaling inversion, the forward operator and the observed data respec-

tively, these are examined in detail in this chapter. We first recall some general features of the two-scale homogenization theory in the context of 1-D periodic and non-periodic media following Capdeville et al. (2020). Then, we extend these results from 1-D to 2-D, which is the case used for the downscaling inversion proposed in this work, and we illustrate the 2-D non periodic homogenization with an application to a subsurface model. The second part is dedicated to FWI, HFWI and associated uncertainty estimation. We first introduce elements of the FWI problem and HFWI, following Capdeville & Métivier (2018), and we present a synthetic application where the final FWI result is homogenized to obtain its smooth equivalent version. Then, because a key aspect in the probabilistic setting of the downscaling inversion is the estimation of observational uncertainty, a strategy to compute uncertainty on the FWI solution and how to map it to its effective equivalent is investigated and discussed.

3.1 Homogenization

3.1.1 1-D periodic homogenization

To introduce the two-scale periodic homogenization we use the simple example of a wave propagating in a 1-D bar with periodic heterogeneities. The elastic wave equation in 1-D is defined as

$$\begin{aligned} \rho(x)\partial_{tt}u(x,t) - \partial_x\sigma(x,t) &= f(x,t), \\ \sigma(x,t) &= E(x)\partial_x u(x,t), \end{aligned} \tag{3.1}$$

and by its initial and boundary conditions. Here, $u(x,t)$ is the displacement of the particle along the $\widehat{x}$ direction as function of space x and time t; $\sigma(x,t)$ is the stress; ρ and E are the density and the Young modulus (or the elastic coefficient) of the bar; and $f(x,t)$ is the external source applied in the bar along the $\widehat{x}$ axis. We consider the source as a point force

$$f(x,t) = F_s\delta(x - x_s)g(t), \tag{3.2}$$

or a moment source

$$f(x,t) = -M_s\partial_x\delta(x - x_s)g(t), \tag{3.3}$$

where F_s is the force strength, M_s is the scalar moment, x_s is the souce location and $g(t)$ is the source time wavelet. In the following, the time dependence t is dropped to ease the notations.

We start by assuming that, given a source with a limited frequency band (i.e., with a maximum frequency f_{max}), we can estimate the wavefield minimum wavelength in the far-field as $\lambda_m = V_{min}/f_{max}$, where V_{min} is the minimum wave velocity. We then assume that the mechanical properties of the bar $\rho(x)$ and $E(x)$ vary periodically with a l-periodicity along $\widehat{x}$ and we assume that l is smaller than λ_m

(Fig. 3.1).
For periodic media, the distinction between small and large scales is made explicit by the periodicity of the mechanical properties: everything that is periodic is considered small scale and all the rest are large scales. Accordingly, we introduce the following quantities: a scaling parameter $\varepsilon = \frac{l}{\lambda_m}$, which measures how small scales are compared to λ_m, and a small-scale space-variable $y = \frac{x}{\varepsilon}$ used to explicitly separate large scales x from small scales y when solving the wave equation (3.1). In practice, when $\varepsilon \to 0$, that is the periodicity is infinitely small compared to λ_m, any change in y induces only a very small change in x. This leads to the assumption that x and y can be treated as independent variables.
The aim of the two-scale homogenization is to find solutions to the wave equation as an asymptotic expansion in ε:

$$\begin{aligned} u^\varepsilon(x) &= \sum_{i=0}^{\inf} \varepsilon^i u^i(x,y), \\ \sigma^\varepsilon(x) &= \sum_{i=-1}^{\inf} \varepsilon^i \sigma^i(x,y), \end{aligned} \tag{3.4}$$

in which coefficients u^i and σ^i depend on both space variables x and y and are λ_m-periodic in y. Note that i are powers in ε but are expansion indices in u and σ. The two-scale assumption and the ansatz (3.4) are mathematically justified with the convergence theorem: as ε becomes smaller and smaller, the true problem converges to the two-scale problem (Nguetseng, 1989, Allaire, 1992). Such a demonstration provides mathematical foundations for the homogenization theory.
The two-scale setting allows us to redefine the quantities ρ and E as the unit cell mechanical properties

$$\begin{aligned} \rho(y) &= \rho^\varepsilon(\varepsilon y), \\ E(y) &= E^\varepsilon(\varepsilon y), \end{aligned} \tag{3.5}$$

which are λ_m-periodic and independent of ε. In other words, we now see the medium as containing only small scales which are repeated in space with a λ_m-periodicity.
The introduction of the two-scale variables x and y also implies that the spatial derivative must be rewritten:

$$\frac{\partial}{\partial_x} \to \frac{1}{\varepsilon}\frac{\partial}{\partial_y} + \frac{\partial}{\partial_x}. \tag{3.6}$$

Thus, introducing the expansions (3.4) in the wave equation, using (3.6) and considering the source as a constant term (i.e., independent of ε, even if this is

not always true, see Capdeville et al. (2010b,c)), we get

$$\begin{aligned} \rho\partial_{tt}u^i - \partial_x\sigma^i - \partial_y\sigma^{i+1} &= f, \\ \sigma^i &= E(\partial_x u^i + \partial_y u^{i+1}). \end{aligned} \tag{3.7}$$

The solution of the homogenization problem is found by solving these equations for each $i \geq -2$ and $i \geq -1$ respectively.

3.1.1.1 Summary of results

In the following we summarize the main results of the 1-D periodic homogenization (Capdeville et al., 2010a):

- the zeroth-order terms u^0 and σ^0 do not depend on the small scale variable y. This result formalizes the poor sensitivity of the wavefield to small heterogeneities.

- u^0 and σ^0 are solution of effective equations at order 0,

$$\begin{aligned} \rho^*\partial_{tt}u^0 - \partial_x\sigma^0 &= f, \\ \sigma^0 &= E^*\partial_x u^0, \end{aligned} \tag{3.8}$$

  which are the same as the original wave equation (3.1), but with the effective properties (which do not depend upon y)

$$\rho^* = \langle\rho\rangle\,, \quad E^* = \left\langle E(1+\partial_y\chi^1)\right\rangle. \tag{3.9}$$

  In (3.9), $\langle\cdot\rangle$ denotes the average over the periodic cell, which for any function $h(x,y)$ λ_m-periodic in y is defined as

$$\langle h\rangle(x) = \frac{1}{\lambda_m}\int_0^{\lambda_m} h(x,y)dy. \tag{3.10}$$

  The term χ^1 in (3.9) is called the first order periodic corrector. It accounts for the heterogeneities at the receiver through its dependence on y, and it is the solution of the so-called cell problem:

$$\partial_y[E(1+\partial_y\chi^1)] = 0, \tag{3.11}$$

  defined on the cell with periodic boundary conditions. A general analytical solution to the cell problem (3.11) exists and leads to estimation of the effective Young modulus E^* through

$$E^* = \left\langle \frac{1}{E} \right\rangle^{-1}. \tag{3.12}$$

- Contrary to u^0 and σ^0, the higher-order terms of (3.4) contain small-scale variations. For instance, the first-order displacement term is y-dependent through the first-order corrector χ^1:

$$u^1(x,y) = \chi^1(y)\partial_x u^0(x) + \left\langle u^1 \right\rangle (x). \tag{3.13}$$

  Examples of the contribution of this last term to the wavefield can be found in Capdeville et al. (2010c).

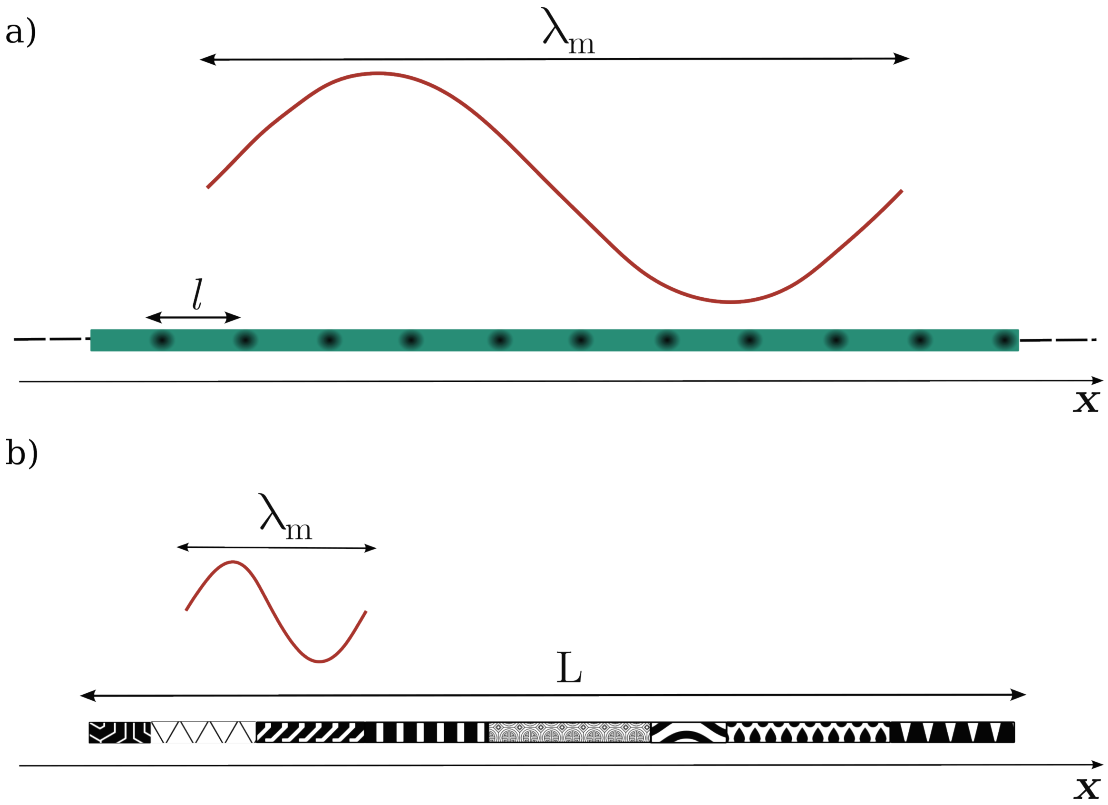

Figure 3.1: a) 1-D periodic homogenization framework: a wavefield having a minimum wavelength λ_m propagates in an infinite 1-D medium made of periodic cells whose size l is much smaller than λ_m. b) 1-D non-periodic homogenization framework: a wavefield having a minimum wavelength λ_m propagates in a finite 1-D medium of size L. Modified from Cupillard & Capdeville (2018).

3.1.2 1-D non-periodic homogenization

For non-periodic media, it is still reasonable to expect, to some sense, that heterogeneities whose size is much smaller than λ_m have a little influence on the wavefield u and that a homogenization procedure can be developed. However, while for the classical periodic homogenization the periodicity makes it possible

to simply and efficiently define the small scales, for the non-periodic case establishing the boundary between small scales and large scales is not straightforward. To find the solution of our homogenization problem, we consider the whole non-periodic bar, whose length is denoted by L, as a single periodic cell and we apply results obtained in the previous section. In this case, the cell can contain various sizes of heterogeneity (Fig. 3.1), so ε defined as the ratio of the periodic cell size L to the minimum wavelength λ_m is no longer a relevant scaling parameter to separate the scales. Therefore, we introduce a user-defined wavelength λ_0 and we measure its position relative to λ_m with

$$\varepsilon_0 = \frac{\lambda_0}{\lambda_m}. \tag{3.14}$$

The wavelength λ_0 therefore is analogous to l in the periodic case, in the sense that all the heterogeneities whose size is smaller than λ_0 are considered small.
While ε is purely formal, ε_0 is a parameter that indicates the degree of smoothness of the homogenized model and accuracy of the homogenized solution: a small ε_0 corresponds to a homogenized model with a lot of details and a precise solution; a large ε_0 corresponds to a smooth homogenized model and a less precise solution. In the following, we assume $\varepsilon \leqslant \varepsilon_0 < 1$.
Given this definition of ε and ε_0, we can define the small-scale variable $y = \dfrac{x}{\varepsilon_0}$ and we can treat x and y independently. To separate these scales, the non-periodic homogenization method requires the use of a low-pass filter $\mathcal{F}^{\varepsilon_0}$, which for any function $h(x) : \mathbb{R} \to \mathbb{R}$ is given by

$$\mathcal{F}^{\varepsilon_0}\{h(x)\} = \int_{\mathbb{R}} h(x')w^{\varepsilon_0}(x - x')dx'. \tag{3.15}$$

where the wavelet w^{ε_0} in its spectral domain $\overline{w}$ is

$$\overline{w}(k) = \begin{cases} 1 & \text{for } k \leq \ k_0 = \dfrac{2\pi}{\lambda_0}, \\ 0 & \text{otherwise.} \end{cases}$$

This filter thus removes all scales smaller than λ_0. Based on this, we can define a functional space Ψ of two-scale functions $h(x, y) : \mathbb{R}^2 \to \mathbb{R}$ that are λ_m-periodic in y, such that $\mathcal{F}_y^{\varepsilon_0}(h)(x, y) = \langle h \rangle(x)$, where $\mathcal{F}_y^{\varepsilon_0}$ is the filter (3.15) applied to variable y. We refer to Capdeville et al. (2010c) for a precise definition of this two-scale functional space.
The non-periodic homogenization needs to determine the two-scale cell properties $\rho^{\varepsilon_0}(x, y), E^{\varepsilon_0}(x, y)$ such that

$$\begin{aligned} \rho(x) &= \rho^{\varepsilon_0}(x, \frac{x}{\varepsilon_0}), \\ E(x) &= E^{\varepsilon_0}(x, \frac{x}{\varepsilon_0}), \end{aligned} \tag{3.16}$$

and such that the coefficients $u^{\varepsilon_0,i}$ and $\sigma^{\varepsilon_0,i}$ of the asymptotic expansions belong to the functional space Ψ, meaning that only fine-scale variations must appear in y. In practice, we solve the asymptotic equations (3.7) and find the necessary conditions on $(\rho^{\varepsilon_0}, E^{\varepsilon_0})$ so that a solution exists; we then construct the cell properties such that these conditions are fulfilled.
In the non-periodic case, the asymptotic equations (3.7) remain unchanged; the only difference with the periodic case is that now ρ^{ε_0} and E^{ε_0} depend on ε_0 and on both (x, y), while they depend only on y in the periodic case. Because the y periodicity is kept in Ψ, the resolution of these equations is almost the same as in the periodic case.

3.1.2.1 Summary of results

Here, we summarize the key findings of the non-periodic homogenization procedure (Capdeville et al., 2010a).

- Similarly to the periodic case, it turns out that the zeroth-order displacement u_0 and stress σ_0 do not depend on y.
- Again, u_0 and σ_0 are the solution of the so-called homogenized problem

$$\begin{aligned} \rho^{\varepsilon_0 *}\partial_{tt}u^0 - \partial_x\sigma^0 &= f, \\ \sigma^0 &= E^{\varepsilon_0 *}\partial_x u^0, \end{aligned} \tag{3.17}$$

  which is a classical elastodynamic problem involving the effective properties

$$\rho^{\varepsilon_0 *}(x) = \langle \rho^{\varepsilon_0}(x, y) \rangle \,, \tag{3.18}$$

  and

$$E^{\varepsilon_0 *}(x) = \langle E^{\varepsilon_0}(x, y)[1 + \partial_y \chi^{\varepsilon_0}(x, y)] \rangle \,. \tag{3.19}$$

  The first-order corrector $\chi^{\varepsilon_0}(x, y)$, which also depends on ε_0, must lie in Ψ and satisfy

$$\partial_y [E^{\varepsilon_0}(1 + \partial_y \chi^{\varepsilon_0})] = 0 \tag{3.20}$$

  with periodic boundary conditions. In 1-D, an analytical solution of (3.20) is available. This solution exists if $\dfrac{1}{E^{\varepsilon_0}}$ lies in Ψ (which is the condition to meet for building $E^{\varepsilon_0}(x, y)$ properly), so that χ^{ε_0} is in Ψ and

$$\partial_y \chi^{\varepsilon_0}(x, y) = -1 + \left\langle \frac{1}{E^{\varepsilon_0}} \right\rangle^{-1}(x) \frac{1}{E^{\varepsilon_0}(x, y)}. \tag{3.21}$$

  This last equation implies that (3.19) reduces to

$$E^{\varepsilon_0^*}(x) = \left\langle \frac{1}{E^{\varepsilon_0}(x, y)} \right\rangle^{-1}. \tag{3.22}$$

Similarly, we meet the condition that $\rho^{\varepsilon_0}(x,y)$ must lie in Ψ for building $\rho^{\varepsilon_0}(x,y)$ properly. Based on these constraints, we can propose

$$E^{\varepsilon_0}(x,y) = \left[\mathcal{F}^{\varepsilon_0}\left\{\frac{1}{E}\right\}(x) + \left[\frac{1}{E} - \mathcal{F}^{\varepsilon_0}\left\{\frac{1}{E}\right\}\right](\varepsilon_0 y)\right]^{-1} \tag{3.23}$$

and

$$\rho^{\varepsilon_0}(x,y) = \mathcal{F}^{\varepsilon_0}\{\rho\}(x) + [\rho - \mathcal{F}^{\varepsilon_0}\{\rho\}](\varepsilon_0 y). \tag{3.24}$$

Finally, including (3.23) into (3.19) leads to

$$E^{\varepsilon_0 *}(x) = \left[\mathcal{F}^{\varepsilon_0}\left\{\frac{1}{E}\right\}(x)\right]^{-1}, \tag{3.25}$$

and including (3.24) into (3.18) yields

$$\rho^{\varepsilon_0 *}(x) = \mathcal{F}^{\varepsilon_0}\{\rho\}(x). \tag{3.26}$$

Hence, the zeroth-order effective properties of any non-periodic 1-D medium can be computed by filtering the density and the inverse of the Young modulus.

- In the non-periodic case, it is not possible to go to orders higher than the leading order of the expansion and the first-order corrector term (see Eq. 3.13). In general, some of the order 2 correctors do not belong to Ψ, which implies that there is no general non-periodic solution beyond order 1, at least in this form. In practice, this is not a strong limitation as only the leading and first order are useful for most applications (Capdeville et al., 2020).

3.1.3 2-D non-periodic homogenization

In 2-D, the elastic wave equation is defined as

$$\begin{aligned} \rho\partial_{tt}\mathbf{u} - \nabla\cdot\boldsymbol{\sigma} &= \mathbf{f}, \\ \boldsymbol{\sigma} &= \mathbf{c} : \boldsymbol{\epsilon}(\mathbf{u}). \end{aligned} \tag{3.27}$$

In this case, Hooke's law (second equation) involves a fourth-order elastic tensor $\mathbf{c}$ to relate the stress $\boldsymbol{\sigma}$ and the strain $\boldsymbol{\epsilon}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} +^T \nabla\mathbf{u})$, such that $\boldsymbol{\sigma}_{ij} = \sum_{kl} c_{ijkl}\epsilon_{kl}$. Together with the initial and boundary conditions, equations in (3.27) uniquely specify the displacement field $\mathbf{u}$ in the domain Ω. The elastic tensor $\mathbf{c}$ is positive-definite and satisfies the symmetries $c_{ijkl} = c_{jikl} = c_{ijlk} = c_{klij}$, which reduces the maximum number of independent parameters necessary to characterize $\mathbf{c}$ to 6 in 2-D.
The homogenization theory developed for 1-D non-periodic media is still valid up

to the zeroth-order effective properties

$$\rho^{\varepsilon_0 *}(\mathbf{x}) = \langle \rho^{\varepsilon_0}(\mathbf{x}, \mathbf{y}) \rangle \,, \tag{3.28}$$

and

$$\mathbf{c}^{\varepsilon_0 *}(\mathbf{x}) = \left\langle \mathbf{c}^{\varepsilon_0}(\mathbf{x}, \mathbf{y}) : \left[\mathbf{I} + \frac{1}{2} (\nabla_y \boldsymbol{\chi}^{\varepsilon_0}(\mathbf{x}, \mathbf{y}) +^T \nabla_y \boldsymbol{\chi}^{\varepsilon_0}(\mathbf{x}, \mathbf{y})) \right] \right\rangle , \tag{3.29}$$

where $\mathbf{I}$ is the fourth order identity tensor, ^{T} is the transpose operator and $\langle \cdot \rangle$ is the average on $\mathbf{y}$ of any function $h(\mathbf{x}, \mathbf{y}) : \mathbb{R}^{2\times 2} \to \mathbb{R}$ over the periodic cell. The third-order tensor $\boldsymbol{\chi}^{\varepsilon_0}$ necessarily belongs to the functional space Ψ_2 (analogous to Ψ in 2-D) and it is the solution of the following cell problem:

$$\nabla_y \cdot \left\{ \mathbf{c}^{\varepsilon_0} : \left[\mathbf{I} + \frac{1}{2} (\nabla_y \boldsymbol{\chi}^{\varepsilon_0} +^T \nabla_y \boldsymbol{\chi}^{\varepsilon_0} \right] \right\} = \mathbf{0}, \tag{3.30}$$

with periodic boundary conditions.
Differently from the 1-D case, an analytical solution to equation (3.30) only exists for layered transverse isotropic media, as demonstrated by Guillot et al. (2010). In the general case, a different procedure, inspired by Papanicolaou (1979), has been proposed to solve the cell problem. This procedure is summarized in the following section (Capdeville et al., 2010b).

3.1.3.1 Construction of the effective elastic tensor and density

While the effective density can be simply obtained with

$$\rho^{\varepsilon_0 *} = \mathcal{F}^{\varepsilon_0}(\rho). \tag{3.31}$$

The construction of $\mathbf{c}^{\varepsilon_0 *}$ is not trivial and requires the introduction of two intermediate fields in Ψ_2

$$\mathbf{G}^{\varepsilon_0} = \mathbf{I} + \frac{1}{2} (\nabla \boldsymbol{\chi}^{\varepsilon_0} +^T \nabla \boldsymbol{\chi}^{\varepsilon_0}), \tag{3.32}$$

and

$$\mathbf{H}^{\varepsilon_0} = \mathbf{c}^{\varepsilon_0} : \mathbf{G}^{\varepsilon_0}. \tag{3.33}$$

The two tensors $\mathbf{G}$ and $\mathbf{H}$ express, respectively, the stress and strain associated with the first-order corrector $\boldsymbol{\chi}^{\varepsilon_0}$. The procedure to build $\mathbf{c}^{\varepsilon_0 *}$ can be summarized with the following steps:

- Step 1. We first solve numerically the cell problem with periodic boundary conditions for the entire domain using $\mathbf{c}(\mathbf{x})$ instead of $\mathbf{c}^{\varepsilon_0 *}(\mathbf{x}, \mathbf{y})$ (i.e., all scales of the elastic tensor are considered to be small):

$$\nabla \cdot \left\{ \mathbf{c} : \left[\mathbf{I} + \frac{1}{2} (\nabla \boldsymbol{\chi}_s +^T \nabla \boldsymbol{\chi}_s \right] \right\} = \mathbf{0}, \tag{3.34}$$

to find the initial guess corrector $\boldsymbol{\chi}_s$. Solving (3.34) usually requires a numerical solver (apart for the simple layered case). Two solvers can be used: a solver based on a finite element method (Capdeville et al., 2010b, Cupillard & Capdeville, 2018) or based on an iterative Fast Fourier Transform (FFT) scheme (Capdeville et al., 2015b) derived from Moulinec & Suquet (1998), Michel et al. (1999).

- Step 2. Once $\boldsymbol{\chi}_s$ is obtained, we compute the initial strain $\mathbf{G}_s(\mathbf{x})$ (similarly to 3.32) and stress $\mathbf{H}_s(\mathbf{x})$ tensor (similarly to 3.33) associated with $\boldsymbol{\chi}_s$ using $\mathbf{c}(\mathbf{x})$.

- Step 3. The effective elastic tensor $\mathbf{c}^{\varepsilon_0 *}$ can then be directly obtained using $\mathcal{F}^{\varepsilon_0}$ as

$$\mathbf{c}^{\varepsilon_0 *}(\mathbf{x}) = \mathcal{F}^{\varepsilon_0}(\mathbf{H}_s) : \mathcal{F}^{\varepsilon_0}(\mathbf{G}_s)^{-1}(\mathbf{x}). \tag{3.35}$$

From $\mathbf{H}_s$ and $\mathbf{G}_s$, one can also build the two-scale tensors $\mathbf{G}^{\varepsilon_0}$ and $\mathbf{H}^{\varepsilon_0}$. These would be necessary to obtain the first-order corrector $\boldsymbol{\chi}^{\varepsilon_0}$ through the construction of $\mathbf{c}^{\varepsilon_0} = \mathbf{H}^{\varepsilon_0} : \mathbf{G}^{\varepsilon_0^{-1}}$ and (3.30), in order to access the first-order displacement. Moreover, additional considerations should be made for effective boundary conditions and the source correction term (Capdeville & Marigo, 2008, 2013), which we neglect in this study because we will only consider the leading order of the expansion.

3.1.4 Numerical example

In this section we illustrate the 2-D non-periodic homogenization technique with an application to a realistic geological medium. The reference model, shown in Figure 3.2a, is 4 km × 2.5 km large and it is composed of multiple faulted and folded horizons with a fine-layered stratigraphy. This model was built from the analog lab model of Colletta et al. (1991) by assigning constant velocity values in each major layer and adding thin-layering structures inside each of them (Irakarama et al., 2019). The medium is elastic and isotropic with compressional velocity V_p ranging from 1 500 m s^{-1} to 4 800 m s^{-1} and shear velocity V_s ranging from 937 m s^{-1} to 3 000 m s^{-1}. A homogeneous density equal to 2 600 kg m^{-3} is defined for the whole medium.
In this study, we employ the FFH method from Capdeville et al. (2015b) to solve the cell problem (3.34) and derive the effective elastic tensor $\mathbf{c}^*$ (the superscript ε_0 is omitted hereafter). We present homogenization results in terms of effective shear-wave velocity V_s^* and effective anisotropy. For this purpose we project the anisotropic elastic tensor $\mathbf{c}^*$ to the nearest isotropic tensor $\mathbf{c}^{*,iso}$, following Browaeys & Chevrot (2004), Capdeville et al. (2020), and we define the effective shear-wave velocity as

$$V_s^*(\mathbf{x}) = \sqrt{(c_{1212}^{*,iso}(\mathbf{x})/\rho(\mathbf{x}))}, \tag{3.36}$$

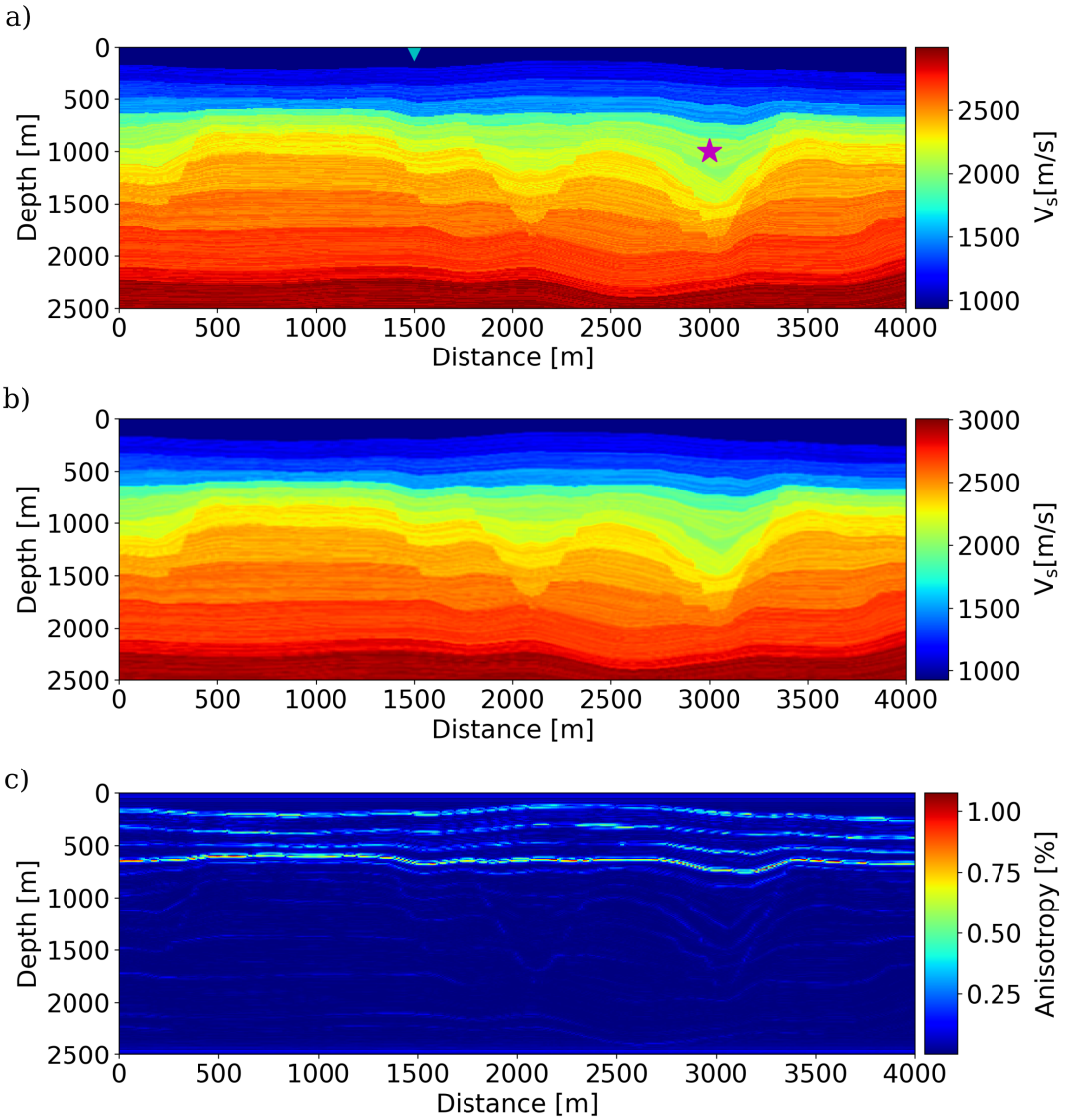

Figure 3.2: a) S-wave velocity of the original fine-scale elastic medium. For the seismic wave simulation, a receiver (green triangle) is positioned on the surface, while the seismic source (blue star) is located within the medium. b) Effective S-wave velocity (Eq. 3.36) obtained by homogenizing the original fine-scale medium with $\varepsilon_0 = 0.5$. c) Effective anisotropy (Eq. 3.37) of the homogenized medium.

and the amount of effective anisotropy as

$$aniso(\mathbf{x}) = \frac{\sqrt{\sum_{ijkl}(c^{iso}_{ijkl}(\mathbf{x}) - c_{ijkl}(\mathbf{x}))^2}}{\sqrt{\sum_{ijkl}(c^{iso}_{ijkl}(\mathbf{x}))^2}}. \tag{3.37}$$

Because fine-scale isotropic heterogeneities lead to anisotropic effective heterogeneities (Backus, 1962, Capdeville et al., 2020), even if the original model is isotropic, its effective properties are anisotropic. The "smoothness" of the effec-

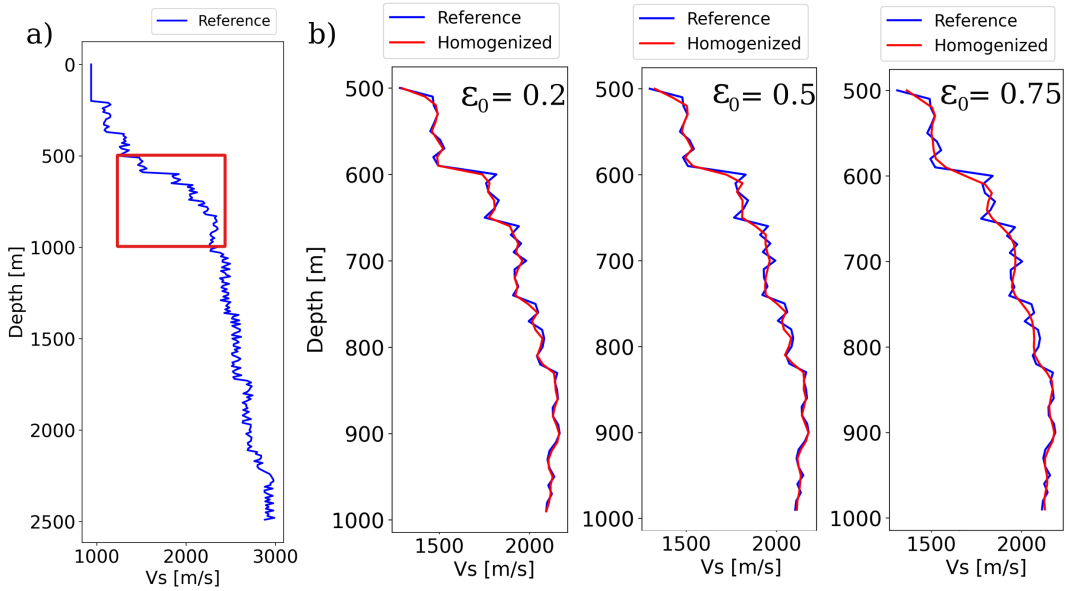

Figure 3.3: a) S-wave velocity profile, extracted at x = 1 000 m, from the original fine-scale elastic medium. b) Zoomed-in portion of the S-wave original velocity, indicated by the red box in (a), and S-wave velocity of the effective medium, homogenized using $\varepsilon_0 = 0.2$, $\varepsilon_0 = 0.5$ and $\varepsilon_0 = 0.75$.

tive properties depend on the user-defined scale λ_0 selected for the separation of microscopic and macroscopic scales, which is linked to λ_m through ε_0 (3.14). According to Capdeville et al. (2010b), for a fixed λ_m value, an ε_0 lying between 0.25 and 0.5 would be a good choice, depending on the desired accuracy, on the signal duration and on the particular elastic model in consideration. Figure 3.2b,c shows the effective shear velocity (3.36) and anisotropy (3.37) computed with the FFH approach using a regularly spaced grid of 1 m × 1 m and a scaling factor $\varepsilon_0 = 0.5$, thus no spatial variations smaller than $\lambda_0 = \lambda_m/2$ are present in the effective medium. This dense FFT grid is adopted to efficiently sample the medium and obtain more accurate homogenization results. The effective models (Fig. 3.2b,c) show interfaces that are blurred and a higher percentage of anisotropy localized around interfaces characterized by strong velocity contrasts. Smaller amount of anisotropy is also observed along faults and at velocity discontinuities within major layers. Figure 3.3a shows a V_s profile, extracted from the original model at $x = 1\,000$ m along the horizontal direction. A zoomed-in portion of the original (indicated by the red box in Fig. 3.3a) and associated homogenized V_s^* profiles, computed with three different values of ε_0 (0.2, 0.5 and 0.75), are provided in Figure 3.3b, showing how large ε_0 lead to smoother elastic media (lower resolution).

To assess the quality of the effective medium (and thus the choice of the scale separation factor ε_0), we perform a seismic wave simulation and we compare the obtained seismograms with traces computed in the original fine-scale model. We

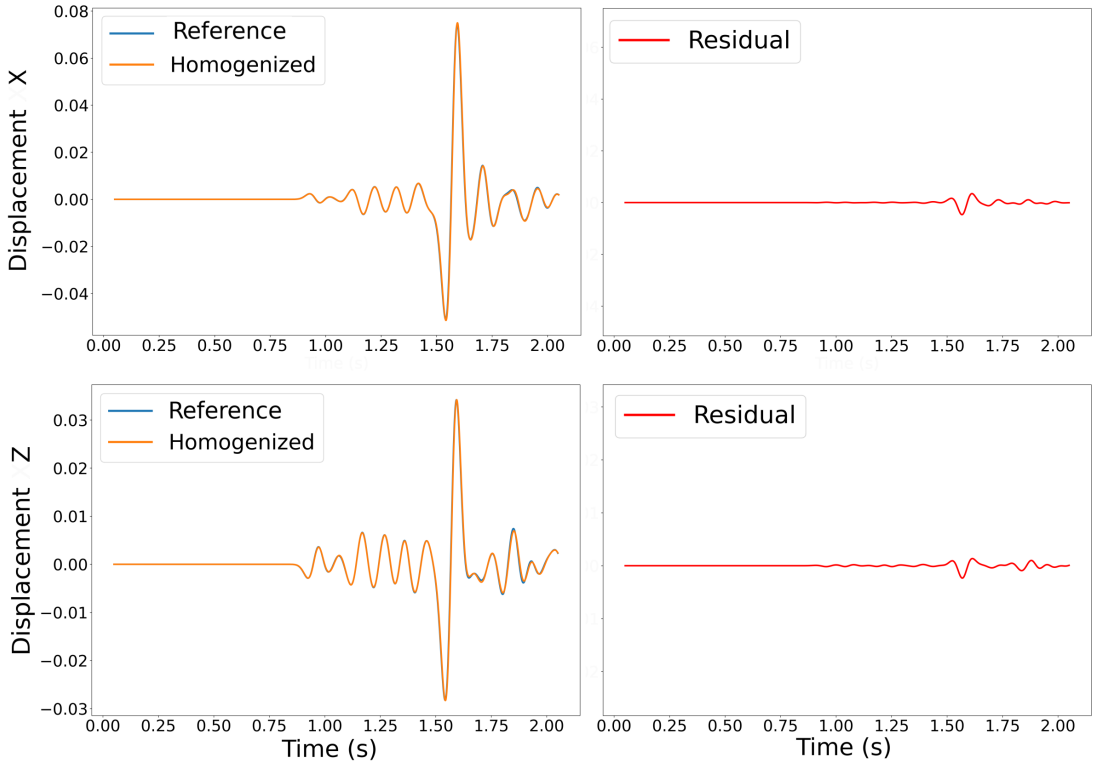

Figure 3.4: Left: x- and z-components of the displacements $\mathbf{u}$ (blue) in the reference medium, and $\mathbf{u}^*$ (orange) in the homogenized medium (using $\varepsilon_0 = 0.5$), recorded at a receiver located on the surface at a $x = 1\,500$ m along the horizontal direction. Right: Difference between $\mathbf{u}$ and $\mathbf{u}^*$.

numerically simulate wave propagation (P-SV waves) based on the spectral element method (SEM) (e.g., Komatitsch & Vilotte, 1998, Komatitsch & Tromp, 1999, Chaljub et al., 2007) with the open source package SPECFEM2D (Tromp et al., 2008, Peter et al., 2011). With this approach, displacements are approximated using Lagrange polynomials at Gauss-Lobatto-Legendre (GLL) points. Free surface boundary condition is applied to the top of the model while absorbing boundary conditions (Stacey, 1988) are applied to the other three sides. A receiver is placed on the surface and a source is placed inside the medium at a depth of 1 000 m, as shown in Figure 3.2a. In both the original and homogenized cases, we compute the wavefield induced by a vertical point-force. The source-time function is a Ricker wavelet (second derivative of a Gaussian) with a source central frequency of 8 Hz (maximum frequency of 20 Hz). The minimum wavelength λ_m is around 40 m, 4 times bigger than the size of heterogeneities in the fine-scale medium (Fig. 3.3). For the spectral element wavefield simulation in the original isotropic medium, to make the computation accurate, we design a mesh that takes into account the fine-scale heterogeneities in Figure 3.2a, i.e., 10 m $\times$ 10 m elements and a polynomial degree equal to 4 in each direction.

For the corresponding (fully anisotropic) effective medium, the mesh could be designed based on the homogenization scale λ_0, allowing for much larger element sizes and, consequently, a reduced computational cost. However, because we are not interested here in quantifying the computational gain, but just in waveform comparisons, we adopt the same mesh for both original and effective media. The displacement measured by the receiver in the original medium ($\mathbf{u}$) and in the effective medium obtained using $\varepsilon_0 = 0.5$ ($\mathbf{u}^*$) are shown in Figure 3.4. One observes that the effective medium generates waveforms that are almost identical to the original solution, with well-matching amplitudes and phases. This means that $\varepsilon_0 = 0.5$ represents a proper choice since it captures heterogeneity scales small enough to accurately reproduce the wavefield. For both vertical and horizontal components the correlation between $\mathbf{u}$ and $\mathbf{u}^*$ is around 0.99 (with the value of 1 indicating a perfect match) and the L^2 norm difference is around 0.06. The existence of small waveform misfit (Fig. 3.4, right-hand side) can be explained by the fact that we only use zeroth-order homogenization, which implies that no corrector is used. Indeed, in theory the accuracy of recovered wavefield could be further improved by considering higher order terms. As demonstrated by Capdeville et al. (2015b), Cupillard & Capdeville (2018), waveform discrepancy also decreases with decreasing ε_0, since more details of the original model are preserved (Fig. 3.3). With the parametrization chosen for this application a smaller ε_0, equal to 0.25, offers a negligible improvement in the accuracy of recovered wavefield, at the expense of a higher computational cost. This points out the poor wavefield sensitivity to the fine-scale heterogeneities that have been smoothed out with $\varepsilon_0 = 0.5$.

3.2 Full waveform inverse problem

The aim of FWI is to recover information on the physical properties of an Earth model $\mathbf{m} = (\rho, \mathbf{c})$ or (ρ, V_p, V_s) across a given domain $\mathbf{\Omega}$ from seismic wave observations $\mathbf{d}$. This is a nonlinear inverse problem (Section 2.1) since the relationship between $\mathbf{d}$ and the model $\mathbf{m}$ is nonlinear.
In the frame of local optimization, FWI is defined as the minimization problem

$$\tilde{\mathbf{m}} = \underset{\mathbf{m}\in\mathcal{M}}{\arg\min}\, \phi(\mathbf{m}), \tag{3.38}$$

$$\phi(\mathbf{m}) = \sum_{r,s} \int_0^T (\mathbf{d}_s(\mathbf{x}_r, t) - \mathbf{u}_s(\mathbf{x}_r, t, \mathbf{m}))^2 dt, \tag{3.39}$$

where $\phi(\mathbf{m})$ is the least-squares misfit function measuring the misfit between observations $\mathbf{d}_s(\mathbf{x}_r, t)$, generated by source number s, ($s \in 1, \ldots, N_s$) and recorded by receiver number r ($r \in 1, \ldots, N_r$) for a time duration T, and the modeled displacement $\mathbf{u}_s(\mathbf{x}_r, t, \mathbf{m})$ obtained by solving the wave equation (3.27) on $\mathbf{m}$. $\mathcal{M}$ defines the model space, i.e., the ensemble of all the physically admissible Earth models.
This optimization problem is solved by employing the iterative optimization methods described in Section 2.2. Thus, for a given model of dimension M, at iteration k, the updated model is

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k. \tag{3.40}$$

In its most general form, $\Delta\mathbf{m}_k$ is given by the Newton equation (2.10)

$$\Delta\mathbf{m}_k = -(\nabla^2_{\mathrm{m}}\phi_k)^{-1}\nabla_{\mathrm{m}}\phi_k, \tag{3.41}$$

where $\nabla_{\mathrm{m}}\phi_k$ and $\nabla^2_{\mathrm{m}}\phi_k$ are respectively the the gradient ($\mathbf{g}$) and Hessian ($\mathbf{H}$) of the misfit function ϕ with respect to the model parameters $\mathbf{m}$.
These two key quantities can also be expressed as

$$\mathbf{g} = \frac{\delta\phi(\mathbf{m})}{\delta\mathbf{m}} = \left(\frac{\delta\mathbf{u}(\mathbf{m})}{\delta\mathbf{m}}\right)^T (\mathbf{d} - \mathbf{u}(\mathbf{m})) = \mathbf{J}^T\mathbf{\Delta d}, \tag{3.42}$$

and

$$\mathbf{H} = \frac{\delta^2\phi(\mathbf{m})}{\delta\mathbf{m}^2} = \left[\mathbf{J}^T\mathbf{J} + \left(\frac{\delta\mathbf{J}}{\delta\mathbf{m}}\right)^T \mathbf{\Delta d}\right], \tag{3.43}$$

where T denotes the transpose operator, $\mathbf{\Delta d}$ the difference between the simulated and the recorded data, and $\mathbf{J}$ the Jacobian matrix or Fréchet derivative matrix. Because the direct calculation of $\mathbf{J}$ is computationally untractable in the majority of applications, the gradient is computed with the adjoint state strategy to avoid explicitly calculating any partial derivatives of the data (Tromp et al., 2005,

Plessix, 2006). This approach requires, instead, solving the forward problem and its adjoint problem just once per gradient calculation. In practice, the adjoint wavefield is generated by adjoint sources, which use the time-reversed residual waveforms as source time functions. These sources are activated simultaneously at the receiver positions and their fields superimpose to produce the total adjoint field. The gradient is then built as a zero-lag correlation of the forward propagated and adjoint backpropagated wavefields and as a summation over sources.
For the Hessian calculation, the second term in (3.43) is usually neglected for nonlinear inverse problems (for linear inverse problems it is equal to zero) resulting in the Gauss-Newton Hessian approximation. The Gauss-Newton method and various quasi-Newton methods, which use iterative approaches to directly update an initial approximation, are employed to approximate the Hessian matrix (Section 2.2.1). This matrix, as it directly measures the change of the data misfit function ϕ with respect to changes in pairs of model parameters $\mathbf{m}$, defines the local curvature of ϕ, and consequently it is often used for resolution and uncertainties analysis in FWI (Fichtner & Trampert, 2011).

3.2.1 Local uncertainty estimation

In the context of local optimization, in multiple studies the uncertainty associated to the FWI solution $\tilde{\mathbf{m}}$ is derived by computing local uncertainties around the global minimum associated to this solution through the inverse Hessian matrix, which approximates the posterior covariance matrix (e.g., Fichtner & Trampert, 2011, Martin et al., 2012b, Bui-Thanh et al., 2012, Gejadze et al., 2013, Bousserez et al., 2015, Fichtner & Leeuwen, 2015, Zhu et al., 2016, Liu & Peter, 2019). The theoretical justification for this approximation, under the assumption of linearizable forward modeling and Gaussian model priors, is given in Section 2.3.1. Based on this, we replace the data misfit function in (3.39) with a generalized (or regularized) least-squares formulation (2.27)

$$\phi(\mathbf{m}) = \frac{1}{2}(\mathbf{d}-\mathbf{u}(\mathbf{m}))^T\mathbf{C}_d^{-1}(\mathbf{d}-\mathbf{u}(\mathbf{m})) + (\mathbf{m}-\mathbf{m}_0)^T\mathbf{C}_m^{-1}(\mathbf{m}-\mathbf{m}_0), \tag{3.44}$$

including a data covariance matrix $\mathbf{C}_d$, which weights the data misfit term, and a prior covariance $\mathbf{C}_m$, which weights the model misfit term $(\mathbf{m}-\mathbf{m}_0)$ where $\mathbf{m}_0$ is the prior mean model.
As presented in Section 2.3.1, the posterior covariance matrix $\mathbf{C}_M$ can be expressed as (Tarantola, 2005)

$$\mathbf{C}_M = (\mathbf{J}^T\mathbf{C}_d^{-1}\mathbf{J} + \mathbf{C}_m^{-1})^{-1} = (\mathbf{H}_a + \mathbf{C}_m^{-1})^{-1}, \tag{3.45}$$

or equivalently (Bui-Thanh et al., 2012, Zhu et al., 2016)

$$\mathbf{C}_M = \mathbf{C}_m^{1/2}(\mathbf{C}_m^{1/2}\mathbf{H}_a\mathbf{C}_m^{1/2} + I)^{-1}\mathbf{C}_m^{1/2}, \tag{3.46}$$

where $\mathbf{H}$ is the Gauss-Newton approximation of the data-misfit Hessian. The idea behind this uncertainty quantification strategy is to compute an approximation of the data-misfit Hessian, after FWI has converged, and evaluate the corresponding posterior covariance.
The Hessian matrix is often approximated in a low-rank form in order to derive an estimate of the uncertainty while sparing the high computation and storage cost of the full matrix. For example, Bui-Thanh et al. (2012), Zhu et al. (2016) apply the randomized SVD algorithm (r-SVD), proposed by Halko et al. (2011), to represent the prior-preconditioned Hessian ($\mathbf{C}_m^{1/2}\mathbf{H}_a\mathbf{C}_m^{1/2}$) in SVD form as

$$(\mathbf{C}_m^{1/2}\mathbf{H}_a\mathbf{C}_m^{1/2}) \approx \mathbf{V}_r\mathbf{\Lambda}_r\mathbf{V}_r^T, \tag{3.47}$$

where r is the estimated rank order, $\mathbf{\Lambda}_r$ is a diagonal matrix consisting of the eigenvalues and $\mathbf{V}_r$ is the matrix of associated eigenvectors. In particular, this r-SVD method estimates the singular values and vectors of a matrix through the application of this matrix to one set of random vectors. Details of this algorithm can be found in Appendix a. In this framework, an important step is therefore the computation of Hessian-vector products, for which various strategies can be adopted, such as the second-order adjoint method (Santosa & Symes, 1988, Fichtner & Trampert, 2011, Métivier et al., 2014). An alternative approach is to employ the L-BFGS algorithm (Alg. 1), which allows to calculate the product of a matrix with any vector. In practice, this algorithm is used in FWI to approximate the action of the inverse Hessian $\mathbf{H}^{-1}$ on the gradient vector without building or storing the full matrix (Section 2.2.1.1). As proposed by Liu et al. (2021), the r-SVD algorithm with a full-history L-BFGS scheme, in which the rank number is set equal to the number of FWI iterations and the gradient is replaced with a set of independent random vectors, can be employed to derive a low-rank representation of the BFGS-based inverse Hessian approximation.
By doing do, assuming $\mathbf{C}_m^{1/2}$ is absorbed into $\mathbf{J}$ in (3.46), an approximate posterior covariance can be derived as

$$\mathbf{C}_M \approx \mathbf{H}^{-1}\mathbf{C}_m, \tag{3.48}$$

which, with a low-rank approximation of $\mathbf{H}^{-1}$, becomes

$$\mathbf{C}_M \approx (\mathbf{V}_r\mathbf{\Lambda}_r\mathbf{V}_r^T)\mathbf{C}_m. \tag{3.49}$$

For the purpose of uncertainty analysis, 'standard deviation maps' are computed from the square root diagonals of the posterior covariance $\mathbf{C}_M$ (3.49) to provide a measure of local uncertainty (Tarantola, 2005, Liu et al., 2021). This approach is investigated through a numerical example in Section 3.2.3.

3.2.2 Homogenized full waveform inverse problem

This section introduces key concepts of the HFWI, from Capdeville & Métivier (2018), used in this study.
Based on the homogenization framework discussed in Section 3.1.3, for a given maximum frequency f_m, a solution to the homogenized FWI problem (HFWI) in the effective model space $\mathcal{M}^*$ can be found by solving the following minimization problem

$$\tilde{\mathbf{m}}^* = \underset{\mathbf{m}^* \in \mathcal{M}^*}{\arg\min}\, \phi^*(\mathbf{m}^*), \tag{3.50}$$

$$\phi^*(\mathbf{m}^*) = \sum_{r,s} \int_0^T (\mathbf{d}_s^{f_m}(\mathbf{x}_r, t) - \mathbf{u}_s^*(\mathbf{x}_r, t, \mathbf{m}^*))^2 dt, \tag{3.51}$$

where $\mathbf{u}_s^*$ is the effective displacement at the leading order of the asymptotic expansion (3.17), which is independent on the small-scale heterogeneities. In this case, because an analytical solution to the cell problem (3.30) does not exist in general, an explicit parametrization cannot be defined for $\mathcal{M}^*$. Therefore, to ensure that the final model $\tilde{\mathbf{m}}^*$ is in $\mathcal{M}^*$, a possible strategy consists in homogenizing the estimated model after each iteration, or only at convergence (Capdeville & Métivier, 2018, Hedjazian et al., 2021). The effective solution $\tilde{\mathbf{m}}^*$ is then spatially smooth (it does not contain wavelengths smaller than λ_0) and fully anisotropic.
Inverting seismic waveforms $\mathbf{d}_s^{f_m}$ for a macroscale effective medium $\tilde{\mathbf{m}}^*$ acts as a form of regularization because it reduces the size of the solution space by constraining the solution to the effective model space $\mathcal{M}^*$, yet ensuring the data fit. Furthermore, assuming that the global minimum of the HFWI has been found, a relation can be established between its solution and the true Earth model $\mathbf{m}_t$ (Capdeville & Métivier, 2018):

$$\tilde{\mathbf{m}}^* \approx \mathcal{H}(\mathbf{m}_t) \tag{3.52}$$

where $\mathcal{H}$ denotes the homogenization operator, built following the procedure described in Section 3.1.3.1, applied to estimate the effective properties of $\mathbf{m}_t$ for scale separation position ε_0. This relation (3.52) justifies the downscaling, or inverse homogenization, problem described in the following chapters.
An important point in HFWI is the choice of ε_0 to separate small and large scales. As seen above (Eq. 3.14), this choice is directly linked to the minimum wavelength λ_m, and consequently to the FWI resolution. Based on the desired accuracy, an ε_0 that lies between 1 and 0.25, which corresponds to a homogenization filter wavelength $\lambda_0 = \lambda_m$ and $\lambda_0 = \lambda_m/4$ respectively, could be selected.

3.2.3 Numerical example

In the following, we show a simple idealized 2-D elastic full-waveform inversion (FWI) example. In this experiment, noise is intentionally excluded and an acquisition geometry designed to provide good coverage of the region to be imaged

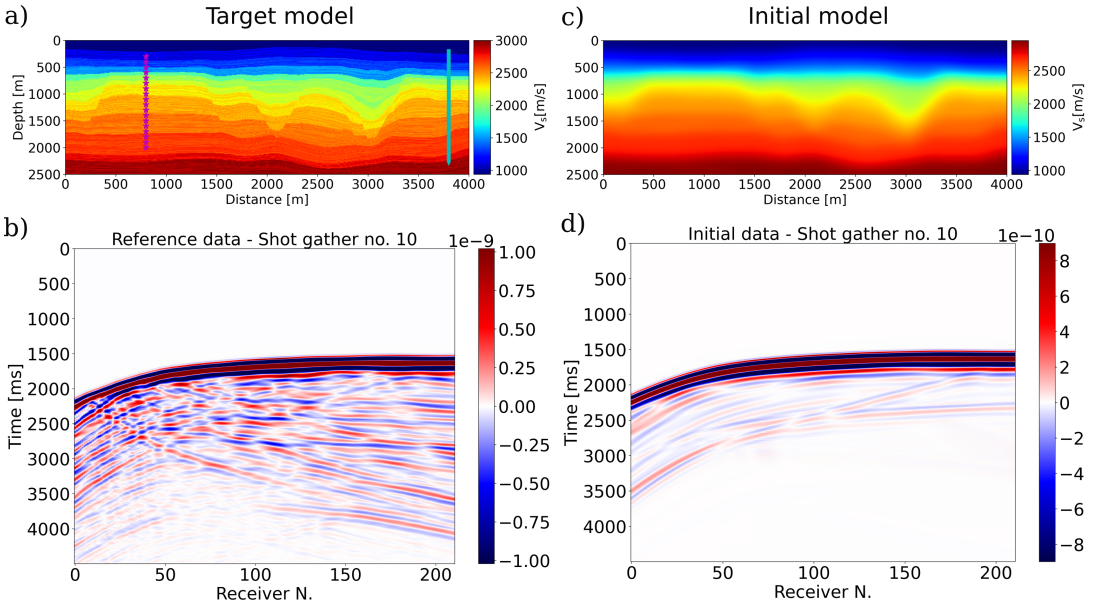

Figure 3.5: a) Reference fine-scale medium shown in terms of V_s. Sources (in purple) and receivers (in light blue) are located in two vertical boreholes at $x = 800$ m and $x = 3800$ m, respectively. b) Reference data: shot gather n.10 (x-component) obtained by simulating wavefield propagation in the reference fine-scale medium. c) Initial FWI V_s model obtained by applying Gaussian smoothing to the reference medium. d) Initial data: shot gather n.10 (x-component) obtained by simulating wavefield propagation in the initial smoothed medium.

is adopted. The objective is to assess the ability of FWI to image small-scale geological features, including faults and fault zones, under optimal conditions.

3.2.3.1 Forward modeling and inversion setup

The 2-D elastic FWI is carried out with the forward modeling/inversion program DENISE (Köhn, 2011), which solves the wave equation by utilization of a finite-difference (FD) time-stepping scheme (Alford et al., 1974, Virieux, 1986). Therefore, the wave equations are discretized in both time and space in accordance with the CFL stability criterion and the accuracy requirements of the FD scheme (Virieux, 1986). Specifically, P-SV wave simulations are performed using an eighth-order FD scheme. For this synthetic experiment, we focus on the inversion of the isotropic P and S-wave velocities in the elastic medium while keeping the density constant. In this simplified example, the same FD grid is used for both generating the data to be inverted and solving the inverse problem; no (white) noise was added to the observed data, and the same source wavelet was used for modeling the observed data and performing elastic FWI (thus intentionally committing an inverse crime). The data misfit function that has to be minimized by the inversion scheme is the least squares norm (Eq. 3.39). Ab-

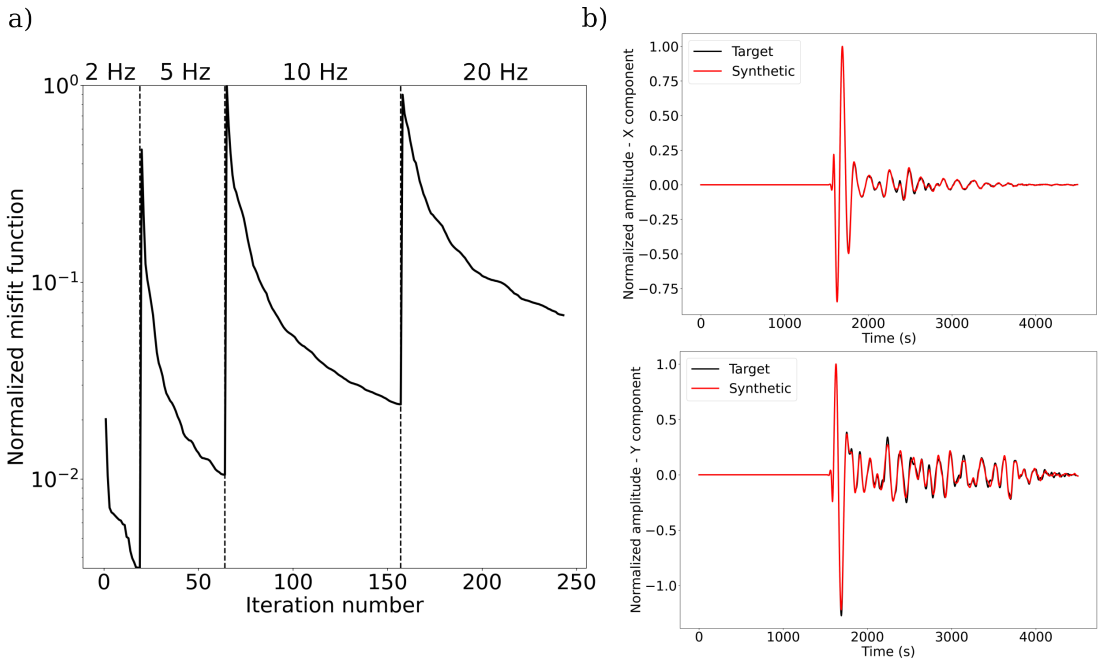

Figure 3.6: a) Normalized misfit function for the different frequency bands. b) Final data residual, i.e., difference between reference and predicted waveforms at the last iteration, for a single trace recorded by receiver n. 100 at the center of the borehole.

sorbing boundaries, based on perfectly matched layers (Festa & Vilotte, 2005), are used all around the domain.

The optimization is carried out with the L-BFGS algorithm (Alg. 1), thus the Hessian is not obtained explicitly and only vector products are computed. The gradient is preconditioned with a diagonal initial inverse Hessian approximation calculated by zero-lag correlation of the absolute value of the forward wavefield with an approximation of the receiver Greens function contribution (Plessix & Mulder, 2004).

The ill-posedness of the inverse problem is then mitigated by applying a multi-stage frequency approach: at each stage, a different low-pass filter $\mathcal{F}^{f_m}$ is applied to the source time function, $g^{f_m} = \mathcal{F}^{f_m}(g)$, and to the observed data, $\mathbf{d}^{f_m} = \mathcal{F}^{f_m}(\mathbf{d})$, such that the inversion starts with a lower maximum frequency and then progressively moves to higher frequencies, thus gradually introducing shorter length scales. In this experiment the inversion is separated into four stages, with maximum frequencies of 2, 5, 10 and 20 Hz. A stopping criterion, which is using the relative misfit change $\Delta\phi$ between the last two iterations, is applied to stop the inversion or to proceed to the next inversion stage ($\Delta\phi < 10^{-3}$).

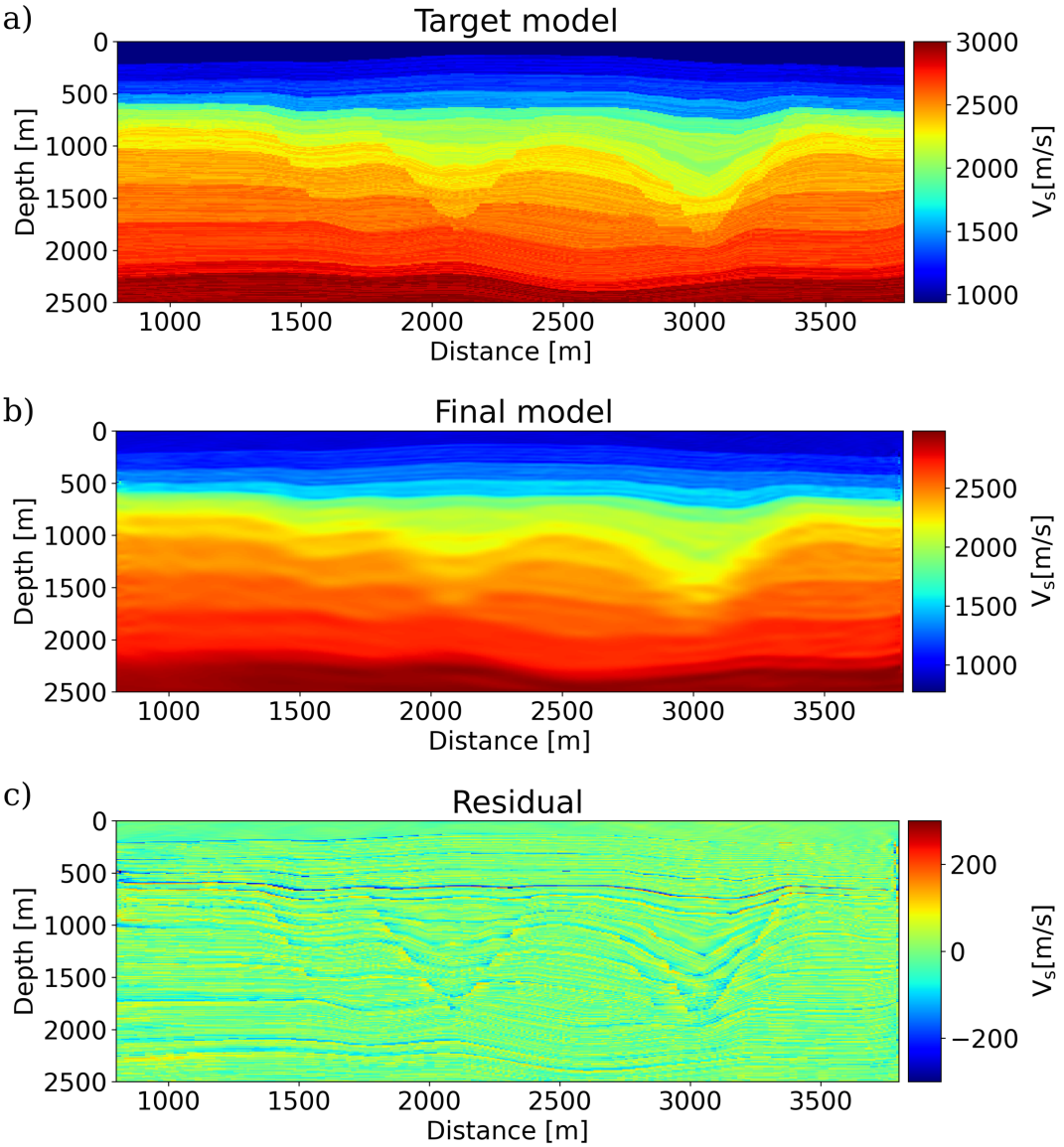

Figure 3.7: a) S-wave velocity of the reference model in the illuminated area (between the source and receiver boreholes) b) Result of the elastic FWI for the S-wave velocity. c) Residual model: difference between reference and predicted model.

3.2.3.2 Velocity model and acquisition geometry

The "observed" seismic data are simulated in the fine-layered reference model (Figures 3.2a and 3.5a), which voluntarily contains heterogeneities of scales smaller than the wavefield minimum wavelength. The initial velocity model (Fig. 3.5c) is derived by a 2-D Gaussian smoothing of the reference model. The FD grid consists of 250 × 400 samples, with a grid spacing of 10 m × 10 m resulting in a model 4 km × 2.5 km large.

The acquisition geometry is arranged in a cross-hole configuration (Fig. 3.5a),

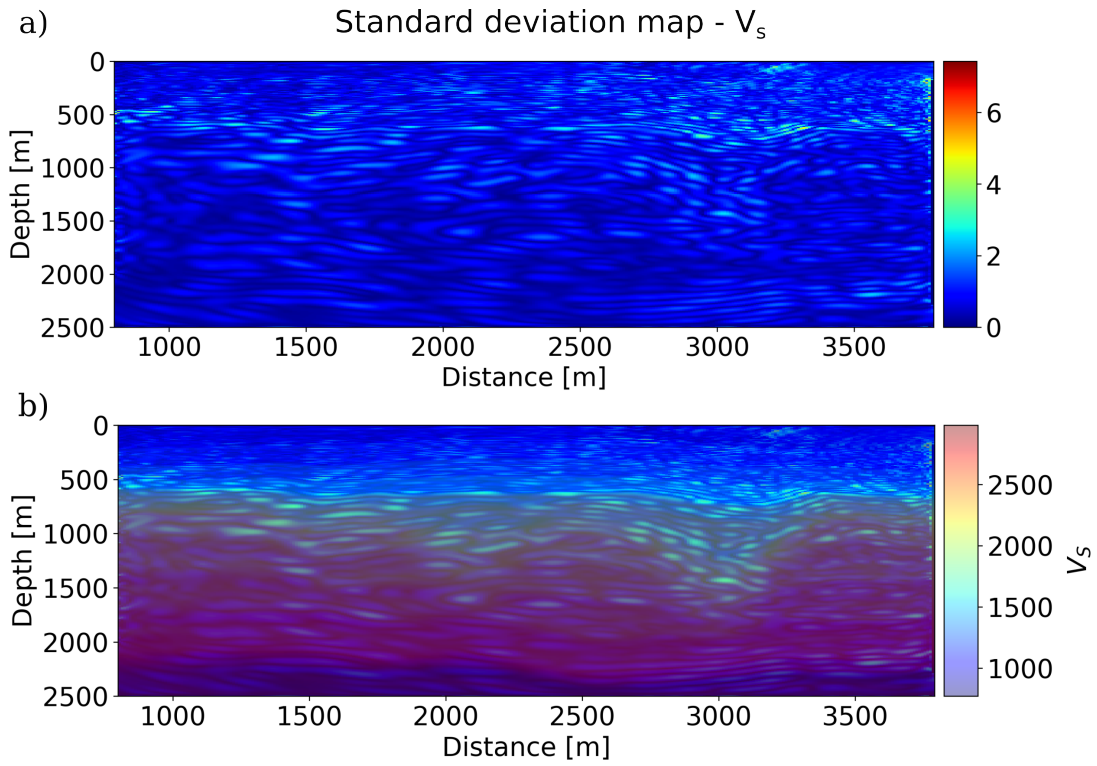

Figure 3.8: a) Standard deviation map computed from the diagonal entries of the approximated inverse Hessian matrix for V_S. Larger standard deviation values indicate higher uncertainty areas. Values range between [0, 7] because gradient values have been normalized by the maximum value. b) Standard deviation map superimposed on the FWI inverted V_S in Figure 3.7b.

including 18 explosive sources, spaced by 100 m, and 210 two-component geophones placed every 10 m. Such source-receiver configuration provides good illumination, even if rarely occur in real field scenarios. Nevertheless data coverage is still sparse as the distance between each each source is equal to 100 m. The sources are point-force sources using a Ricker wavelet with a maximum frequency $f_m = 20Hz$, which is consistent with the maximum frequencies generally achievable in conventional industrial surveys, such that the estimated minimum wavelength is $\lambda_m \sim 40$m.

3.2.3.3 Inversion results

Starting from the initial model (Fig. 3.5b), the convergence is reached after a total of 243 iterations involving all the frequency stages (Fig. 3.6) and resulting in the final model shown for the S-wave velocity in Figure 3.7b within the illuminated area (between the two boreholes). Figure 3.6a shows the normalized misfit function decreasing fast throughout the iterations for the different frequency bands. The multi-stage approach (i.e., starting with low-frequency contents and

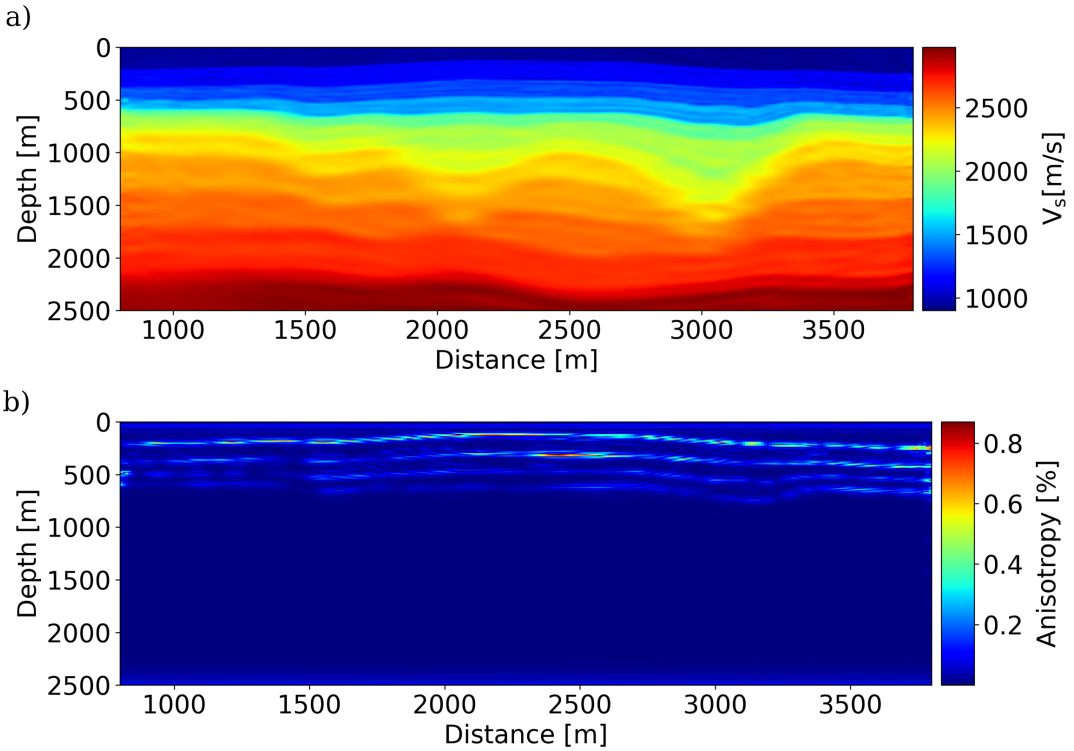

Figure 3.9: Effective S-wave velocity (a) and anisotropy (b) of the homogenized FWI solution model with $\varepsilon_0 = 0.5$, derived respectively with equations (3.36) and (3.37).

gradually introducing high-frequency content) mitigates cycle skipping enough to make the predicted data match the observed data. Indeed, the good fit between the target and predicted data shown in Figure 3.6b indicates that the FWI final model can adequately explain the target data. The inverted model reconstructs the major interfaces, which are hardly visible in the initial smooth model, and reveals the presence of the fault system (Fig. 3.7b). However, the exact extent and displacement of the individual faults remain uncertain even in this ideal setting. This is particularly significant for the graben structure situated around $x = 3\,000$ m, where the geology is more complex and several alternative fault configurations can be assumed. This result could be potentially improved by increasing the number of sources and the maximum frequency content of the wavelet.
The residual in Figure 3.7c, computed as the difference between the target and the inverted model, reveals the small-scale heterogeneities not recovered by the inversion.

3.2.3.4 Standard deviation maps

Following an approach similar to that of Liu et al. (2021), we apply the r-SVD algorithm to retrieve the BFGS inverse-Hessian approximation (referred to as **B**

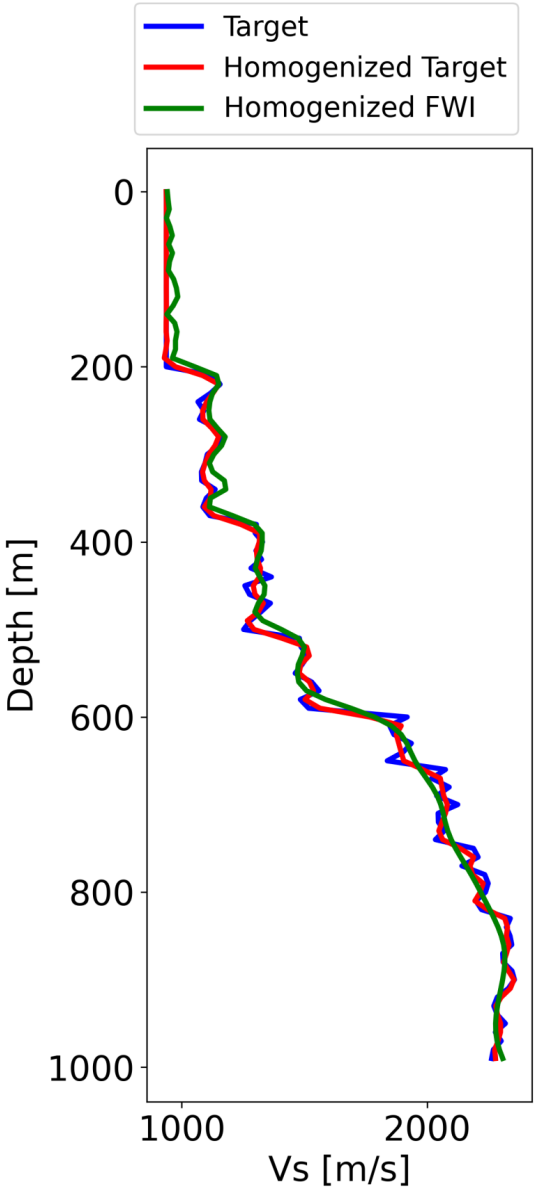

Figure 3.10: Zoom on the S-wave velocity profile for the upper 1 000 m, extracted at $x = 1\,000$ m, of the reference fine-scale elastic medium, its effective medium and the homogenized FWI solution.

in Algorithm 1) in a low-rank form. For this purpose, during the optimization process, at the last frequency stage (when all the frequency components up to $f_m = 20$ Hz are considered in the inversion), all the L-BFGS vector pairs $\mathbf{s}_k, \mathbf{y}_k$ $(k = 1, \ldots, K)$ (2.16) and the initial diagonal inverse Hessian approximation are stored in memory. Here K indicates the number of FWI iterations at the last stage (equal to 85 for this experiment). Once FWI has reached convergence, by choosing the rank number equal to K (which is generally smaller than the dimension of the problem: $K << M$), the approximate inverse Hessian $\mathbf{H}_{\mathrm{K}+1}^{-1}$ can

be probed with a set of independent random vectors $\mathbf{X}$ ($K \times M$), by replacing the gradient $\mathbf{g}_k$ at each k iteration of the recursion scheme in Algorithm 1 with $\mathbf{x}_k \in \mathbf{X}$. This probing, stated in the first line of Algorithm 4 in Appendix a, only requires in practice vector-vector products, resulting in high computational efficiency. Then, the subsequent steps of the r-SVD algorithm (Alg. 4) compute the eigenvectors $\mathbf{V}_\mathrm{K}$ and eigenvalues $\mathbf{\Lambda}_\mathrm{K}$ of $\mathbf{H}^{-1}_{\mathrm{K}+1}$, from which the matrix columns and/or the diagonal elements can be easily extracted. For simplicity, we use an identity matrix as prior covariance matrix $\mathbf{C}_m$ in equation (3.49), thus assuming no knowledge on the prior model uncertainty. Therefore, assuming that the main features of $\mathbf{C}_M$ are reflected by $\mathbf{H}^{-1}$, a measure of the uncertainty is directly obtained from the square root of the diagonal entries of the low-rank approximation $\mathbf{H}^{-1}$, computed from $\mathbf{V}_\mathrm{K}$ and $\mathbf{\Lambda}_\mathrm{K}$. In this sense, this approximation, shown as standard deviation map in Figure 3.8 for the S-wave velocity, offers a spatially qualitative indication of uncertainty rather than a quantitative measure. Figure 3.8 shows higher value of uncertainty in the vicinity of fault structures and at the bending of the layers in the upper part of the model. Hence, uncertainty grows locally in regions where there are stronger discontinuities in the medium. These can be interpreted as zones of insufficient coverage due to the bending of the wave in correspondence of these structures. The validity of these measures of uncertainty and how these could be used in our downscaling workflow is discussed in Section 3.3.

3.2.3.5 Homogenized FWI results

Following the procedure and parametrization described in Section 3.1.4, we compute the effective elastic properties of the FWI inverted medium with the non-periodic homogenization operator. In this case, the FWI iterative scheme is applied classically, without using homogenization, and only the solution is homogenized. As for the original fine-scale medium in Figure 3.2a, for which we have verified the accuracy of the homogenized version obtained using $\varepsilon_0 = 0.5$ (Fig. 3.4), we consider results at the resolution limit (i.e., $\lambda_0 = \lambda_m/2$ according to equation (3.14)). The effective V_s^* and anisotropy of the FWI solution are shown in Figure 3.9. By comparing this result with the effective reference medium (Fig 3.2b,c), we can observe that apparent anisotropy recovered by the homogenized FWI solution is evident only in the upper portion of the model, in correspondence of interfaces with strong velocity contrasts. This means that FWI inversion with isotropic parametrization, once homogenized using $\varepsilon_0 = 0.5$, is not able to provide information about the effective anisotropy induced by fault structures. Using an anisotropic parametrization for the FWI, as well as adopting a spatially denser discretization might help in recovering a stronger effective anisotropy. Higher values of ε_0 (i.e., using larger wavelength equivalent media) could also provide larger effective anisotropy.

Figure 3.10a shows for a single V_s profile the comparison between the reference

fine-scale medium (Fig. 3.2a), the homogenized reference medium (Fig. 3.2b), and the homogenized FWI inverted medium (Fig. 3.7b). The homogenized reference medium represents the best solution that FWI could recover with limited frequency-band assuming perfect coverage, noiseless data, and accurate forward physics assumptions. The homogenized FWI velocity profile, which closely follows the FWI, shows lower resolution, and some velocity artifacts (such as in the first 200 m) not related to true subsurface heterogeneities. This mismatch between the homogenized FWI solution and the effective properties of the reference fine-scale earth model, i.e., the meaning of $\approx$ in equation (3.52) represents the data uncertainty in our downscaling inverse problem.

3.3 Summary and discussion

The purpose of this chapter is to introduce the two main ingredients of the downscaling workflow (Fig. 1.5) covered in the following chapters. Hence, we did not present new findings here, but rather key concepts and parameter choices that will be used later.

The chapter starts with an introduction of the theory behind the elastic homogenization, starting from the more intuitive 1-D periodic case and extending to non-periodic material in the 2-D case, which serves as the forward operator in the downscaling inversion. A key aspect of the non-periodic homogenization is that the spatial variations of the original medium do not have to be neither periodic nor smaller than the minimum wavelength λ_m. The medium can contain all the scales, and a scaling parameter ε_0 is introduced to separate small and large scales. In his current state, the theory relies on a λ_m which is constant all over the domain. In some realistic geological scenarios where λ_m significantly varies with location, this restricts its benefit or can affect the quality of the solution if λ_m is not chosen properly (see Capdeville & Marigo (2013), appendix B).
We applied the non-periodic homogenization to a subsurface model to show the effect of the homogenization operator on the elastic properties for various ε_0. Then, with a wavefield comparison, we showed that small-scale heterogeneities filtered out with a scaling parameter $\varepsilon_0 = 0.5$ have little effect on the computed displacement. We used the approach based on FFT from Capdeville et al. (2015b) for the numerical resolution of the elasto-static problem in equation (3.30). This method represents a fast alternative to the finite element method since it is based on a simple regular grid and relies on highly optimized algorithms. However, when dealing with discontinuous media, a dense grid is required to approximate the discontinuities properly. The FFT solver is used in all the tests presented hereinafter to derive the effective properties as having a cheap forward model represents a great advantage in the probabilistic framework of the downscaling inversion.

In the second part we described important features of the FWI inverse problem and introduced results from Capdeville & Métivier (2018) about HFWI, whose solution represents the data in the Bayesian framework of the downscaling. Indeed, the downscaling inverse problem is justified by the assumption that, assuming noise-free data and perfect data coverage, HFWI provides the effective properties of the true Earth model, provided that the global minimum of the HFWI problem has been reached. With a simple experiment, deterministic FWI is carried out to infer the elastic properties of a reference subsurface model containing small-scale heterogeneities. The same homogenization operator is then applied to the reference model and to the medium recovered by FWI to compare the effective elastic properties of the two media.
In general, assuming there are no errors in the homogenization procedure, the mismatch between the two effective media comes from errors in the FWI solution due to noise in the data, lack of data coverage, convergence towards a local minimum in the waveform inversion, incorrect physical assumptions in the forward model, etc. How to compute uncertainty associated to the FWI solution obtained through a local optimization approach is a widespread subject of research in seismic tomography.
In the example proposed in this chapter, for the purpose of quantifying uncertainty in the FWI solution, a strategy that builds on the idea of using the dominant eigenvalues and eigenvectors of the inverse Hessian to approximate the posterior covariance in the vicinity of the optimum FWI result is investigated. Specifically, we derive the diagonal of a low-rank inverse Hessian approximation from L-BFGS-based FWI relying on the single-pass randomized SVD algorithm (r-SVD), from which a measure of the standard deviation of each parameter is computed. In this example, since data are noise free, the spatial standard deviation is primarily related to the acquisition geometry and wavefield illumination. The advantage of this approach is that, since it only requires vector products, it is efficient and suitable for large-dimensional problems. However, the validity of the approximation relies on the set of underlying assumptions on which the method is built. In primis, this approximation is based on the assumption of linearized forward modeling operator and Gaussian model priors, meaning that its validity decreases as the nonlinearity of the inverse problem increases. Then, it implicitly assumes that the BFGS-based inverse Hessian is an accurate approximation of the Hessian and that the r-SVD algorithm computes an accurate estimate of the singular values as in SVD. Furthermore, knowledge of the prior model covariance (which is usually poorly known) is necessary for the absolute values of posterior uncertainty to be meaningful. As a consequence, if any or all of these assumptions are not satisfied, the approximate estimate of the covariance derived may not be representative of the uncertainty of the inversion results.
Estimating uncertainty is even more difficult in the case of the HFWI solution. Indeed, the use of the homogenization step in the FWI scheme, which is a nonlinear step, makes it impossible to work directly on the Hessian. For this reason,

an alternative strategy to map FWI uncertainty in the effective model space and approximate a data covariance matrix in HFWI must be determined. A possible strategy would be, for instance, using the estimated low-rank approximation to the posterior covariance to draw random samples from the Gaussian posterior distribution (as in Zhu et al., 2016, Liu & Peter, 2019). This process results in an ensemble of model perturbations to the final model, that once homogenized, could be used to calculate a covariance matrix on the homogenized ensemble. This ensemble covariance gives an approximate measure of uncertainty, whose reliability also depends on the size of the ensemble. A straightforward but more costly strategy is that employed by Hedjazian et al. (2021), who compute uncertainties performing multiple HFWI for multiple noise realizations on the synthetic data. As such, this approximation considers errors associated to the noise in waveforms and mapped to the HFWI solution through the inversion. Both methods can be considered as practical approximations, but a more accurate estimate of data uncertainties in HFWI should be addressed in future work.

Chapter 4

Downscaling inversion for fault-related uncertainty quantification

4.1 Introduction

Downscaling generally refers to the process of inferring small-scale information from large-scale variables. Within the scope of our work, this process is used to reconstruct a high-resolution model of the subsurface, given an estimation of its smooth effective elastic properties. To compute effective properties, we employ the elastic non-periodic homogenization, introduced in the previous chapter. Therefore, we refer to the downscaling as an inverse homogenization problem, where the homogenization plays the role of the forward operator.
In this chapter, downscaling is applied to an elastic subsurface model with the goal of reducing uncertainty on a fault structure. In particular, from the smooth effective elastic properties of the HFWI solution, representing the observed data, the inverse problem aims to infer fault-related model parameters describing the fault geometry and displacement. A Bayesian formalism is adopted to address the ill-posed nature of this inverse problem and to incorporate prior geological knowledge about faults in the study area. In this case, elastic properties are treated as fixed inputs and are not inverted during the process; downscaling of these properties will be addressed in the next chapter.
One of the main challenges in inverse problems trying to resolve structural uncertainty is to specify a model parametrization that ensures realizations are geologically realistic. In this work, to guarantee that model realizations are structurally compatible, fault displacement is generated using a kinematic modeling approach (Jessell, 1981, Wellmann et al., 2016, Godefroy et al., 2018). As compared to physics-based modeling, kinematic modeling is purely geometric and does not account for stress transfer, making it more computationally efficient. This represents a significant advantage for the integration into the probabilistic framework

of the downscaling.
In the following, we describe in detail the model parametrization and the Bayesian framework of the inverse problem, and then we present a representative application case. Specifically, two distinct experiments investigate the benefits of using the full effective elastic tensor — which encodes anisotropy information — as observations, rather than relying solely on its first component. In the first case, observations derive from a synthetic HFWI medium, while in the second example, these are computed by homogenizing a synthetic FWI result, as described in Section 3.2.3.

4.2 Geological Parametrization

In this section we describe the methodology employed to simulate fault displacement and the fault-related model parameters representing the unknowns in the downscaling inverse problem.
As a first step, the domain is discretized with a regular Cartesian grid $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^2 \mid i = 1, \ldots, N\}$, where the elastic properties $\mathbf{m} = (V_p, V_s, \rho)$ are defined at each grid point. Faulting is then simulated by applying a kinematic fault operator to the set of points $\mathbf{X}$ (as in Godefroy et al., 2018). The kinematic model is characterized by a discontinuous displacement field defined within an elliptic shape centered on the fault (Fig. 4.1a). The location of the fault in the domain and the fault dip are given by the coordinates of the center (x_c, y_c) and the rotation angle ϕ of the ellipse, respectively. The spatial extent of the fault displacement is provided by the lengths of the ellipse semi-axis, controlling the distance between the center and the tip of the fault along the fault line (a) and in the orthogonal direction (b). The displacement magnitude inside the elliptical fault space is controlled by a maximum amount of slip $D_{\max}$ at the fault center, and two parametric profiles, which can be adapted to handle deformations with different complexities. In this work, these profiles are parametric Hermite splines (Fig. 4.1b) which define:

- the displacement along the fault line, which decreases away from the fault center to zero on the tip-line according to a bell-shaped profile;

- the displacement orthogonal to the fault, which is discontinuous at the center to account for the discontinuity of the displacement field across the fault.

Once applied, the kinematic operator yields a mapping of each point's location $\mathbf{x}$ to a post-faulting location $\mathbf{x}^{\boldsymbol{\vartheta}}$, where $\boldsymbol{\vartheta}$ denotes the fault-related parameters $(x_c, y_c, a, b, \phi, D_{\max})$ describing the fault model (or structural model). As a final step, the post-faulting elastic values $\mathbf{m}^{\boldsymbol{\vartheta}}$ are interpolated bilinearly onto the regular grid $\mathbf{X}$ before forward modeling. In the downscaling inverse problem, the

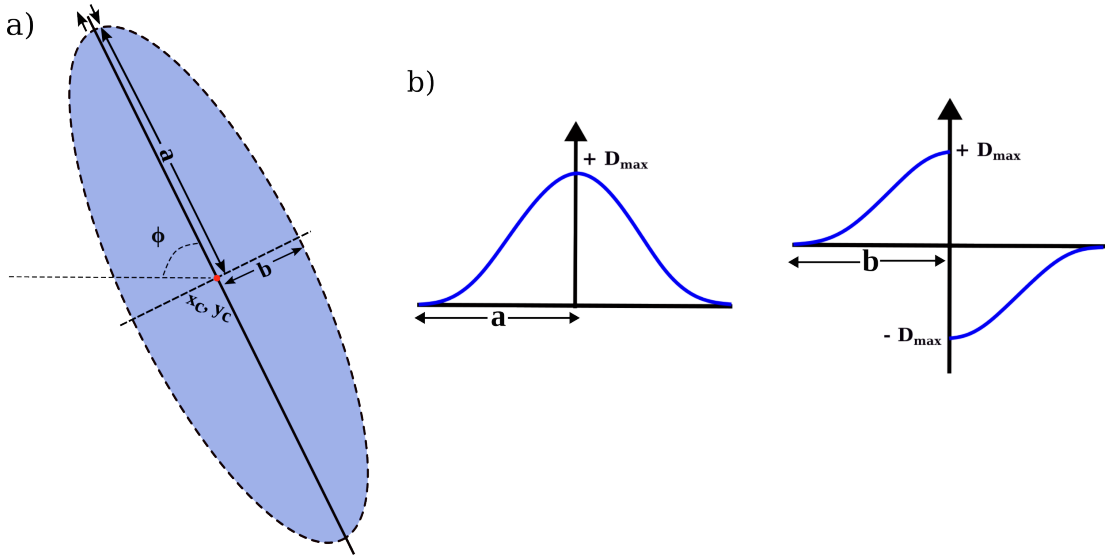

Figure 4.1: a) Elliptical fault space. The major-axis of the ellipse defines the fault line, while the minor-axis defines the thickness of the displacement field. b) Parametric profiles defining the displacement inside the elliptical fault space along the two directions. The displacement decreases smoothly from the center (where the displacement is $D_{\max}$) to the boundary of the ellipse (where the displacement vanishes).

parameters $\boldsymbol{\vartheta}$ will be treated as unknowns and sampled to retrieve an ensemble of possible fault structures configurations, consistent with the HFWI image.

4.3 Bayesian inference with MCMC

Once the model parameters $\boldsymbol{\vartheta}$ are defined, in order to set-up the Bayesian framework of the downscaling inversion, it is essential to specify the a *priori* information, provided as a probability distribution $\pi(\boldsymbol{\vartheta})$. These probabilities are typically derived from a prior knowledge about the geological setting. In this work, uniform prior distributions are chosen, reflecting the assumption of little prior knowledge on the fault structure.
The forward modeling operator is the non-periodic elastic homogenization $\mathcal{H}$, described in Section 3.1.3.1. Therefore, for each proposed fault model $\boldsymbol{\vartheta}$, the corresponding elastic medium $\mathbf{m}^{\boldsymbol{\vartheta}}$ is homogenized to obtain its smooth and anisotropic equivalent medium

$$\mathbf{c}^* = \mathcal{H}(\mathbf{m}^{\boldsymbol{\vartheta}}), \tag{4.1}$$

where $\mathbf{c}^*$ corresponds to the six independent parameters $(\mathbf{c}^*_{11}, \mathbf{c}^*_{12}, \mathbf{c}^*_{13}, \mathbf{c}^*_{22}, \mathbf{c}^*_{23}, \mathbf{c}^*_{33})$, expressed here in Voigt notation, of the symmetric effective elastic tensor. The observed effective medium $\mathbf{c}^*_{\mathbf{obs}}$ refers to the effective properties of the HFWI

deterministic solution (3.50)

$$\mathbf{c}^*_{obs} = \tilde{\mathbf{m}}^* = \mathcal{H}(\mathbf{m}_t) + \boldsymbol{\epsilon}, \tag{4.2}$$

with $\boldsymbol{\epsilon}$ expressing the mismatch between the HFWI solution $\tilde{\mathbf{m}}^*$ and the effective properties of the true earth $\mathbf{m}_t$ (3.52). As discussed in Section 3.3, these errors can derive from both theoretical errors on the homogenization operator $\mathcal{H}$ and, primarily, from errors in the FWI solution owing to noise in the data, lack of data coverage, convergence towards a local minimum in the waveform inversion, etc. Assuming the distribution of errors $\boldsymbol{\epsilon}$ to be multivariate-Gaussian distributed with zero mean and covariance matrix $\mathbf{C}_d$, the likelihood is expressed as

$$\pi(\mathbf{c}^*_{\mathbf{obs}} \mid \boldsymbol{\vartheta}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{c}^*_{\mathbf{obs}} - \mathcal{H}(\mathbf{m}^{\boldsymbol{\vartheta}})\right)^T \mathbf{C}_d^{-1}\left(\mathbf{c}^*_{\mathbf{obs}} - \mathcal{H}(\mathbf{m}^{\boldsymbol{\vartheta}})\right)\right). \tag{4.3}$$

Given the likelihood and prior distributions, Bayesian inference aims to determine the posterior distribution $\pi(\boldsymbol{\vartheta} \mid \mathbf{c}^*_{\mathbf{obs}})$ of the parameter set $\boldsymbol{\vartheta}$ according to the well-known Bayes equation:

$$\pi(\boldsymbol{\vartheta} \mid \mathbf{c}^*_{\mathbf{obs}}) \propto \pi(\mathbf{c}^*_{\mathbf{obs}} \mid \boldsymbol{\vartheta})\pi(\boldsymbol{\vartheta}), \tag{4.4}$$

which is then used to quantify the uncertainty of the parameters of interest $\boldsymbol{\vartheta}$. A popular method for characterizing $\pi(\boldsymbol{\vartheta} \mid \mathbf{c}^*_{\mathbf{obs}})$ is the Markov Chain Monte Carlo (MCMC) sampling algorithm (see Section 2.3.2.3). This stochastic process starts from an initial guess of the model parameters $\boldsymbol{\vartheta}$, randomly chosen from the uniform prior $\pi(\boldsymbol{\vartheta})$, and samples the model space, at each iteration k, according to the proposal distribution $q(\boldsymbol{\vartheta}_{k+1} \mid \boldsymbol{\vartheta}_k)$, with the probability of the proposal state $\boldsymbol{\vartheta}_{k+1}$ being accepted depending on $\boldsymbol{\vartheta}_k$ only, according to the Metropolis-Hastings criterion (2.38). For this study, we make use of a Gaussian proposal $q(\boldsymbol{\vartheta}_{k+1} \mid \boldsymbol{\vartheta}_k) \sim \mathcal{N}(\boldsymbol{\vartheta}_k, \sigma^2\mathbf{I})$, and of an adaptive Metropolis sampling scheme, which combines the traditional Metropolis-Hastings with the capacity to tune the step size σ during burn-in until a desired acceptance rate is achieved. After the burn-in phase, the sampling process eventually reaches its stationary regime (i.e., when the distribution of the chain's states stabilizes) and begins sampling from the posterior $\pi(\boldsymbol{\vartheta} \mid \mathbf{c}^*_{\mathbf{obs}})$.

4.4 Examples

In the following examples, inverse homogenization is applied to downscale a localized area in the reference fine-layered subsurface model (shown within the black rectangle in Figure 4.2a), in which the elastic properties $\mathbf{m}_t = (V_p, V_s, \rho)$ are defined. In particular, the objective is to quantify uncertainty on the fault structure in this area, which is assumed to be of interest for supporting decision making in this application scenario.

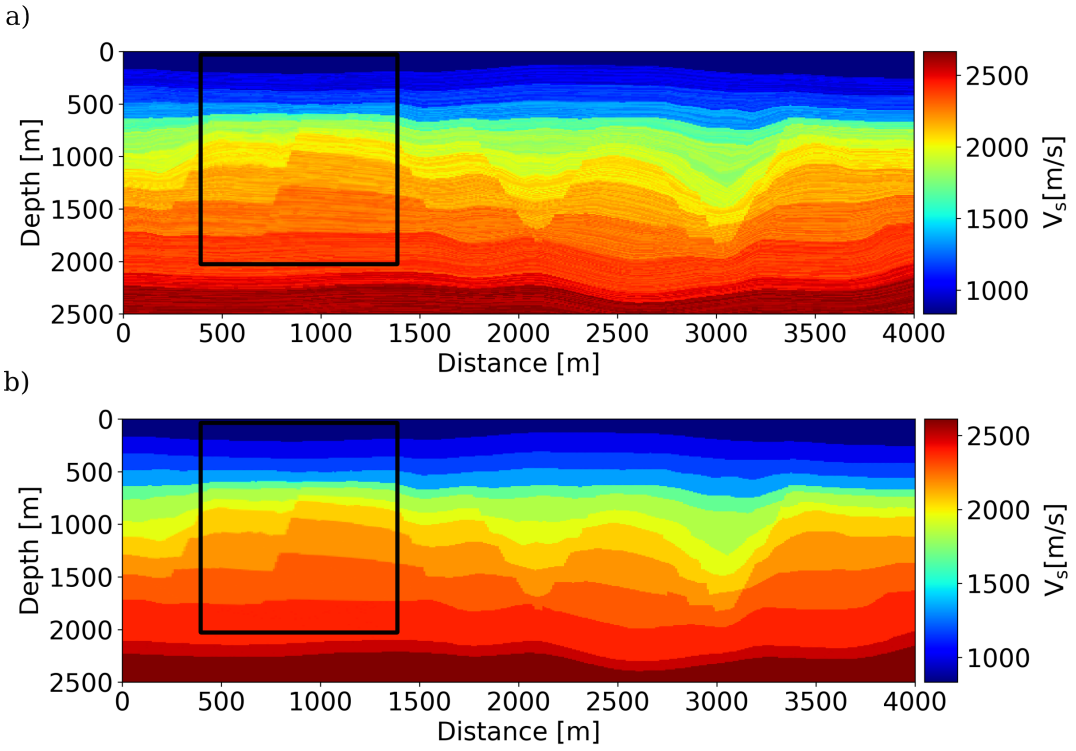

Figure 4.2: a) Reference fine-layered elastic medium. b) Associated macro-layered medium. The elastic medium is illustrated here in terms of S-wave velocity V_s. The black rectangle delineates the region of the medium where downscaling is carried out to infer the target fault structure.

The medium is discretized using a regular grid of 250 × 400 nodes, with a uniform grid spacing of 10 m × 10 m resulting in a model 4 km × 2.5 km large. In this case, the S-wave velocity V_s and density ρ are calculated from the P-wave velocity V_p using the ratio $V_p/V_s = 1.8$ and Gardner's relation $\rho = 0.31\ V_p^{0.25}$ (Gardner et al., 1974). We assume that the macro-scale elastic properties $\mathbf{m}$ (i.e., the properties of the macro-layered earth model (Fig. 4.2b) associated with the reference fine-scale version) are known. This macro-layered model captures the major geological interfaces and, traditionally, can be built by defining discontinuities from structural interpretations and by assigning velocity to layers based on tomography, or using average interval velocities estimated from well logs or vertical seismic profiles (e.g., Irakarama et al., 2019).

To the macro-scale elastic properties in the pre-faulting state (Fig. 4.3b), the kinematic fault operator, described in Section 4.2 and defined by the model parameters $\boldsymbol{\vartheta}$, is applied to simulate fault displacement. Therefore, in the MCMC sampling scheme of the downscaling inversion, for each proposed $\boldsymbol{\vartheta}$, the faulted macro-scale elastic properties $\mathbf{m}^{\boldsymbol{\vartheta}}$ are homogenized to derive their effective properties $\mathbf{c}^*$ in which the sharp discontinuities between layers and at the fault location

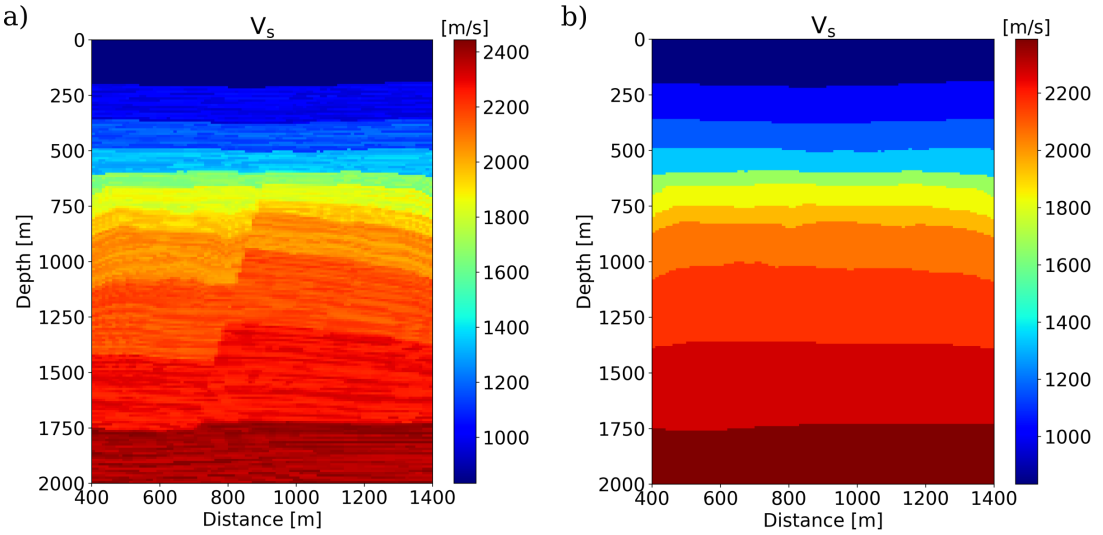

Figure 4.3: a) Region of the reference fine-layered medium, inside the black rectangle in Figure 4.2, including the target fault structure for which the downscaling inverse problem aims to reduce uncertainty. b) Initial macro-layered medium whose elastic properties (P-wave velocity, S-wave velocity and density) are treated as known in the inversion. The elastic medium is illustrated here in terms of S-wave velocity V_s.

are smoothed. Homogenization is carried out with the FFH approach as described in Section 3.1.4. In the following, we compare results of two downscaling applications involving distinct types of observational data $\mathbf{c}^*_{obs}$.

4.4.1 Application with synthetic HFWI observations

In this first application, we approximate the observations $\mathbf{c}^*_{obs}$ as

$$\mathbf{c}^*_{obs} = \mathcal{H}(\mathbf{m}_t) + \boldsymbol{\eta}, \tag{4.5}$$

where the homogenization operator $\mathcal{H}$ with a scale-separation spatial wavelength $\lambda_0 = 200$ m is applied to the reference medium $\mathbf{m}_t$ (Fig. 4.3a), and spatially correlated Gaussian noise $\boldsymbol{\eta} \sim N(0, \boldsymbol{\Sigma})$ is added to the resulting effective medium. In these data, no spatial variations of wavelength smaller than λ_0 are kept in the smooth effective properties. A large λ_0 was selected to investigate whether the methodology is capable of accurately reconstructing the fault structure despite the strong smoothing applied to the data. The noise $\boldsymbol{\eta}$ is modeled using a Gaussian covariance function, with variance set to 4% of the mean value of the effective elastic properties, and characteristic length-scales $l_x = 400$ m and $l_y = 200$ m, in order to reproduce possibly large-scale structured correlations that may be present in an effective elastic medium obtained from HFWI. The resulting

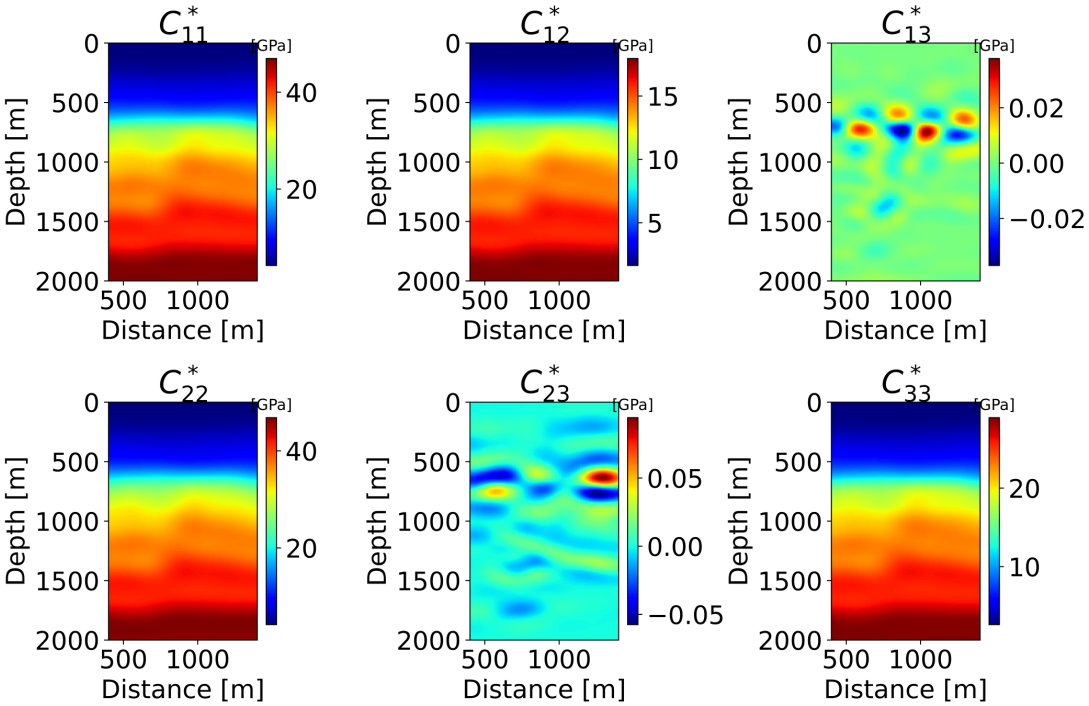

Figure 4.4: The six independent components of the effective elastic tensor obtained by homogenizing the reference model (Fig.4.3a) with a scale-separation spatial wavelength $\lambda_0 = 200$ m.

effective elastic tensors, with and without noise, are shown in Figures 4.4 and 4.5 respectively.

As can be observed, the off-diagonal terms $\mathbf{c}_{13}^*$ and $\mathbf{c}_{23}^*$ of the effective elastic tensor, which contain essential information about the effective medium anisotropy, are non-zero where significant velocity variations occur both vertically and laterally. For calculating the likelihood (4.3), we use an approximate data covariance matrix $\mathbf{C}_d$ instead of using the full exact covariance of the noise $\boldsymbol{\eta}$, which would represent an idealized scenario. In particular, $\mathbf{C}_d$ considers errors in the six elastic coefficients of the data vector $\mathbf{c}_{\mathbf{obs}}^*$ are correlated, but does not account for spatial correlation between data points, meaning that errors are considered independent for each location.

To verify the convergence, we run 4 MCMC parallel chains starting from different initial fault models $\boldsymbol{\vartheta}$ (Fig. 4.8), randomly sampled from the uniform prior distributions. The convergence plots of the 4 runs are shown in Figure 4.6, which displays the negative log-likelihood, equivalent of the misfit function in Bayesian inference, and the acceptance rate of all the realizations. To assess convergence we calculate the Gelman–Rubin statistic $\hat{R}$ (Gelman & Rubin, 1992), which provides an estimate of how well the Markov chains have mixed and whether they

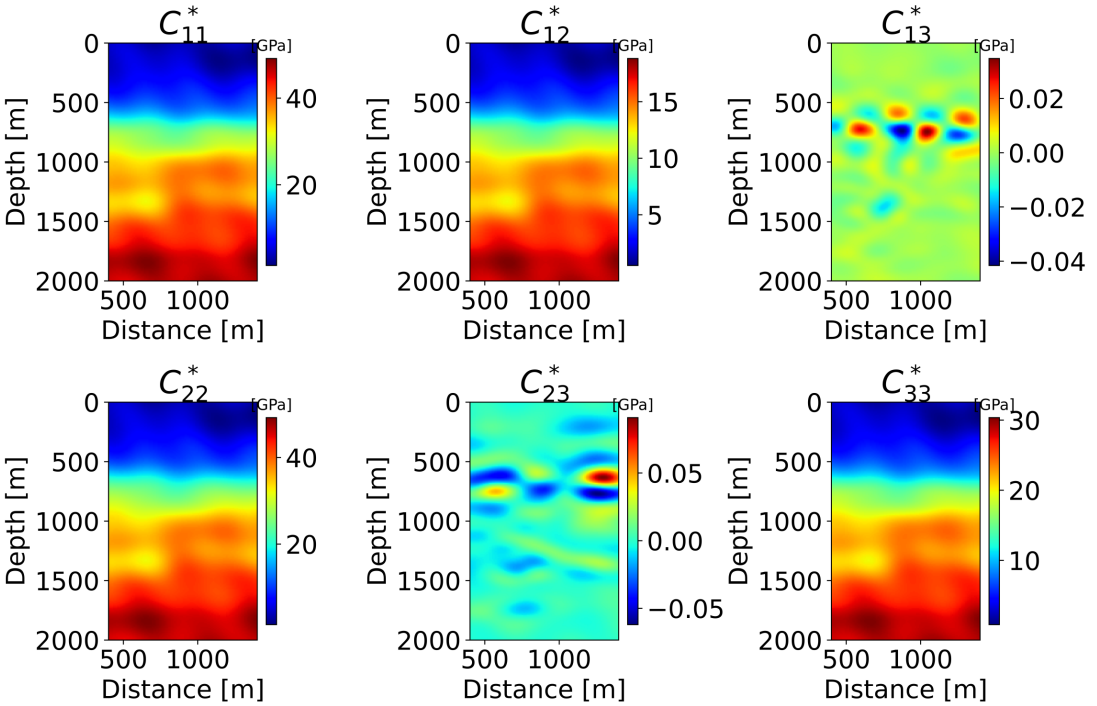

Figure 4.5: The six components of the effective elastic tensor $\mathbf{c}^*_{\mathbf{obs}}$ representing the observation in the synthetic application. These are obtained by homogenizing the reference model (Fig. 4.3a) with a scale-separation spatial wavelength $\lambda_0 = 200$ m (Fig. 4.4) and by adding spatially correlated Gaussian noise $\boldsymbol{\eta}$.

have converged to the target distribution (typically when $\hat{R} < 1.1$). For all the fault-related parameters $\boldsymbol{\vartheta}$, the estimated $\hat{R}$ values are close to 1 after 10 000 iterations, indicating convergence. The MCMC run completed in 10 hours on a cluster using a single node with four MPI tasks, where the MCMC chains were executed in parallel. We discard the first 4 000 realizations, considered part of the burn-in period, and compute the posterior probabilities based on the remaining realizations in the stationary regime, for which the acceptance rate is around 5 %. The minimum and maximum values of the uniform prior distributions and the marginal posterior distributions of the fault parameters $\boldsymbol{\vartheta}$ are shown in Figure 4.7. Despite using large uniform priors, the reference value for all the fault-related parameters $\boldsymbol{\vartheta}$ (marked by the dashed line on the plots) lies within the posterior distribution. This indicates that data, reflected through the likelihood function, have sufficiently informed the posterior, even in presence of noise $\boldsymbol{\eta}$. In particular, the posterior densities suggest how the likelihood function tightly constrains the posterior for all fault parameters, leading to sharp and concentrated distributions and low standard deviation in the posterior samples. A larger standard deviation can be observed for the semi-minor axis b of the ellipse, defining the lateral spatial extent of the fault-induced deformation. This is reasonable because

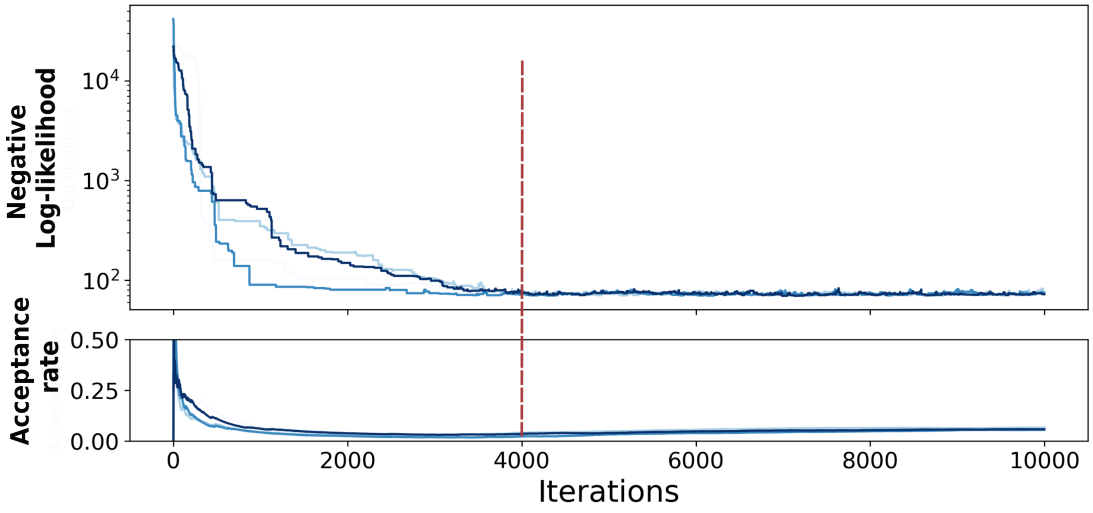

Figure 4.6: Convergence diagnostics of the downscaling inversion with synthetic HFWI observations. Top graph: Negative log-likelihood functions of 4 parallel MCMC chains for number of iterations. The burn-in phase is considered to finish after 4 000 iterations (indicated by the red dotted line) when the functions reach a plateau and start sampling the posterior. Bottom graph: Acceptance rate of 4 parallel MCMC chains for number of iterations. As the number of iterations increases, the acceptance rates stabilize around 5 %.

the fault displacement is assumed to be maximum at the center of the fault and decrease towards the tips of the elliptic fault zone. Hence, a small variation of b does not produce a significant change in the smoothed effective properties, thus data are less informative. The mean model shown in Figure 4.8 confirms how, in this synthetic data case study, the target fault structure (Fig. 4.2a) has been successfully recovered. The good accuracy of the results stems in particular from the fact that the key features in the $\mathbf{c}^*_{13}$ and $\mathbf{c}^*_{23}$ components of the elastic tensor (Fig. 4.5), revealing anisotropic information about the fault structure, remain intact despite the noise. Because these informative features are not distorted, the inversion can correctly infer the model parameters, leading to data predictions that, for these two components, closely resemble the observations. As a result, the fault-related parameters are well constrained.

4.4.2 Application with homogenized FWI observations

Differently from the previous application, we now consider observations $\mathbf{c}^*_{obs}$ as

$$\mathbf{c}^*_{obs} = \mathcal{H}(\tilde{\mathbf{m}}_{\mathrm{FWI}}), \tag{4.6}$$

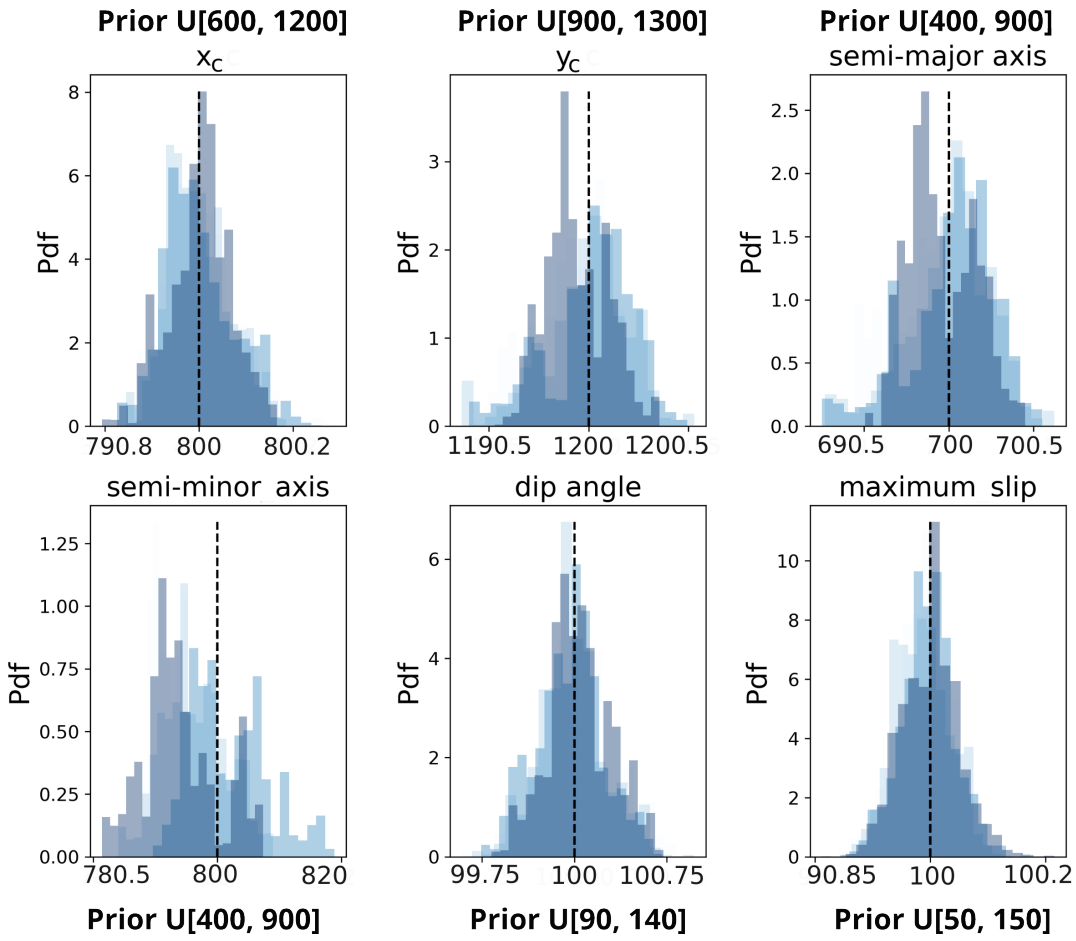

Figure 4.7: Inversion results of the downscaling inversion with synthetic HFWI observations: posterior densities of the fault-related parameters $\boldsymbol{\vartheta}$ from 4 MCMC chains (shown in different shades of blue) after burn-in. Black dotted lines indicate target values from the reference fault structure in Figure 4.3a. The minimum and maximum values of the uniform priors for each model parameter are indicated in bold.

where the homogenization operator $\mathcal{H}$ is applied to the inverted FWI medium $\tilde{\mathbf{m}}_{\text{FWI}}$ (3.38). As presented in Section 3.2.3, we carried out a synthetic 2-D isotropic FWI experiment where the fine-layered elastic medium in Figure 4.2a represents the target solution. The initial FWI medium is computed with a 2-D Gaussian smoothing of the reference medium (Fig. 3.5c). We adopt exactly the same FD discretization, (two-borehole) acquisition geometry, source characteristics and inversion scheme described in Section 3.2.3. The resulting medium $\tilde{\mathbf{m}}_{\text{FWI}}$, for the V_p and V_s elastic properties, is shown in Figure 4.9, with the black rectangle highlighting the area where the target fault is located. As can be observed, given the adopted survey and inversion design parameters in FWI, even under ideal conditions with noise-free data, the inverted elastic properties do not allow for a reliable interpretation of the fault geometry and displacement. The down-

scaling inversion is therefore carried out as a second stage inversion to quantify fault-related uncertainty. For this purpose, $\tilde{\mathbf{m}}_{\mathrm{FWI}}$ is homogenized with the FFH approach using a regularly spaced grid of 1 m $\times$ 1 m to derive the anisotropic effective elastic properties $\mathbf{c}^*_{obs}$.
Given the wavefield's estimated minimum wavelength $\lambda_m = V_{min}/f_{max} \sim 40$ m, we set $\varepsilon_0 = 1$ to homogenize the FWI solution, which corresponds to a scale separation $\lambda_0 = \lambda_m$ (3.14). Differently from the previous application, in this case we

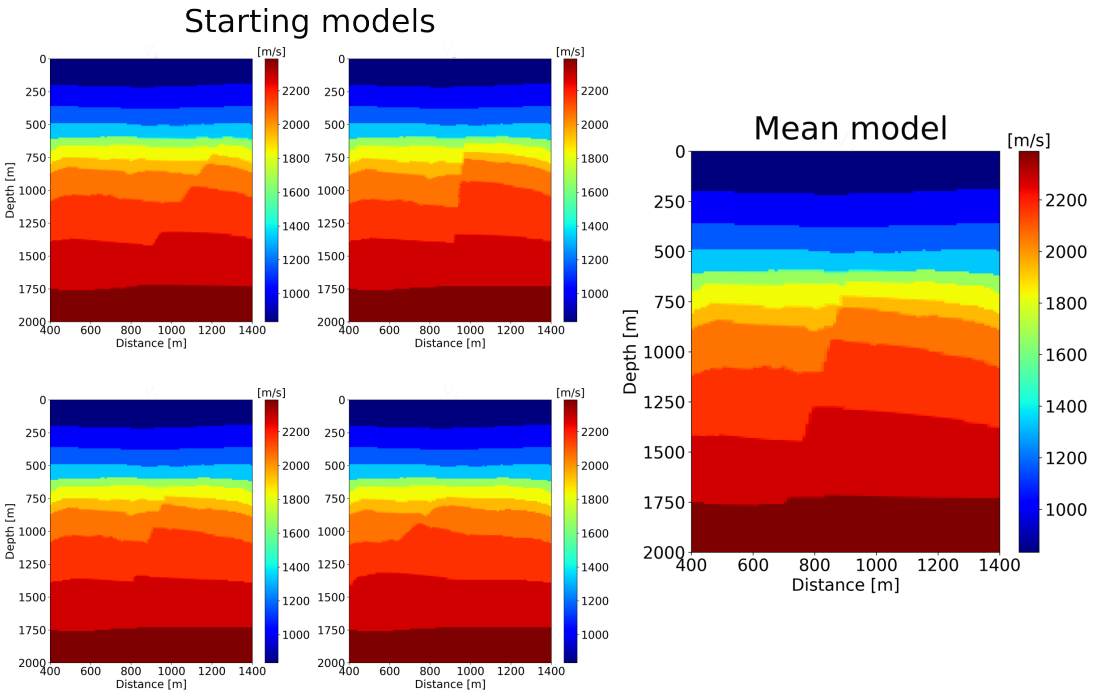

Figure 4.8: Fault velocity models of the first downscaling application with synthetic HFWI observations: the 4 starting models, in which the fault structure is generated using fault-related variables randomly sampled from uniform prior distributions, and the mean model derived from post-burn-in posterior samples across the 4 MCMC chains.

restrict the observations to the first component of the elastic tensor $\mathbf{c}^*_{obs} = \mathbf{c}^*_{11}$, shown in Figure 4.10, instead of the full tensor. For the likelihood estimation (4.3), we use a diagonal data covariance matrix $\mathbf{C}_d$ whose variance is derived as a fixed percentage of the observation variance.
Again, we run 4 MCMC parallel chains starting from different initial fault models $\boldsymbol{\vartheta}$, randomly sampled from the uniform prior distributions. The negative log-likelihood and the acceptance rate of all the realizations for the 4 runs are shown in Figure 4.11. Convergence was reached after 5 000 iterations, with the Gelman–Rubin statistic below 1.1 for all fault-related parameters. We discard the first 1 000 realizations and compute the posterior probabilities based on the remaining realizations for which the acceptance rate stabilizes around 47 %. As

compared to the first application, where each proposed model was better constrained by information coming from observations and therefore more easily rejected, now the proposals are more often compatible with the $\mathbf{c}_{11}^*$ observations, giving a much higher acceptance rate. The minimum and maximum values of the uniform prior distributions and the posterior marginal distributions of fault parameters $\boldsymbol{\vartheta}$ are shown in Figure 4.12. For the location (x_c, y_c), dip angle (ϕ) and displacement $(D_{\max})$ of the fault structure, posterior distributions are concentrated around, or near, the reference parameter values; whereas the semi-major axis a and the semi-minor axis b of the ellipse are shifted with respect to the reference, indicating that the spatial extent of the fault-induced deformation, i.e., the size of the elliptic fault space, is less constrained by the $\mathbf{c}_{11}^*$ data. As expected, all the posterior distributions are wider (and therefore yield higher uncertainty) than in the previous example since the smooth effective $\mathbf{c}_{11}^*$ is less informative than the full elastic tensor of the synthetic data-case. This is also confirmed by the mean and standard deviation models, shown in Figure 4.13, obtained with the posterior samples (after burn-in) of the 4 chains.

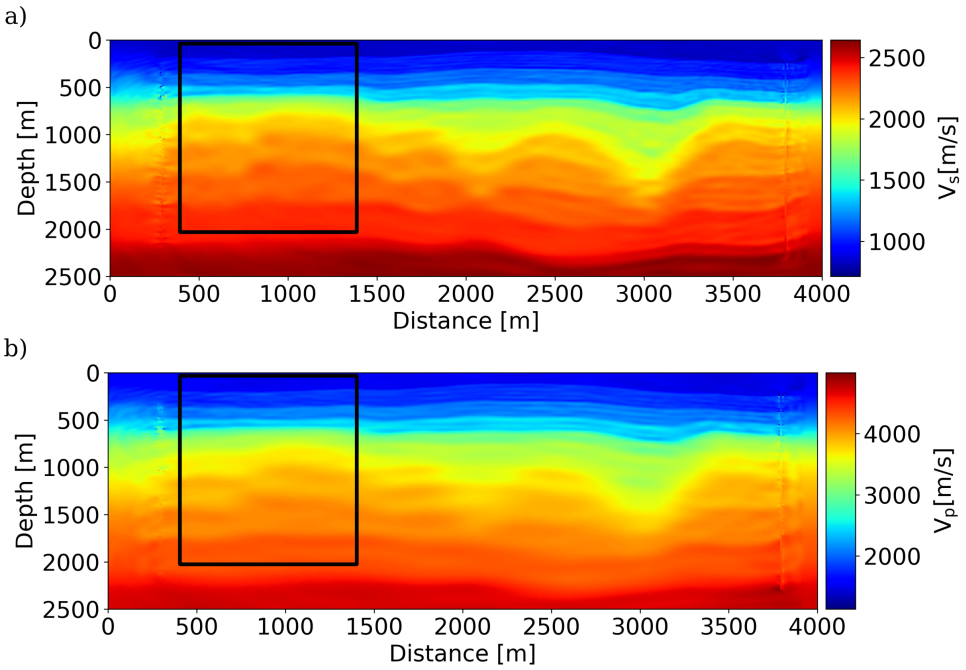

Figure 4.9: a) Result of the elastic FWI for the S-wave velocity. b) Result of the elastic FWI for the P-wave velocity. The black rectangle delineates the region of the medium where the downscaling inverse problem is performed to characterize the target fault structure. The sources and receivers used in the synthetic FWI experiment are positioned in two vertical boreholes located at $x = 300$ m and $x = 3800$ m, respectively.

4.5 Discussion

In this chapter, we illustrated the downscaling inverse problem in its formulation as an inverse homogenization problem. For the purposes of the application presented here, downscaling aims at resolving uncertainty on fault structures from a smooth effective medium, derived via HFWI, that gathers the information present in the seismic data. Specifically, the inversion is carried out within a stochastic

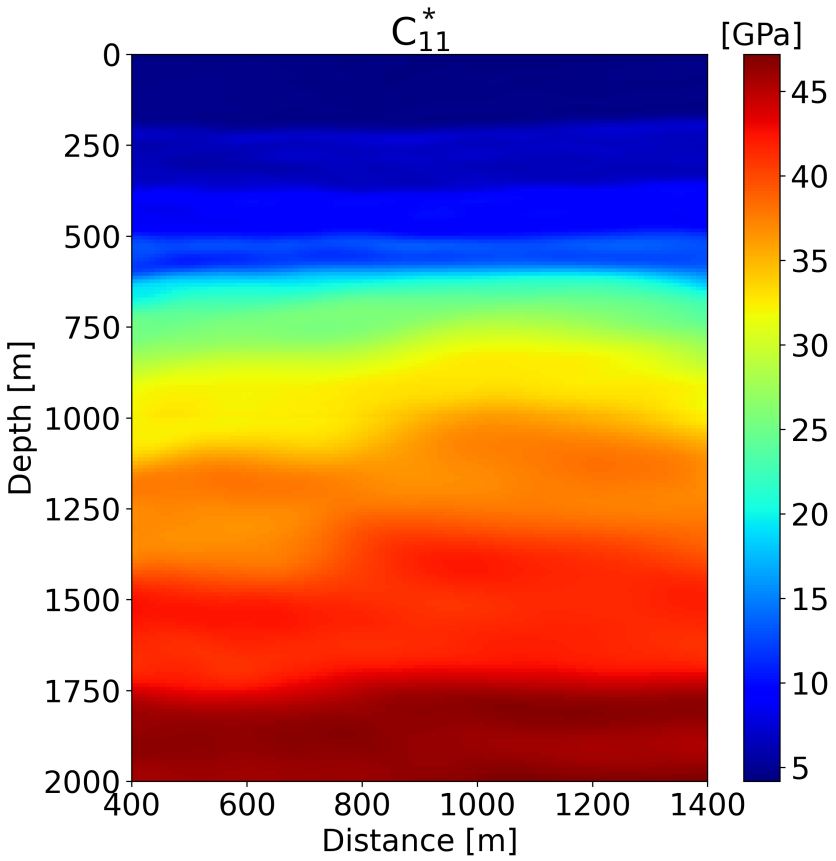

Figure 4.10: First component of the effective elastic tensor $\mathbf{c}^*_{\mathbf{obs}}$, derived by homogenizing, with a scale-separation spatial wavelength $\lambda_0 = 40$ m, the elastic properties of the FWI solution in Figure 4.9. This represents the observations in the second application of the downscaling inversion.

Bayesian framework with a fault-related parametrization in order to directly incorporate geological knowledge into the inversion process.
We presented an application in which, assuming perfect knowledge on a macro-layered subsurface model, the objective was to quantify uncertainty on a target fault structure in the reference fine-layered subsurface. In the first example, the HFWI medium is synthetically generated by homogenizing the elastic properties of the reference subsurface model, to which noise has been introduced in order to make the inversion more challenging. Moreover, uninformative prior distributions

of fault parameters are chosen to represent realistic conditions. In this case, the approach is highly effective in recovering the exact fault structure, owing to the informative content of the data, particularly the effective anisotropy information provided by the off-diagonal components $\mathbf{c}^*_{13}$ and $\mathbf{c}^*_{23}$ of the full elastic tensor. In a second experiment, we tested the same algorithm with less constraining observed data, reflecting a more realistic scenario. In this example, the smooth elastic properties used to inform the downscaling inversion are given by the first component of the effective elastic tensor $\mathbf{c}^*_{11}$ obtained by homogenizing the elastic properties of the subsurface model resulting from FWI, rather than the reference ones. Even with larger uncertainty, the mean model reflects the target fault structure, demonstrating the effectiveness of the methodology.
Although the method has shown good performance in resolving structural uncertainty, we are aware of some assumptions and simplifications made in the above synthetic tests, as well as of the additional challenges to address the case of real data applications. These aspects are further discussed below.

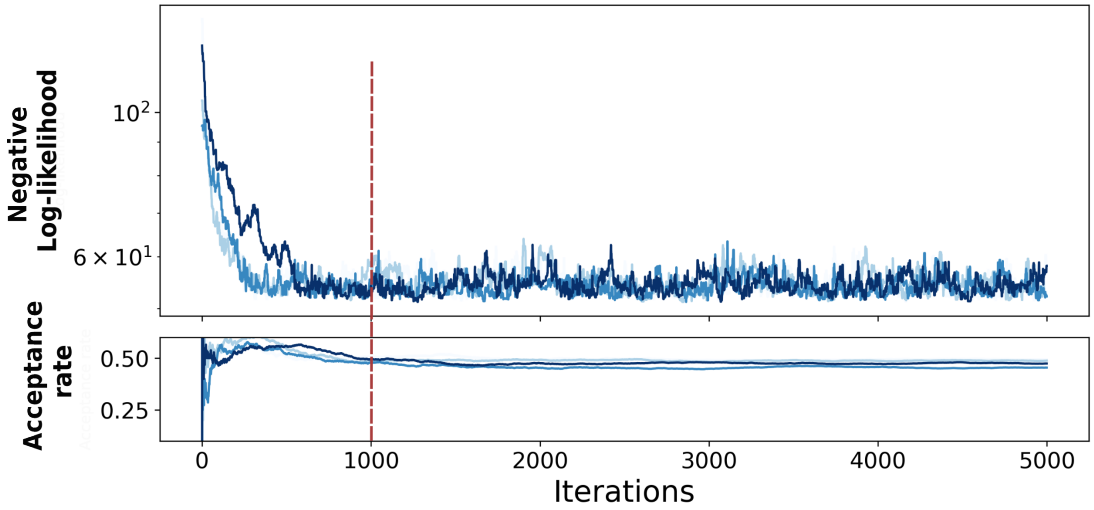

Figure 4.11: Convergence diagnostics of the downscaling inversion with homogenized FWI observations. Top graph: Negative log-likelihood functions of 4 parallel MCMC chains for number of iterations. The burn-in phase is considered to finish after 1 000 iterations (indicated by the red dotted line) when the functions reach a plateau and start sampling the posterior. Bottom graph: Acceptance rate of 4 parallel MCMC chains for number of iterations. As the number of iterations increases, the acceptance rates stabilize around 47 %.

4.5.1 Geological complexity

To parametrize the fault model, we adopt a kinematic operator where fault deformation is modeled in a symmetric ellipse centered on the fault line. For simplicity,

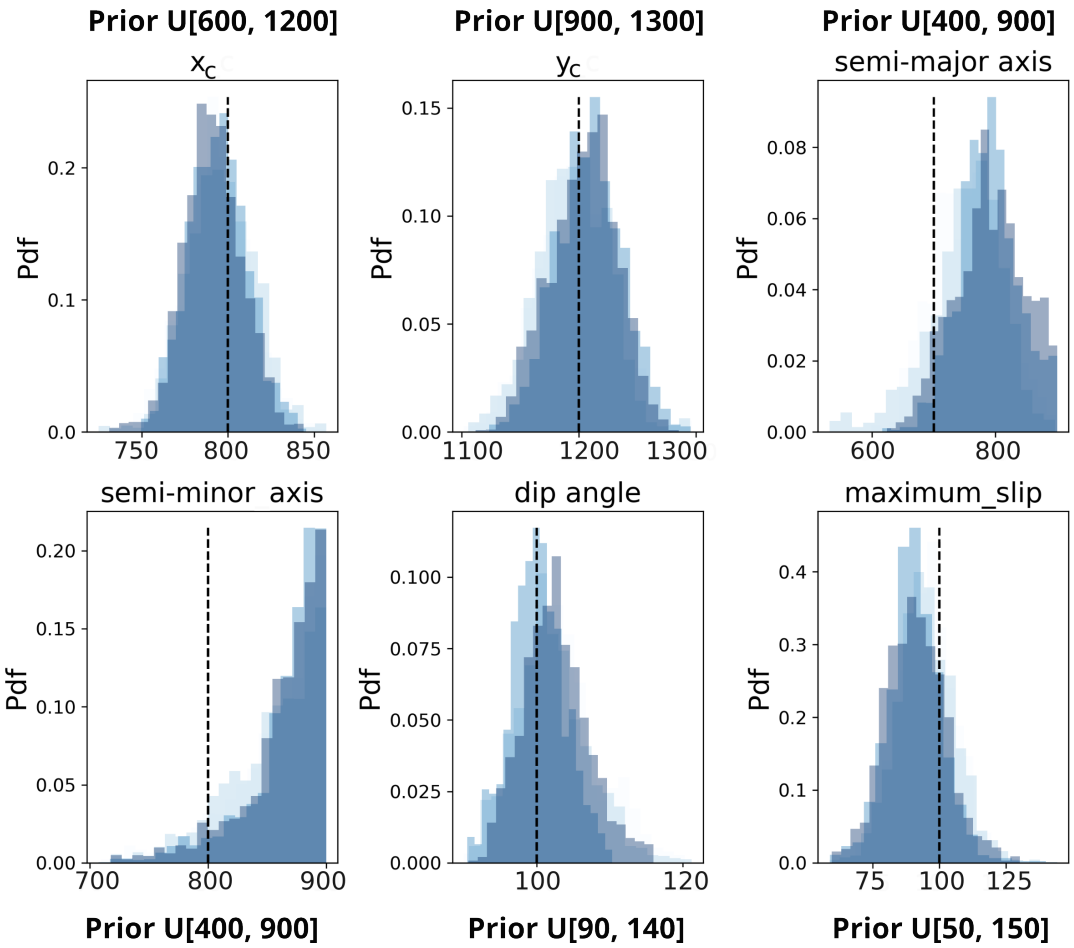

Figure 4.12: Inversion results of the downscaling inversion with homogenized FWI observations: posterior densities of the fault-related parameters $\boldsymbol{\vartheta}$ from 4 MCMC chains (shown in different shades of blue) after burn-in. Black dotted lines indicate target values from the reference fault structure in Figure 4.3a. The minimum and maximum values of the uniform priors for each model parameter are indicated in bold.

we tested the approach with a normal fault inducing a displacement symmetric and parallel to the fault line, and we used the same kinematic fault operator to simulate fault displacement in both observations and predictions. While convenient for testing, this setup is a simplification relative to real-world cases, where faults are often nonlinear and more complex. In this case, by defining asymmetric displacement profiles, the fault model can be modified to account for asymmetric displacement and more complex fault geometries, such as listric faults (as in Godefroy et al., 2018). In addition, to generalize the findings presented here to a more realistic context, the downscaling inversion could be tested in the case of several faults where fault displacements are applied sequentially and the relationship between faults is controlled by the sequence order (Cherpeau et al., 2010). In such case, the number of faults in the model and the order in which they

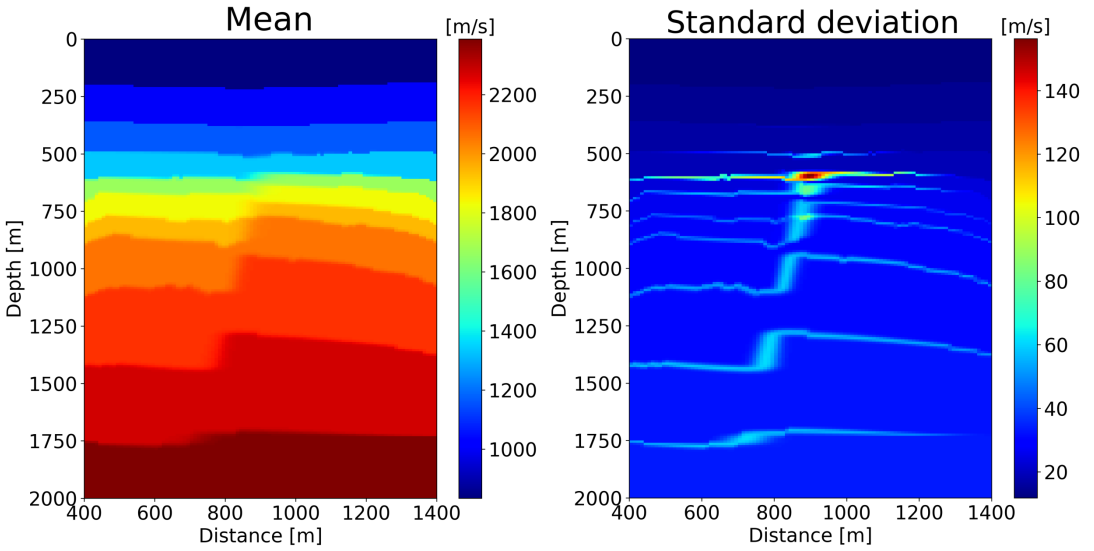

Figure 4.13: Posterior statistics of the downscaling inversion with homogenized FWI observations: mean and standard deviation model in which the fault structure is derived from posterior samples post-burn-in of 4 MCMC chains.

truncate each other, for instance, can be considered uncertain and more sophisticated MCMC algorithm, such as rj-MCMC (Sambridge et al., 2006), or more informative priors could be adopted to address these problems (e.g., Cherpeau & Caumon, 2015, Aydin & Caers, 2017, Taty Moukati et al., 2025, Holden et al., 2003).
A further simplification we made was to assume as known the exact interface depths and elastic properties of the macro-layered subsurface model. Although interval velocity models are typically built from prior geological knowledge, often using well logs and seismic interpretations, these are inherently uncertain because based on incomplete or imperfect data and modeling approximations. As a further extension, downscaling could be used to infer the interface geometries and layer velocities of the macro-layered model, which FWI recovers only in a smoothed form. Additionally, the proposed approach could be also employed to infer the fine-layered velocity, thus enabling quantification of both structural and stratigraphic uncertainty. This opens the way for using geostatistical techniques to model spatial variations in the velocity field and to incorporate hard data, such as well logs, to constrain velocity realizations. This topic will be explored in the next chapter.

4.5.2 HFWI as observations

A central aspect of the described methodology is the choice of the cutoff wavelength λ_0, setting the separation between the small and the large scales, for com-

puting the effective elastic properties of the FWI and of the proposed subsurface models. In the second experiment presented in this chapter, the homogenization operator is applied to the FWI solution with a scale separation factor $\varepsilon_0 = 1$, corresponding to a wavelength $\lambda_0 = \lambda_m$. In general, for shorter λ_0, geological heterogeneities at smaller scales would be preserved, and the solution would be better constrained by the observations. However, assuming that FWI can only reach a resolution down to half the smallest propagated wavelength $\lambda_{min}/2$, a $\lambda_0 \geq \lambda_{min}/2$ is a reasonable choice, because smaller scales are poorly resolved by the seismic data, as seen in the numerical example shown in Section 3.1.4.
In these synthetic tests, the data were sufficiently informative for the MCMC algorithm to recover the fault structure. In particular, from results of the first experiment, it is clear that the fault-induced effective anisotropy present in the synthetic HFWI solution plays a key role in constraining the inversion. Examples presented in Capdeville & Métivier (2018), Hedjazian et al. (2021) show that performing HFWI with an anisotropic parametrization and a finer spatial discretization helps, at the expenses of an increased computational cost, in recovering the correct effective anisotropy, which could then be used to inform the downscaling inversion about the small-scale heterogeneities observed in the waveforms.
Another point to consider is that, in both applications described above, we compute an approximate data covariance matrix considering that errors in the six independent elastic coefficients obtained by homogenization are correlated with each other. However, we do not account for spatial correlation between data points. This approximation was made for simplicity although spatial correlation induced by the homogenization (variables are spatially correlated on a length scale λ_0) should be accounted for by using a full covariance matrix.
For real data application, downscaling becomes more challenging since the HFWI solution may be affected by noisy data and lack of data coverage. These sources of uncertainty must therefore be properly accounted for in the data covariance matrix to avoid biased inversion estimates. Possible strategies for estimating and defining uncertainty in HFWI have been discussed in Section 3.3.

4.5.3 Computational aspects

In this work, we performed the downscaling inversion using MCMC sampling based on an adaptive Metropolis-Hastings algorithm, which automatically adjusts the proposal step size during the burn-in phase according to the acceptance rate. In general, MCMC sampling is known to become computationally intensive for high-dimensional parameter spaces due to the curse of dimensionality (Curtis & Lomax, 2001). In our examples, given the limited number of fault-related parameters to be inferred, the sampler efficiently reaches the stationary distribution, eliminating the need for more advanced strategies to accelerate convergence, such as parallel tempering, or more sophisticated MCMC algorithms (see Section

2.3.2.3). In particular, we observe that, using only the $\mathbf{c}_{11}^*$ component of the effective elastic tensor as data (as for the second experiment), leads to a higher acceptance rate and faster convergence with respect to using the full tensor case. Indeed, off-diagonal components of the effective elastic tensor exhibit localized variations with sharp transitions, which make the inversion become more constrained, and the sampled posterior distribution sharper. In this case, further reducing the proposal step size or adopting a data covariance matrix that incorporates spatial correlations helps increase the acceptance rate.

Chapter 5

Downscaling inversion for stratigraphic and fault-related uncertainty quantification

5.1 Introduction

In this chapter, we extend the work presented in Chapter 4 with a downscaling application whose objective is to resolve uncertainty on both small-scale fault and stratigraphic structures. For this purpose, the methodology combines stochastic fault modeling with geostatistical simulations for generating multiple realizations of faulted subsurface models incorporating small-scale variations in elastic properties.

A wide range of geostatistical simulation techniques have been developed and commonly adopted to represent the model spatial variability in Earth-science inverse problems. These include variogram-based methods, routinely used to simulate permeability and porosity fields (Goovaerts, 1997); multipoint geostatistical methods, used to model complex spatial patterns such as meandering channelized aquifers (Strebelle, 2002); probability field simulations (PFS, Srivastava, 1992) or sequential Gaussian simulation (SGS, Deutsch et al., 1992), often employed to simulate stochastic realizations of petrophysical properties (e.g., Grana et al., 2021). One of the main advantages of using a geostatistical approach is that simulated realizations can easily be used to reflect property variations at different scales, from local (small-scale heterogeneities) to regional (large-scale trends). Moreover, these techniques enable the integration of different data sources, such as well logs, seismic tomography, and rock physics constraints, ensuring that the generated spatial distribution of the modeled properties honors the data (e.g., Hansen et al., 2006, Grana et al., 2023).

In this work, for its simplicity and computational efficiency, we adopt Turning Bands Simulation (TBS) (Matheron, 1973, Carr, 2002) for generating Gaussian realizations of continuous properties. In particular, assuming prior knowledge on

the macro-scale elastic properties, Gaussian random fields (GRF) are applied to generate small-scale perturbations, thus deriving more realistic subsurface models. In the following, we first describe the model parametrization and the Bayesian framework of the inverse problem; then, we present results of a synthetic application.

The results presented in this chapter have been published in the following journal paper:

- Ruggiero G., Cupillard P. and Caumon G. (2026) Downscaling FWI images for reducing structural uncertainty, *Geophysical Journal International*, Volume 244, Issue 3.

5.2 Model Parametrization

In this section, we present the methodology employed to generate subsurface models characterized by faults and small-scale stratigraphic features, and we detail the model parameters that constitute the unknowns in the downscaling inverse problem.
Before applying the kinematic operator described in Section 4.2 to simulate fault displacement, we generate a finely stratified elastic medium in the pre-faulting space, using piecewise constant values in each layer and Gaussian random fields (GRF). We describe here the methodology using the P-wave velocity V_p, noting that the same approach can be applied for V_s and ρ.
We begin by discretizing the domain with a regular Cartesian grid $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^2 \mid i = 1, \dots, N\}$, and subdivide it into J initially flat rock units separated by $J-1$ internal horizons. The P-wave velocity within the domain is decomposed into a macro-layered velocity field $\mathbf{W}$, defined by a homogeneous velocity w_j in each rock unit j, and a residual field simulated by generating samples from an anisotropic Gaussian random field $\mathbf{Z}$ of zero mean. The residual field $\mathbf{Z}$, generated with Turning Bands Simulation (TBS) (refer to the Appendix b for details), is added to the macro-layered velocity to reproduce thin-layering structures inside each major layer (Fig. 5.1). The total fine-layered velocity at $\mathbf{x}$ can be thus expressed as

$$V_p(\mathbf{x}) = \sum_{j=1}^{J} W_j \mathbb{1}_j(\mathbf{x}) + \ Z(\mathbf{x}), \tag{5.1}$$

where $\mathbb{1}_j(\mathbf{x})$ is the indicator function equal to 1 if $\mathbf{x}$ belongs to unit j and 0 elsewhere. The characterization of the residual field $\mathbf{Z}$ is governed by the parameters defining the variogram — such as sill, correlation range and anisotropy — and describing its spatial variability. In general, the specific parametrization of the simulation depends on the geological settings and on data available. For this work, we assume a flat stratigraphy in the geological model, so we consider only

vertical changes in the velocity, which are controlled by the vertical correlation length l_y of $\mathbf{Z}$, and we neglect lateral velocity variations within each stratigraphic layer. Assuming prior knowledge coming from indirect data on the macro-layered velocity model (Fig. 5.1a), downscaling aims to reconstruct the 1-D fine-scale velocity (Fig. 5.1b) through inference of the small-scale perturbations produced by the residual field $\mathbf{Z}$. In order to reduce the model space dimensions and facilitate an efficient sampling, we consider a subset of points $\mathbf{s} = \{s_i \mid i = 1, \ldots, N_i\}$ where the residual field $\mathbf{Z}$ is inferred, as done in the pilot point method (de Marsily et al., 1984). The choice of the $\mathbf{s}$ pilot points is guided by the vertical (since we neglect lateral variations) spatial variability, defined by l_y, and by the accuracy we aim to achieve in the inference of the residual velocity field. The residual velocity in the whole domain is then obtained with a conditional turning bands simulation at the values $\mathbf{Z}(\mathbf{s})$.
Within the Bayesian framework of this downscaling problem, the model is then defined as a structural-velocity model consisting of two components $\boldsymbol{\theta} = \{\boldsymbol{\vartheta}, \mathbf{Z}(\mathbf{s})\}$, where $\boldsymbol{\vartheta}$ are variables describing the fault model (i.e., location, dip, maximum slip and spatial extent of the fault-induced deformation), and $\mathbf{Z}(\mathbf{s})$ are the residual velocity values at the defined pilot points.

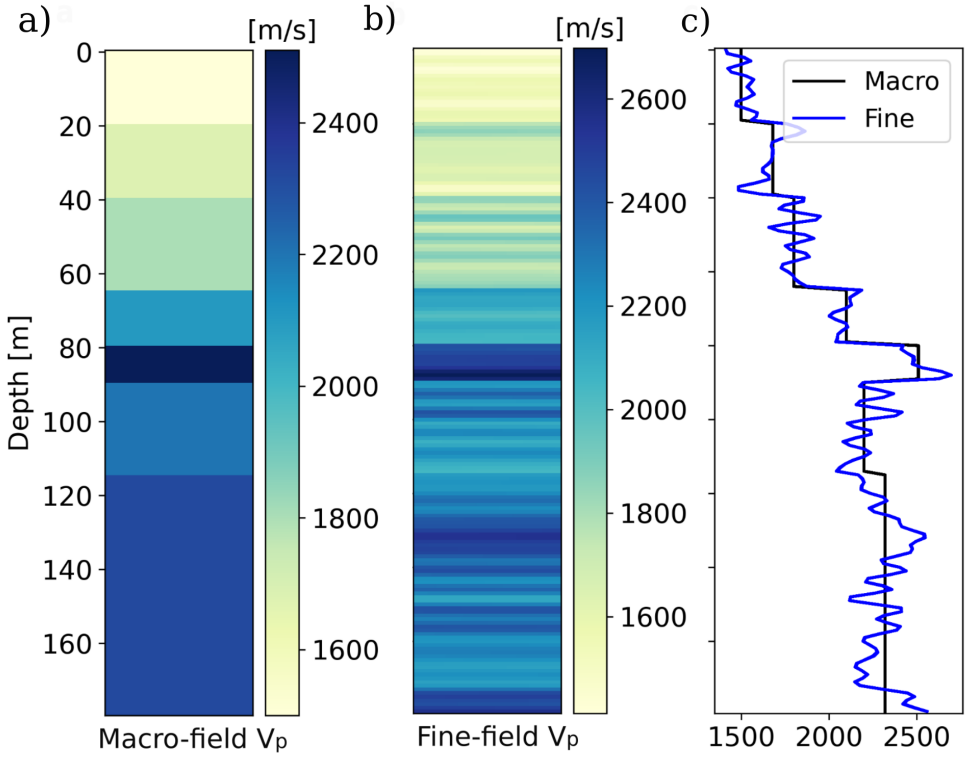

Figure 5.1: a) 1-D macro-layered velocity field, obtained by defining a constant velocity in each macro layer. b) Fine-layered velocity field, generated as described in Eq. (5.1). c) Corresponding vertical velocity profiles.

5.3 Bayesian framework with MCMC

The objective of the downscaling is to quantify uncertainty on both P-wave residual velocity $\mathbf{Z}$ and fault structure $\boldsymbol{\vartheta}$ of a fine-scale elastic medium $\mathbf{m}(\boldsymbol{\theta} = \{\boldsymbol{\vartheta}, \mathbf{Z}(\mathbf{s})\})$, knowing its smooth effective medium $\mathbf{c}^*_{\mathbf{obs}}$. As for the downscaling framework described in Section 4.3, the observed effective medium $\mathbf{c}^*_{\mathbf{obs}}$ refers to a smooth and homogenized FWI model, achieved via HFWI, which gathers the information present in the seismic data and does not contain heterogeneities of scale smaller than the homogenization wavelength λ_0; whereas the predicted data $\mathbf{c}^*$ is computed with the non-periodic homogenization operator $\mathcal{H}$ as $\mathbf{c}^* = \mathcal{H}(\mathbf{m}(\boldsymbol{\theta}))$. In this case as well, we specify uninformative prior distributions $\pi(\boldsymbol{\theta})$ for the model parameters and Gaussian noise assumptions for evaluating the likelihood $\pi(\mathbf{c}^*_{\mathbf{obs}} \mid \boldsymbol{\theta})$ (4.3).
Bayesian inference then determines the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{c}^*_{\mathbf{obs}})$ of the parameter set $\boldsymbol{\theta}$ according to Bayes' equation:

$$\pi(\boldsymbol{\theta} \mid \mathbf{c}^*_{\mathbf{obs}}) \propto \pi(\mathbf{c}^*_{\mathbf{obs}} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{5.2}$$

expressing the probability that $\boldsymbol{\theta}$ provides a fine scale elastic medium $\mathbf{m}(\boldsymbol{\theta})$ whose effective medium is $\mathbf{c}^*_{\mathbf{obs}}$.
An adaptive Metropolis-Hastings MCMC scheme is employed to characterize $p(\boldsymbol{\theta} \mid \mathbf{c}^*_{\mathbf{obs}})$. Specifically for the implementation in this study, the Markov chain consists of two possible updates:

- the velocity model: at the iteration k, one pilot point $s_i \in \mathbf{s}$ is randomly selected and the residual velocity $Z(s_i)$ is updated according to a Gaussian proposal distribution $q(Z(s_i))$. To simplify the model and reduce computational complexity, the velocity at the selected pilot point is perturbed independently, without regard for the velocity values of other pilot points. The residual velocity in the whole medium $\mathbf{Z}^k$ is simulated conditioned on the values at all pilot points, $\mathbf{Z}(\mathbf{s})^k$. In this way, the Markov chain explores the parameter space by adjusting the values $\mathbf{Z}(\mathbf{s})$, which in turn influences the simulated values at other locations. The total fine-layered velocity field $\mathbf{V}_p^k$ is then obtained with equation (5.1).

- the structural model: at the iteration k, all the fault-related parameters $\boldsymbol{\vartheta} = (x_c, y_c, a, b, \phi, D_{\max})$ are perturbed according to a Gaussian proposal distribution $q(\boldsymbol{\vartheta})$. The kinematic fault operator, defined by the proposed fault model $\boldsymbol{\vartheta}^k$, is then applied to simulated the fault displacement in the fine-scale elastic medium $\mathbf{m}(\boldsymbol{\theta})^k$.

A simplified version of the implemented algorithm is sketched in Algorithm 3.

Algorithm 3 Sketch of the MCMC algorithm

Input: Observed data $\mathbf{c}^*_{\mathbf{obs}}$; Data covariance matrix $\mathbf{\Sigma}$; $\mathbf{s} = \{s_i \mid i = 1, \ldots, N_i\}$ Pilot point set; Proposal distributions $q(\boldsymbol{\vartheta})$ and $q(Z(s_i))$;
Output: Posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{c}^*_{\mathbf{obs}})$

Initialize random fault parameters $\boldsymbol{\vartheta}^0$ from prior $p(\boldsymbol{\vartheta})$;
Initialize residual velocity at all pilot points with $\mathbf{Z}(\mathbf{s})^0 = 0$;
for $k = 1$ to K **do**
    if $k \bmod 2 = 0$ **then** ▷ Fault Step
        Draw random perturbations $\mathbf{x}$ from $q(\boldsymbol{\vartheta})$;
        Update fault parameters $\boldsymbol{\vartheta}^k \leftarrow \boldsymbol{\vartheta}^{k-1} + \mathbf{x}$;
        Define $\mathbf{V}_p^k \leftarrow \mathbf{V}_p^{k-1}$;
    else ▷ Velocity Step
        Randomly select pilot point index $i \sim \mathcal{U}(\{1, \ldots, N_i\})$
        Draw random perturbation x from $q(Z(s_i))$;
        Update residual velocity at selected point $Z(s_i)^k \leftarrow Z(s_i)^{k-1} +$ x
        Perform TBS conditioned on the new set $\mathbf{Z}(\mathbf{s})^k$ to derive $\mathbf{Z}^k$
        Compute the fine-layering velocity $\mathbf{V}_p^k \leftarrow \mathbf{W} + \mathbf{Z}^k$
        Define $\boldsymbol{\vartheta}^k \leftarrow \boldsymbol{\vartheta}^{k-1}$
    end if
    Update model parameters $\boldsymbol{\theta}^k \leftarrow \{\boldsymbol{\vartheta}^k, \mathbf{v}_p^k\}$;
    Derive fine-scale elastic medium $\mathbf{m}(\boldsymbol{\theta})^k$;
    Compute predicted data $\mathcal{H}(\mathbf{m}(\boldsymbol{\theta})^k)$;
    Evaluate the likelihood probability $p(\mathbf{c}^*_{\mathbf{obs}} \mid \boldsymbol{\theta}^k)$ (4.3);
    Compute acceptance probability α (2.38);
    Draw random number $r \sim \mathcal{U}(0, 1)$;
    if $r < \alpha$ **then**
        Accept sample $\boldsymbol{\theta}^k$
    else
        Reject sample $\boldsymbol{\theta}^k \leftarrow \boldsymbol{\theta}^{k-1}$
    end if
end for

5.4 Example

The methodology is tested on a synthetic data application. A model involving a single fault that offsets a fine-layered stratigraphy is used to represent the reference subsurface model (Fig. 5.2b). It is parametrized with a regular grid where each cell has a 1 m × 1 m size for a total domain of 180 m × 100 m. To build the reference fine-scale elastic medium $\mathbf{m}(\boldsymbol{\theta}_{ref})$, we define a constant P-wave velocity in each macro-layer w_j and a covariance or variogram model describing the spatial structure of the Gaussian random field $\mathbf{Z}$ employed to simulate the fine-layering velocity model $\mathbf{V}_p$. For this study, $\mathbf{Z}$ is obtained by Turning Bands Simulation performed using the gstlearn Python package from MINES Paris (gstlearn, 2023). Because we aim at reproducing a horizontal stratigraphic orientation, we select the variogram model parameters to account for anisotropy: the horizontal direc-

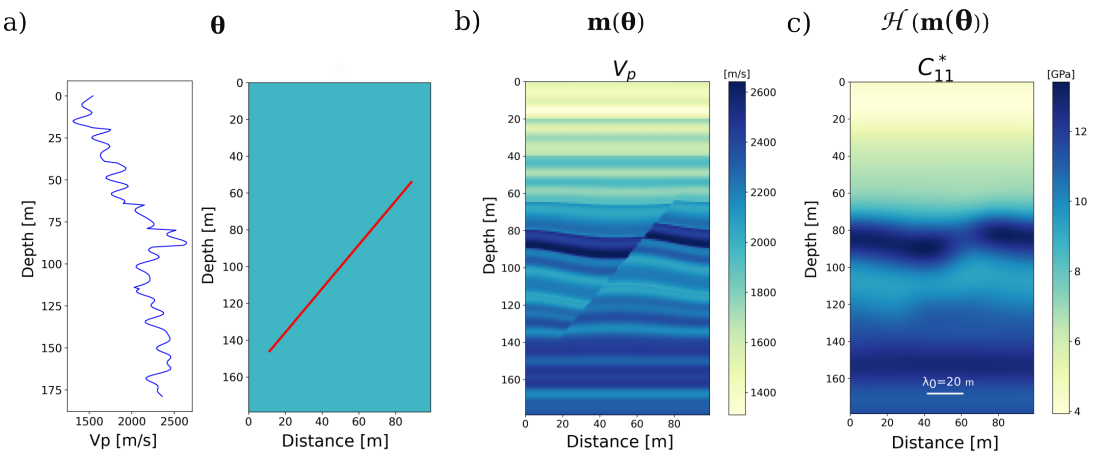

Figure 5.2: a) Reference 1-D velocity model $\mathbf{V}_p$ and structural model $\boldsymbol{\vartheta}$. b) Fine-scale elastic medium $\mathbf{m}(\boldsymbol{\theta})$ used as reference medium for the synthetic application. The elastic medium is illustrated here in terms of P-wave velocity V_p. The elastic properties V_s and ρ, required for the elastic homogenization, are computed from V_p as explained in Section 5.4. c) First component of the effective elastic tensor $\mathbf{c}^* = \mathcal{H}(\mathbf{m}(\boldsymbol{\theta}))$ obtained by homogenizing the reference fine-scale medium with $\lambda_0 = 20$ m.

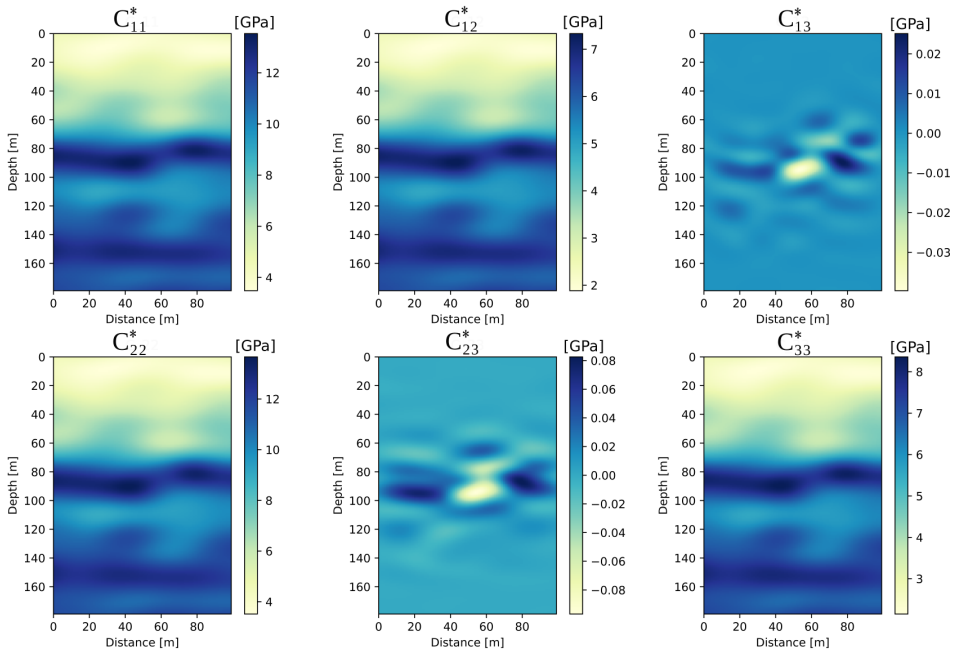

Figure 5.3: The six components of the effective elastic tensor $\mathbf{c}^*_{\mathbf{obs}}$ representing the observation in the synthetic application. These are obtained by homogenizing the reference medium with a scale-separation spatial wavelength $\lambda_0 = 20$ m and by adding spatially correlated Gaussian noise $\boldsymbol{\eta}$.

tion is chosen as the primary orientation where spatial correlation is stronger. Thus, we choose a Gaussian covariance function with vertical correlation length $l_y = 5$ m and a horizontal to vertical anisotropy ratio of 500 : 1, typical of deep-water or deltaic depositional environments (Pyrcz & Deutsch, 2014). For the simplified test of this study, we use a set of deterministic values for the geostatistical variables of the variogram model, including l_y. Hence we assume that this information is known from other data sources.

Once the velocity model is defined, fault displacement is applied as described in Section 4.2. The S-wave velocity V_s and density ρ properties, also needed for the forward step, are calculated from V_p using the fixed ratio $V_p/V_s = 1.8$ and Gardner's relation $\rho = 0.31\ V_p^{0.25}$.

In this synthetic application, we approximate the observation (the smooth HFWI model) with the homogenized reference medium $\mathcal{H}(\mathbf{m}(\boldsymbol{\theta}_{ref}))$, to which we add spatially correlated Gaussian noise $\boldsymbol{\eta}$ to increase the complexity of the inverse problem. Homogenization is carried out with the FFH approach as described in Section 3.1.4 using a regularly spaced grid of 1 m × 1 m. Specifically, $\mathcal{H}(\mathbf{m}(\boldsymbol{\theta}_{ref}))$ is computed from V_p, V_s and ρ by applying the homogenization operator with a scale separation parameter $\lambda_0 = 20$ m. As so, no spatial variations of wavelength smaller than λ_0 are kept in the smooth effective properties. For the correlated noise $\boldsymbol{\eta} \sim N(0, \boldsymbol{\Sigma})$, we choose a Gaussian covariance function with characteristic length scales $l_x = l_y = 20$ m equal to λ_0, in order to reproduce possibly structured correlations that may be present in an elastic model obtained from HFWI, and noise variance equal to 4% of the mean value of the effective elastic properties. The resulting tensor $\mathbf{c}^*_{\mathbf{obs}} = \mathcal{H}(\mathbf{m}(\boldsymbol{\theta}_{\mathbf{ref}})) + \boldsymbol{\eta}$ is shown in Figure 5.3. As can be observed, the off-diagonal terms $\mathbf{c}^*_{13}$ and $\mathbf{c}^*_{23}$ of the effective properties, which give information about effective medium anisotropy, exhibit higher values at the velocity discontinuities produced by the fault and by the high-contrast velocity layer in the central part of the model.

In this case as well, we compute an approximate data covariance matrix $\boldsymbol{\Sigma}$ instead of using the exact covariance of the generated noise $\boldsymbol{\eta}$, which considers errors in the six elastic coefficients of the data vector $\mathbf{c}^*_{\mathbf{obs}} = (\mathbf{c}^*_{11}, \mathbf{c}^*_{12}, \mathbf{c}^*_{13}, \mathbf{c}^*_{22}, \mathbf{c}^*_{23}, \mathbf{c}^*_{33})$ are correlated, but does not account for spatial correlation between data points, meaning that errors are considered independent for each location.

A key aspect of the parametrization we adopt is the choice of the pilot points $\mathbf{s}$ where the residual velocity $\mathbf{Z}$ is sampled, which may significantly affects the feasibility and efficiency of the method. For this application we choose to use points regularly spaced along the vertical direction with a spacing between two successive points $\Delta s = 2$ m, i.e. approximately half the vertical correlation length l_y, to control the recovery of the reference fine-scale velocity heterogeneities through the inversion process. The choice of this parametrization will be further discussed in Section 5.5.

5.4.1 Results

We apply the MCMC sampling scheme described in Algorithm 3 to sample posterior models. To verify the convergence, we run 4 MCMC parallel chains starting from different initial models. Examples of starting models are shown in Figure 5.6a. The convergence plots of the 4 runs are shown in Figure 5.4, which displays the negative log-likelihood, equivalent of the misfit function in Bayesian inference, and the acceptance rate of all the realizations. We discard the first 50 000 realizations, considered part of the burn-in period, and compute the posterior probability based on the remaining 180 000 realizations in the stationary regime, for which the acceptance rate is around 10 %. MCMC convergence is assessed with the Gelman–Rubin statistic $\hat{R}$, which is below 1.1 for all the fault-related parameters $\boldsymbol{\vartheta}$, indicating convergence. For the residual velocity defined at the pilot points $\mathbf{Z}(\mathbf{s})$, convergence is inherently slower because only one pilot point is perturbed at each iteration. This behavior is reflected by the small decreasing trend of the log-likelihoods after the burn-in (Fig. 5.4b). Consequently, although not all velocities sampled at the 90 pilot points achieved $\hat{R} < 1.1$, we chose to terminate the sampling as a computational compromise.

The minimum and maximum values of the uniform prior distributions and the marginal posterior distributions of the fault parameters $\boldsymbol{\vartheta}$ are shown in Figure 5.5b. It is evident from the figure that reference values $\boldsymbol{\vartheta}_{ref}$ are accurately recovered. In particular, all the posterior distributions are sharply peaked and tightly concentrated around the reference values; and in agreement with results presented in Chapter 4, the semi-major a and semi-minor b axis of the ellipse, defining the spatial extent of the fault-induced deformation, appear less constrained by the data. In this application as well, the effective anisotropy information provided by the off-diagonal components $\mathbf{c}^*_{13}$ and $\mathbf{c}^*_{23}$ of the full elastic tensor is highly informative especially about the fault position, dip angle, and its maximum displacement at the center.

In addition to a reduction of uncertainty in the structural model, downscaling provides a reduction of uncertainty in the fine-layering velocity model $\mathbf{V}_p$. The inverted $\mathbf{V}_p$ is expressed as the posterior mean computed with posterior samples of the 4 chains. Figure 5.5a shows the reference, prior, and posterior velocity with depth at $x = 50$ m. While the prior mean reflects the macro-layered velocity (Fig. 5.1a), since the initial residual velocity field is set to zero (see Alg. 3), the mean of the posterior follows the shorter wavelength trend of the reference model's velocity profile. However, the posterior mean velocity doesn't align with the reference profile at all depths. Some regions — more affected by data noise — exhibit higher uncertainty, as indicated in Figure 5.5a by a larger confidence interval, whereas others are better constrained.

The mean, standard deviation and relative error maps are computed and displayed in Figure 5.7. The relative error is calculated as the difference between the reference model and the mean model, divided by the standard deviation. These

posterior statistics confirm that the fault model is successfully recovered with low uncertainty, whereas a relatively large uncertainty is associated to the fine-layered velocity, especially in the upper and lower parts of the domain. Overall, these results demonstrate the ability of the methodology to resolve uncertainty on the unknown small-scale features in the synthetic case, in spite of the spatially correlated noise added to the reference homogenized model. The accuracy of the results stems from the fact that this correlated noise has a smaller horizontal variogram range than the prior horizontal range of the residual velocity field. Consequently, the proposed model samples manage to filter out part of the noise at the expense of a lower data fit. A complementary explanation for the obtained accuracy comes from the relatively strong control of the elastic parameters provided by the deterministic macro-layered velocity. Indeed, the continuous velocity residual field tends to preserve the strong anisotropy at the macro-layer boundaries. These aspects are further discussed in the following section.

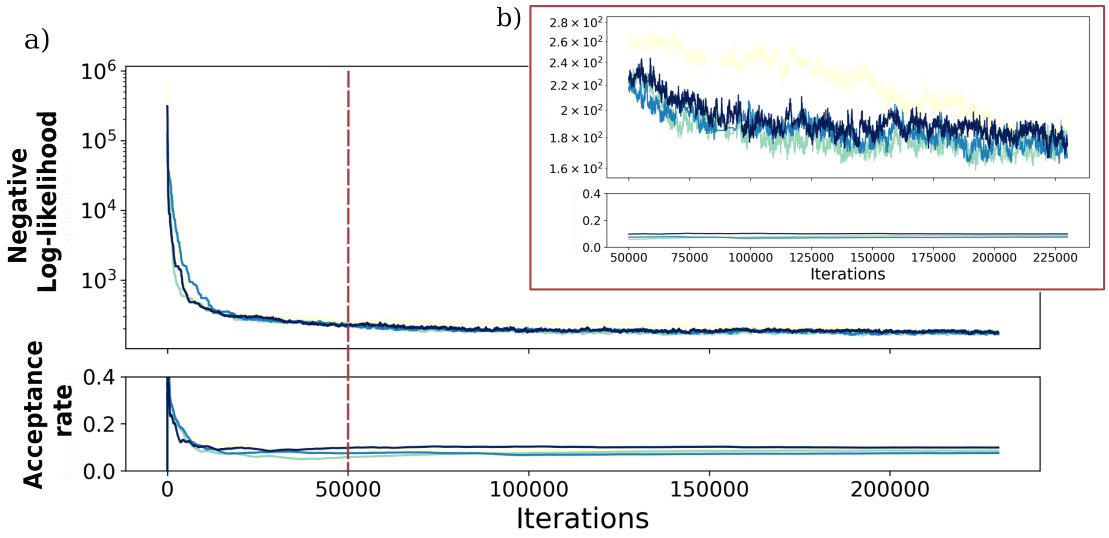

Figure 5.4: a) Top graph: Negative log-likelihood functions of 4 parallel MCMC chains for number of iterations. The burn-in phase is considered to finish after 50 000 iterations when the functions reach a plateau and start sampling the posterior. Bottom graph: Acceptance rate of 4 parallel MCMC chains for number of iterations. As the number of iterations increases, the acceptance rates stabilize around 10 %. b) Zoomed-in view of the two graphs post burn-in.

5.5 Discussion

In this chapter the downscaling inverse problem aims to infer, in addition to fault-related model parameters, the small-scale vertical variations in the medium's elastic properties, specifically its P-wave velocity, in order to jointly account for

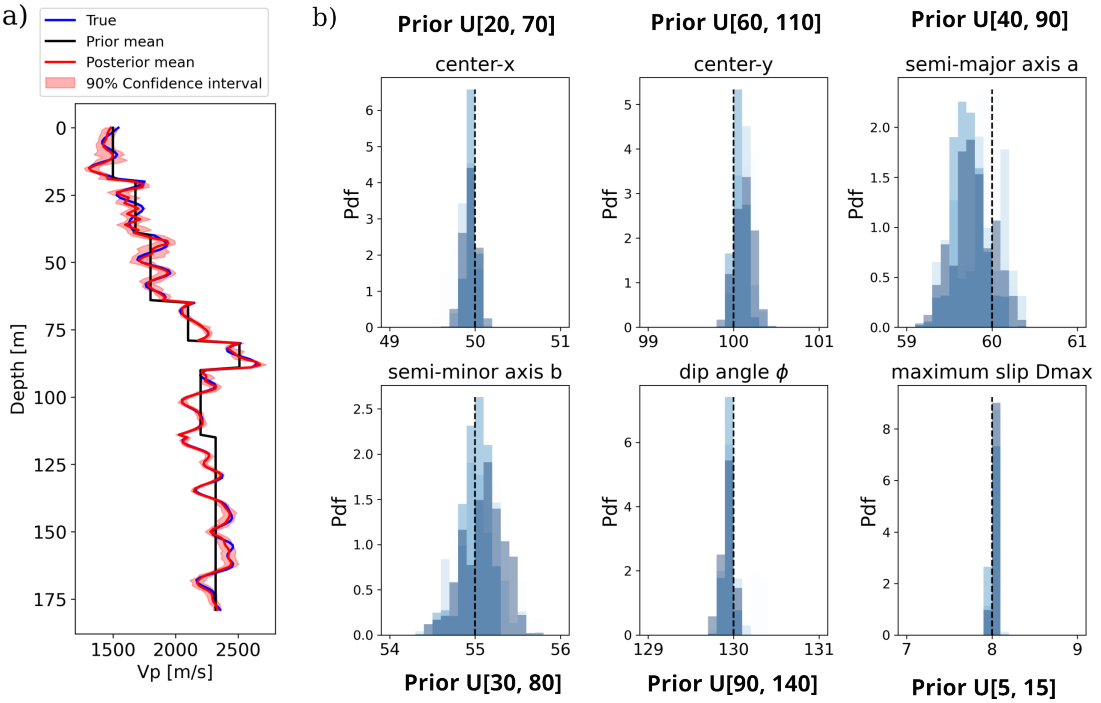

Figure 5.5: Inversion results of 4 MCMC chains. a) V_p velocity profiles extracted at x = 50 m. The blue line is the reference fine-scale velocity profile. The black and red lines show the mean of the prior and posterior distributions. Red-shaded region marks the 90 % confidence interval (5th and 95th percentile range), capturing the most probable range of velocity values. b) Posterior densities of the fault-related parameters $\boldsymbol{\vartheta}$ from 4 MCMC chains (shown in different shades of blue) after burn-in. Black dotted lines indicate target values from the reference fault structure in Figure 5.2. The minimum and maximum values of the uniform priors for each model parameter are indicated in bold.

fault-related and stratigraphic uncertainty. The inverse problem is cast within a Bayesian probabilistic framework to retrieve an ensemble of possible subsurface velocity models at the small-scale consistent with the smooth HFWI solution and the a *priori* knowledge. In the presented synthetic application, in which the HFWI medium is synthetically generated by adding Gaussian noise to the homogenized reference elastic medium, the proposed approach has shown good results in capturing both fine-layering stratigraphy and fault-related variables. The notably good performance is attributed in part to specific simplifications that were made; these, together with other relevant considerations, are analyzed in the next sections.

5.5.1 Geological complexity

In addition to the simplified assumptions relative to the fault structure, already discussed in Section 4.5, additional comments need to be done for what con-

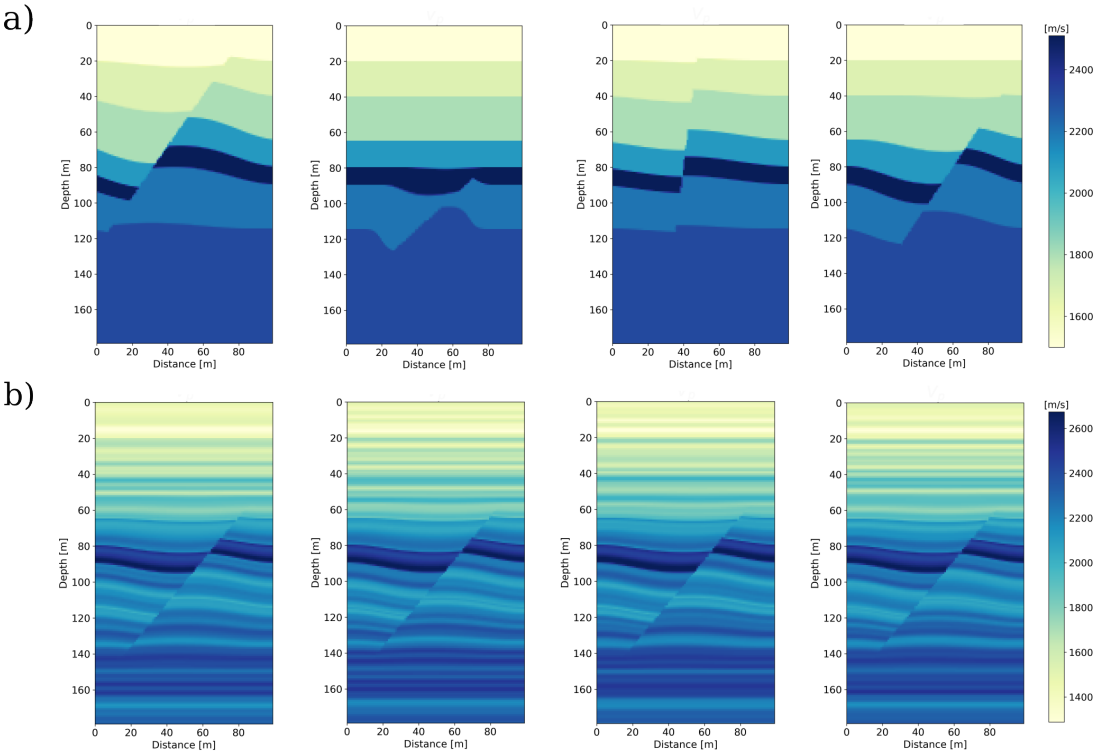

Figure 5.6: a) Starting models for different chains. The initial fault parameters $\boldsymbol{\vartheta}$ are randomly sampled from the uniform prior distributions, whereas the initial velocity $\mathbf{V}_p$ corresponds to the macro-scale velocity field. b) Example of posterior models for different chains. These models show a fault that matches the target structure and a fine-layering velocity $\mathbf{V}_p$ that closely resembles the reference one (Fig. 5.2a).

cerns the velocity model. For the sake of simplicity, we decide to use a reference model in which the velocity varies only along the vertical direction to reproduce a horizontal stratigraphy and we derive the S-wave velocity from P-wave velocity values assuming a fixed ratio of $V_p/V_s = 1.8$. When only V_p is known, a V_p/V_s ratio of roughly $1.7 - 1.8$ is commonly assumed in numerical simulations since many crustal rocks exhibit a Poisson's ratio near $\nu \approx 0.25$, which corresponds to $V_p \approx 1.73\, V_s$ (Sheriff & Geldart, 1995, Komatitsch & Tromp, 1999, Gercek, 2007). Alternatively, empirical $V_p - V_s$ relationships derived from lab data, well logs, or field measurements are adopted, such as the one proposed by Castagna et al. (1985) for clastic sediments or by Brocher (2005) for crustal rocks. The total velocity is generated by combining a macro-layered velocity field, defined by a homogeneous velocity in each layer, with a continuous fine-scale residual field. One of the main simplifications in this application is the use of the correct macro-layered velocity model, which yields predictions highly consistent with the observations, despite the added noise. The residual velocity is simulated with a GRF through geostatistical approach. Hence, its spatial variability is defined by

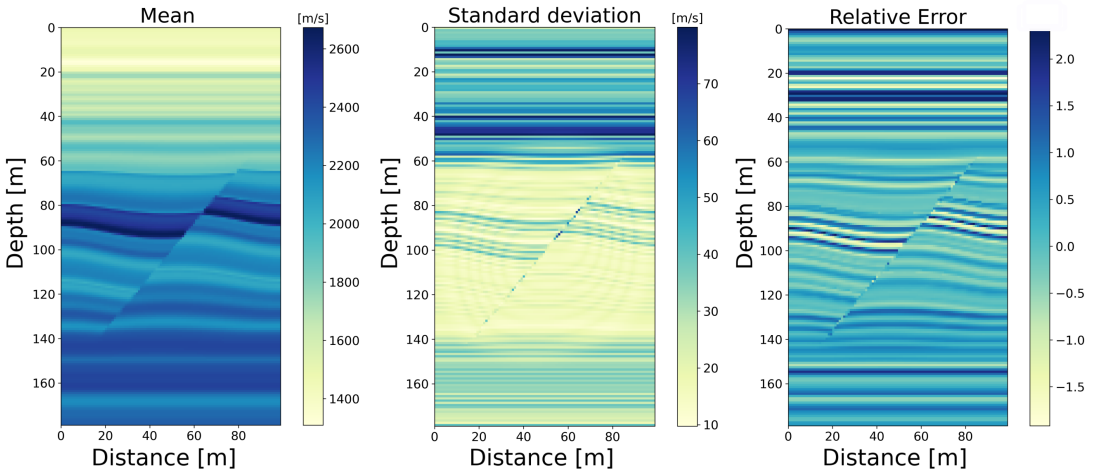

Figure 5.7: Posterior statistics post burn-in: mean, standard deviation and relative error of the small-scale elastic medium $\mathbf{m}(\boldsymbol{\theta})$ obtained from posterior samples of 4 MCMC chains. The fault structure has been successfully recovered and well-constrained, whereas higher uncertainty is observed for some fine-scale velocity layers.

the vertical correlation length of the variogram model. In this framework, having an accurate estimate of the geostatistical variables describing the variogram model is determinant in recovering the fine-scale velocity through downscaling inversion. In this synthetic test, we assume that the variogram model (e.g., sill, correlation length, anisotropy) is known, which is a further significant hypothesis. In a real case study, a common practice is to fit a chosen model to an empirical variogram computed using the available borehole data. In this case, the approach can be easily extended to modeling lateral heterogeneities provided there is sufficient data to define the horizontal correlation length. If instead data are not available or insufficient to reliably constrain the variogram model, variables defining the variogram (such as range, sill, and anisotropy) could also be treated as random variables and sampled through inversion. Another limitation of this parametrization is that the residual velocity field obtained with geostatistical simulation relies on the hypothesis of stationarity of the model; however, in many real scenarios, the velocity profile is non-stationary and shows depth trends. In this case, several residual fields with varying statistical properties could be simulated in each macro-layer or alternatively, geostatistical algorithms for non-stationary models could be adopted.

To infer the velocity profile with MCMC, we use a pilot point parametrization strategy, wherein the inversion focuses on a reduced set of velocities defined at selected point locations. The velocity field in the whole domain is then obtained with the geostatistical simulation conditioned on the velocity values sampled at the pilot point locations. This approach reduces the effective parameter space

explored during MCMC sampling, while approximately honoring the spatial correlation structure. In this scheme, a critical choice is the number and spacing of pilot points, which affects the trade-off between model accuracy and computational feasibility. Indeed, including as many pilot points as possible for calibration leads to a good fit of observed data but increases the computational burden. In a stationary domain, the choice of the pilot points spacing should be guided by the spatial correlation structure of the velocity field (i.e., the variogram range) (e.g., Alcolea et al., 2006, Gómez-Hernánez et al., 1997, Aissat et al., 2023). For the synthetic application of this work, we define regularly spaced pilot points with a spacing of approximately half the variogram range to adequately capture the spatial heterogeneity of the velocity field. An adaptive approach with progressive refinement of the homogenization wavelength λ_0 and pilot points grid could be also adopted in a complex real-study case. In such case, the MCMC inversion would start with a large λ_0 (smoother effective properties) and a coarse pilot points grid (fewer points where velocity is inferred) and, as the algorithm converges, gradually decrease λ_0 and refine the pilot points grid. This multi-scale strategy would allow to progressively increase the resolution and the level of small-scale details as the posterior distribution becomes better constrained, thus improving convergence efficiency.
It should also be noted that, for simplicity, in the MCMC algorithm we adopt a proposal scheme where the velocity at pilot points is perturbed independently (see Section 5.3). However, because these points are spatially correlated ($\Delta s < l_y$), independent perturbations could violate the assumed spatial structure of the variogram model if the perturbations are too large or uncorrelated with nearby points. We have assessed the limited impact of this simplification in our experiments by comparing the empirical variogram of conditional simulations from the posterior samples with the theoretical (assumed) model (see Appendix c for additional information). As an alternative, pilot points values could be sampled from the conditional probability distribution defined by kriging and the estimation variance computed from the neighboring pilot points (Freulon & de Fouquet, 1993).

5.5.2 Homogenization and geological heterogeneities scales

As discussed in Section 4.5, a key component of the described methodology is the cutoff wavelength λ_0 used to homogenize the elastic properties. In this synthetic test, we approximate the observation (homogenized FWI model) with the homogenized reference model using $\lambda_0 = 20$ m, to which we add Gaussian noise. Choosing λ_0 as four times bigger than the vertical spatial correlation range of the velocity field, and considering a spatially correlated Gaussian noise are meant to increase the complexity of the inverse problem for the synthetic case. We would expect that for shorter λ_0, geological heterogeneities at smaller scales would be preserved, and the solution would be better constrained by the observations; while larger λ_0 lead to greater uncertainty in the inferred model parameters. Never-

theless, the fine-scale structures of the target medium are recovered. This can be attributed to the significant effective anisotropy observed at the fault position (Fig. 5.3), providing valuable information to the downscaling procedure when constraining this structure. The effective anisotropy is enhanced by the presence of a high-contrast velocity layer in the reference model which is offset by the fault. In addition, the use of the same geostatistical parametrization to simulate fine-layering velocity in both predictions and observations and of a dense pilot-points grid facilitates the accurate recovery of the velocity field.
As explained in Section 4.5, in the case of real data application, λ_0 is chosen according to the minimum resolvable wavelength λ_{m} in order to ensure that the original and homogenized models produce similar wavefields.

5.5.3 Computational aspects

In this application, the sampling involves a much larger number of model parameters (96) compared to the applications presented in Chapter 4, where only the fault-related parameters were inferred, leading to slower convergence. This is aggravated by the fact that velocity is perturbed at a single pilot point per iteration. Indeed, updating all 90 velocity values simultaneously can lead to faster exploration of the parameter space, but it may reduce the acceptance rate and require more careful tuning of the proposal distribution. Alternatively, advanced techniques such as parallel tempering or gradient-based proposals can accelerate convergence and enable more efficient sampling of the solution ensemble (see Section 2.3.2.3).
An advantage of the methodology is that the forward solver (the FFT-based elastic homogenization) is based on a regular-grid formulation, is computationally much faster than, for instance, spectral element methods that require mesh generation. The cost of one forward modeling for the two-dimensional model used in this study is less than one second on a laptop equipped with 20 cores. As a result, even though MCMC requires a large number of iterations to converge, the computational cost remains affordable. In addition, the downscaling approach proposed in this work can be used as a target-oriented uncertainty characterization framework by applying it on a subset of a larger-scale inverted model. This allows to further spare computational time by reducing the dimension of the model space.

Chapter 6

Conclusions

6.1 Motivations for this thesis

The proposed approach attempts to bridge the gap between full-waveform inversion (FWI) results, which often suffer from a lack of geological realism due to smoothing of sharp discontinuities, and geological interpretations. Indeed, subsurface structural models are conventionally built from structural interpretations derived from reflection seismic images, or directly from FWI elastic models. However, it is well known that, due to the noisy and non-unique nature of the seismic inversion process, these interpretations are inherently uncertain. One of the reason is that seismic inversion is performed with limited frequency-band data; hence, geological structures below the resolution limit (e.g., thin layers, small faults, sharp boundaries) cannot be resolved, leading to ambiguity in their interpretations.

Despite structural interpretation uncertainty from seismic imaging is a well-known and current topical issue, we believe that the existing strategies still lack a methodology that delivers accurate geological information at small scales (without scarifying resolution) and ensures feasibility of uncertainty assessments while maintaining reasonable computational costs. This provides the motivation for the downscaling, or inverse homogenization, technique presented in this thesis.

Downscaling is used in this work to derive an ensemble of plausible geological interpretations of the subsurface at the small-scale, given an estimation of its smooth elastic properties. Because the proposed downscaling framework could be applied locally, for example to specific areas of interest where imaging is deemed reliable but the resolution is insufficient to support decision making, and uses a computationally affordable forward modeling (the non-periodic homogenization), the use of stochastic inversion techniques, such as MCMC sampling schemes, becomes feasible. This represents a significant advantage compared to challenges encountered, for instance, in stochastic FWI schemes, for which the computational cost of solving the forward problem can be prohibitively expensive for large-scale problems.

6.2 Thesis achievements

In this thesis the downscaling inverse problem is applied to a 2-D subsurface model with the aim of resolving uncertainty in fault structures and small-scale (below seismic resolution) vertical variations in the elastic properties, thus reflecting uncertainty in the subsurface stratigraphy. A central aspect of this work is therefore the implementation of a downscaling Bayesian framework that explicitly encodes geological knowledge into the inversion process. In a first application, this is accomplished by adopting a fault-related model parametrization describing fault geometry and displacement and by employing a kinematic fault modeling technique to stochastically simulate the fault-induced deformation of the elastic properties. In a second application, this implementation is extended to account for small-scale vertical variations in the subsurface velocity, modeled through geostatistical simulations. In this case, by adopting a pilot-point parametrization, the downscaling inverse problem is applied to infer both the fault-related model parameters and the velocity values at the pilot points. A simple Metropolis-Hastings algorithm is used to characterize the posterior probabilities of the unknown model parameters describing the fault and velocity model. The downscaling implementation was evaluated through applications involving synthetic subsurface models and observed data. In all the synthetic tests presented in this thesis, the proposed methodology showed good efficiency, producing accurate results together with consistent and informative uncertainty estimates, suggesting its potential for practical applications.

6.3 Critical Assessment and Future Perspectives

In this thesis, we evaluated the proposed methodology using synthetic tests to assess its performance in simplified scenarios. Nevertheless, certain limitations remain, and several aspects require further investigation before the methodology can be reliably applied to more general and realistic geological cases.
The most relevant points to be explored in future research are outlined below.

- **Geological complexity**. Real geology is considerably more complex than the synthetic cases studied here and accurate prior geological information is not always available. Consequently, parametrizing the subsurface in such conditions becomes substantially more challenging.
  An important aspect to consider is that, in nature, fault geometries are often irregular and may intersect with one another, forming complex networks. Therefore, extending the approach to account for multiple intersecting faults would be a first step to better reflect natural complexity. In addition, fault damage zones often exhibit a visco-elastic behavior, which is not modeled by the kinematic fault operator, and not accounted in the wavefield simulation or homogenization process performed in this study, which treat the

subsurface as purely elastic.
To improve realism and robustness, the velocity model should also be extended to incorporate lateral velocity variations. Furthermore, the inversion strategy adopted here, which relies on perfect knowledge of a macro-layered velocity and aims to infer its fine-scale perturbations, has inherent limitations. In practice, the macro-layered structure is not known: FWI typically recovers only a smooth approximation, and uncertainty in the interface geometries and layer velocities of this macro-layered model must also be evaluated.

- **Data uncertainty**. The accuracy of the results obtained in the synthetic examples proposed in this thesis can be partly attributed to the strong prior information assumed for the velocity model (macro-layered structure and geostatistical variogram model), and to the data providing substantial information about the fault structure, in particular the fault-induced effective anisotropy provided by the HFWI elastic properties. In real applications, the observed data given by the HFWI solution may be less informative, and this data uncertainty must be integrated into the Bayesian framework. Obtaining a reliable quantification of HFWI uncertainty is itself challenging: uncertainty analysis in FWI solutions obtained via local optimization is is well known to be a difficult task, and the nonlinear nature of the homogenization operator further complicates the problem. These aspects should be explicitly addressed in future developments.

- **Sampling efficiency and convergence assessment**. Under more complex geological or data conditions, a basic Metropolis-Hastings sampling scheme may be insufficient to ensure proper exploration of the posterior distribution and more advanced sampling strategies (e.g., Hamiltonian Monte Carlo (HMC), Sequential Monte Carlo (SMC), Parallel tempering MCMC (PT-MCMC)) could be considered. Alternatively, reversible-jump MCMC (rj-MCMC) could be used to treat as random variables quantities such as the number of faults in the model, the sequence in which they truncate one another, or the number of layers in the macro-layered velocity field (e.g., Herrero et al., 2024) when no prior knowledge on these quantities is available. Future work should also explore more robust convergence assessment tools in high-dimensional problems.

- **Limitations of the homogenization operator**. A key aspect of the non-periodic homogenization is that it currently relies on a fixed (constant all over the domain) minimum wavelength to define the scaling parameter ε_0 separating small and large scales. In realistic subsurface models, however, velocity (and therefore the minimum wavelength) usually increases with depth. As a result, the ε_0 , and therefore the quality of the solution, changes with depth. An "adaptive" homogenization strategy that adjusts

to spatial changes in wavelength should be developed to obtain accurate and physically consistent effective-medium solutions.

Once these issues have been addressed for the 2-D case, extending the methodology to 3-D could be a further step toward practical applicability.

Appendix a

Single-pass randomized SVD algorithm

Algorithm 4 Randomized SVD algorithm

1. $\mathbf{Y} = \mathbf{BX}$
2. $\mathbf{QR} = \mathbf{Y}$
3. $\mathbf{A}(\mathbf{Q^T X}) = \mathbf{Q^T Y}$
4. $\mathbf{A} = \mathbf{U\Lambda U^T}$
5. $\mathbf{V} = \mathbf{QU}$
6. $\mathbf{B} = \mathbf{V\Lambda V^T}$

The single-pass randomized SVD proposed in Halko et al. (2011), presented in Algorithm 4, is applied to estimate the singular values and vectors of an input symmetric matrix $\mathbf{B}$ of size $M \times M$. In the first step of the algorithm the matrix $\mathbf{B}$ is multiplied to a $M \times r$ random matrix $\mathbf{X}$ where r has the same order as the rank of the matrix $\mathbf{B}$. This produces a matrix $\mathbf{Y}$, of the same size as $\mathbf{X}$, that captures the main column space of $\mathbf{B}$. As a second step, a $\mathbf{QR}$ decomposition of $\mathbf{Y}$ is performed, where $\mathbf{Q}$ is an orthonormal basis for $\mathbf{Y}$. Next, an approximation to $\mathbf{B}$ in the basis $\mathbf{Q}$ is computed as

$$\mathbf{A} = \mathbf{Q}^T\mathbf{B}\mathbf{Q}. \tag{a.1}$$

Multiplying both sides of the above equation with $\mathbf{Q}^T\mathbf{X}$ yields

$$\mathbf{A}\mathbf{Q}^T\mathbf{X} = \mathbf{Q}^T\mathbf{B}\mathbf{Q}\mathbf{Q}^T\mathbf{X}, \tag{a.2}$$

and knowing that $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ and $\mathbf{Y} = \mathbf{BX}$ we derive

$$\mathbf{A} = (\mathbf{Q}^T\mathbf{Y})(\mathbf{Q}^T\mathbf{X})^{-1}, \tag{a.3}$$

which is the third line in Alg. 4. After solving for $\mathbf{A}$, we can decompose the matrix as:

$$\mathbf{A} = \mathbf{U\Lambda U}^T. \tag{a.4}$$

The dominant vectors of $\mathbf{B}$ are then returned as $\mathbf{V} = \mathbf{QU}$, with eigenvalues on the diagonal of $\mathbf{\Lambda}$. The algorithm finally outputs the desired decomposition of the matrix $\mathbf{B}$:

$$\mathbf{B} \approx \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T. \tag{a.5}$$

Appendix b

Turning Band Simulation

Turning bands simulation is a commonly applied technique in geostatistics and it is often preferred to other simulation methods because the computational effort related to this method is proportional to a 1-D process (Tompson et al., 1989). In fact, the key idea behind the algorithm is that of simulating values along a set of directions, typically radial directions, and then combining these simulated values to create a spatially correlated field. Consequently, the computational amount of the process is comparable to a 1-D simulation process. In the following, we briefly introduce the main steps of the TB algorithm:

1. Definition of the domain where random functions are generated;
2. Selection of a straight line (turning band) i, passing through the domain, accordingly to a fixed direction u_i;
3. Generation of a one-dimensional random function, $\mathbf{Z}_i(\xi_i)$, on the line i with fixed mean and variance;
4. Orthogonal projection of each grid point, x_k of the domain to the line i, $\xi_{ik} = x_k * u_i$, and estimation of a random function $\mathbf{Z}_i(\xi_{ik})$;
5. Re-composition of the components $\mathbf{Z}_i(\xi_{ik})$ to get back the 2-D random function $\mathbf{Z}(x_k)$;
6. Iteration of steps 2 to 5 onto N different straight lines;
7. Normalization of $\mathbf{Z}(x_k)$ accordingly to the $\sqrt{N}$ factor.

This outline can be summarized with the expression:

$$\mathbf{Z}(x_k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_i(x_k * u_i). \tag{b.1}$$

Appendix c

Empirical and theoretical variograms

The choice of independent perturbations of the residual velocity field at pilot points is made for convenience. However, as mentioned in Section 5.5, this could lead to correlation ranges approximately equal to the spacing between pilot points in the simulated random fields, rather than to the range specified by the covariance model used in the turning bands simulation. We check quantitatively the impact of this simplification by comparing the empirical variogram of conditional simulations from the posterior samples with the theoretical (assumed) model. The two figures below (Fig. c.1) show the theoretical variogram model (solid blue line), with sill= 1 and correlation range= 5, used to generate the random fields with conditional simulations, together with the mean empirical variogram (solid orange line) computed from multiple simulations (grey dotted lines) conditioned to the pilot points values. To check the empirical variograms, we have first generated independent perturbations at the pilot points for 100 000 iterations where all perturbations are accepted because no MCMC acceptance-rejection step is applied. In this case (Fig. c.1, top graph), empirical variograms computed from the independently perturbed pilot points show a rapid increase and high variance at short distances, and a correlation range approaching the distance between pilot points (2 meters), thus smaller than the theoretical one. This suggests, as expected, independent or weakly correlated perturbations between pilot points. We have then computed empirical variograms of multiple posterior simulations conditioned by the pilot points values from the last MCMC iteration of the experiment described in Section 5.4. As illustrated (Fig. c.1, bottom graph), differently from the previous case, at short distances the mean variogram from the conditional simulations follows the theoretical model, showing a similar rapid increase and a comparable, but slightly longer correlation range, suggesting that the spatial correlation at small lags is well reproduced by the simulations. Thus, the MCMC acceptance step filters out the spatially incoherent (independent) proposals and keeps realizations that are conform to the spatial structure encoded

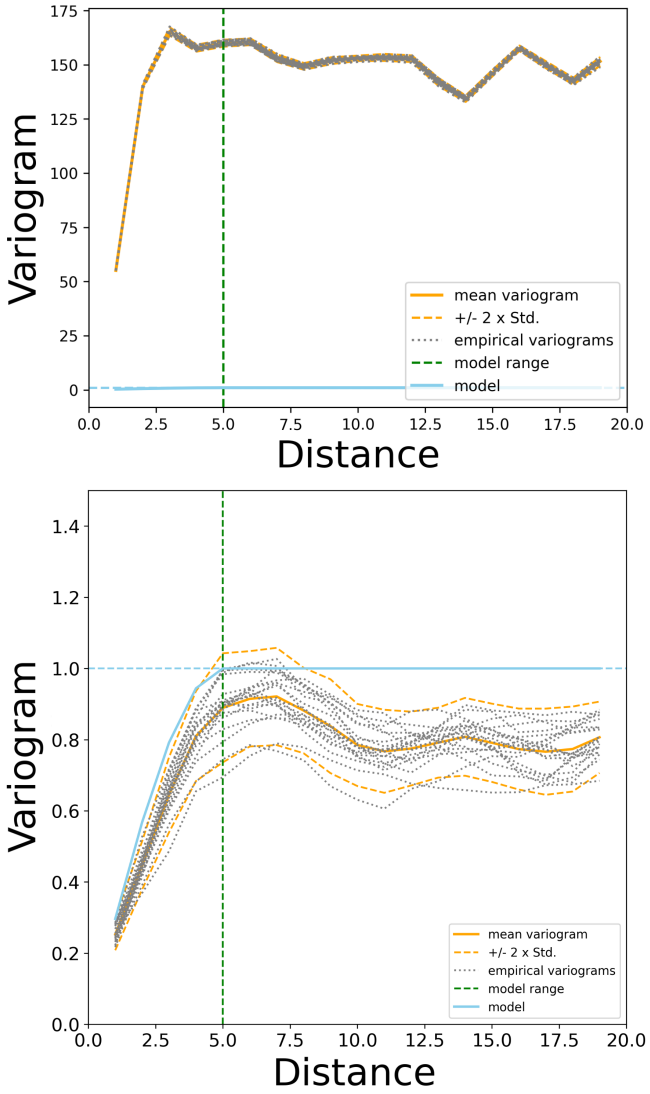

Figure c.1: Theoretical variogram model (solid blue line), with sill= 1 and correlation range= 5, used to generate the random fields with conditional simulations and the mean empirical variogram (solid orange line) computed from multiple simulations (grey dotted lines) conditioned to the pilot points values. Top graph: Empirical variograms of multiple simulations conditioned by the pilot points values independently perturbed. In this case all the perturbations are accepted because no MCMC acceptance-rejection step is applied. Bottom graph: Empirical variograms of multiple posterior simulations conditioned by the pilot points values from the last MCMC iteration of our experiment.

in the model and likelihood, so the accepted ensemble inherits the model's correlation. The amount of this regularization probably depends on the choice of the prior covariance structure, on the cutoff wavelength λ_0 used to homogenize the elastic properties and on the data which tend to "naturally" introduce a spatial dependency between the posterior values between adjacent pilot points.

List of Figures

1 Overview of the methodology. Downscaling is cast into a Bayesian framework where information coming from the observed data, a smooth effective medium derived from HFWI, is combined with prior knowledge on the small-scale elastic and structural properties to retrieve the posterior distribution of all possible small-scale models that fit the data set. 4

2 a) S-wave velocity of the original fine-scale elastic medium. For the seismic wave simulation, a receiver (green triangle) is positioned on the surface, while the seismic source (blue star) is located within the medium. b) Effective S-wave velocity of the homogenized medium. c) Effective anisotropy of the homogenized medium. 6

3 Posterior statistics of the downscaling inversion with homogenized FWI observations: mean and standard deviation model in which the fault structure is derived from posterior samples post-burn-in of 4 MCMC chains. 8

4 Posterior statistics post burn-in: mean, standard deviation and relative error of the small-scale medium $\mathbf{m}(\boldsymbol{\theta})$ obtained from posterior samples of 4 MCMC chains. The fault structure has been successfully recovered and well-constrained, whereas higher uncertainty is observed for some fine-scale velocity layers. 9

5 Vue d'ensemble de la méthodologie. Le downscaling est formulé dans un cadre bayésien où les informations provenant des données observées, c'est-à-dire un milieu effectif lisse issu de la HFWI, sont combinées avec les connaissances a priori sur les propriétés élastiques et structurales à petite échelle afin d'obtenir la probabilité a posteriori de tous les modèles à petite échelle compatibles avec le jeu de données. . . 14

6 a) Vitesse des ondes S du milieu élastique d'origine à fine échelle. Pour la simulation des ondes sismiques, un récepteur (triangle vert) est positionné à la surface, tandis que la source sismique (étoile bleue) est située à l'intérieur du milieu. b) Vitesse effective des ondes S du milieu homogénéisé. c) Anisotropie effective du milieu homogénéisé. 17
7 Statistiques a posteriori de l'inversion par réduction d'échelle avec observations FWI homogénéisées : modèle moyen et écart type dans lequel la faille est dérivée des échantillons a posteriori après la phase de burn-in des 4 chaînes MCMC. 19
8 Statistiques a posteriori après burn-in : moyenne, écart type et erreur relative du milieu à petite échelle obtenu à partir d'échantillons a posteriori de 4 chaînes MCMC. La structure de la faille a été récupérée avec succès et bien contrainte, tandis qu'une incertitude plus élevée est observée pour certaines couches de vitesse à petite échelle. 20

1.1 Schematic diagrams illustrating onshore (a) and offshore (b) seismic acquisition surveys. Black arrows represent seismic waves propagating through the subsurface. From Barnier (2022). 25
1.2 Typical seismic imaging workflow. a) Portion of the Marmousi velocity model (Versteeg, 1994) used for numerical experiments representing the reference (true) velocity model, which is unknown in practice. b) "Observed" data generated from the reference model. c) Initial velocity model obtained by smoothing the reference model. d) Velocity model estimated from the observed data and the initial velocity by waveform inversion. e) Seismic image of the reference model obtained by migrating the observed data using the velocity model in d). After Irakarama et al. (2019). 26
1.3 Different interpretations of the same seismic image reflecting uncertainty in geological interpretation, such as subjectivity in fault placement, or existence. Modified from Bond (2015). 27

1.4 Homogenization principle illustrated on a portion of the Marmousi model. In wavefield simulations, waves are propagated up to a maximum frequency in a complex elastic model (a). The first and classical solution is to mesh all discontinuities of (a) to obtain the complex mesh (b) and then to perform the waveform modeling (1) to obtain the wavefield (e). The second option is to use homogenization to compute the effective elastic model (c) from (a) valid up to the maximum frequency. (c) is fully anisotropic but smooth (without being spatially uniform), which leads to a simple mesh (d). Finally the waveform modeling (2) is performed with the same numerical solver as the one used for waveform modeling (1) to obtain the wavefield (f), but at a lower numerical cost than (e). From Capdeville et al. (2015a). 30

1.5 Overview of the methodology. Downscaling is cast into a Bayesian framework where information coming from the observed data, a smooth effective medium derived from HFWI, is combined with prior knowledge on the small-scale elastic and structural properties to retrieve the posterior distribution of all possible small-scale models that fit the data set. 32

2.1 Forward and Inverse problem. The forward problem maps parameters of the material **m** into observable measures **d**. The inverse problem aims to estimate unknown material parameters **m** from available observations **d**. In this figure the inverse problem seeks to identify the material properties of a subsurface model based on observed seismic data. Modified from Cupillard et al. (2019). 36

2.2 A gradient descent step (black) and a Newton step (red) on the same misfit function (blue). The gradient step moves the point downwards along the linear approximation of the function. The Newton step moves the point to the minimum of the parabola, which is used to locally approximate the misfit function, leading to a faster convergence rate. 39

2.3 Schematic comparison between Steepest descent (local optimization) and Monte Carlo sampling for a simple two-parameter model space. Steepest descent (right plot) traces a single trajectory from an initial solution $\mathbf{m}_0$ toward the nearest minimum. Thus, depending on the initial model, it may converge to a nearby local minimum rather than the global one. Monte Carlo approach (left plot) explores the model space by randomly generating candidate models according to a given proposal distribution. This stochastic exploration allows the algorithm to probe multiple regions of the model space and avoid being trapped in local minima. 46

3.1 a) 1-D periodic homogenization framework: a wavefield having a minimum wavelength λ_m propagates in an infinite 1-D medium made of periodic cells whose size l is much smaller than λ_m. b) 1-D non-periodic homogenization framework: a wavefield having a minimum wavelength λ_m propagates in a finite 1-D medium of size L. Modified from Cupillard & Capdeville (2018). 58

3.2 a) S-wave velocity of the original fine-scale elastic medium. For the seismic wave simulation, a receiver (green triangle) is positioned on the surface, while the seismic source (blue star) is located within the medium. b) Effective S-wave velocity (Eq. 3.36) obtained by homogenizing the original fine-scale medium with $\varepsilon_0 = 0.5$. c) Effective anisotropy (Eq. 3.37) of the homogenized medium. 64

3.3 a) S-wave velocity profile, extracted at x = 1 000 m, from the original fine-scale elastic medium. b) Zoomed-in portion of the S-wave original velocity, indicated by the red box in (a), and S-wave velocity of the effective medium, homogenized using $\varepsilon_0 = 0.2$, $\varepsilon_0 = 0.5$ and $\varepsilon_0 = 0.75$. 65

3.4 Left: x- and z-components of the displacements $\mathbf{u}$ (blue) in the reference medium, and $\mathbf{u}^*$ (orange) in the homogenized medium (using $\varepsilon_0 = 0.5$), recorded at a receiver located on the surface at a $x = 1\,500$ m along the horizontal direction. Right: Difference between $\mathbf{u}$ and $\mathbf{u}^*$. 66

3.5 a) Reference fine-scale medium shown in terms of V_s. Sources (in purple) and receivers (in light blue) are located in two vertical boreholes at $x = 800$ m and $x = 3800$ m, respectively. b) Reference data: shot gather n.10 (x-component) obtained by simulating wavefield propagation in the reference fine-scale medium. c) Initial FWI V_s model obtained by applying Gaussian smoothing to the reference medium. d) Initial data: shot gather n.10 (x-component) obtained by simulating wavefield propagation in the initial smoothed medium. 72
3.6 a) Normalized misfit function for the different frequency bands. b) Final data residual, i.e., difference between reference and predicted waveforms at the last iteration, for a single trace recorded by receiver n. 100 at the center of the borehole. 73
3.7 a) S-wave velocity of the reference model in the illuminated area (between the source and receiver boreholes) b) Result of the elastic FWI for the S-wave velocity. c) Residual model: difference between reference and predicted model. 74
3.8 a) Standard deviation map computed from the diagonal entries of the approximated inverse Hessian matrix for V_S. Larger standard deviation values indicate higher uncertainty areas. Values range between [0, 7] because gradient values have been normalized by the maximum value. b) Standard deviation map superimposed on the FWI inverted V_S in Figure 3.7b. 75
3.9 Effective S-wave velocity (a) and anisotropy (b) of the homogenized FWI solution model with $\varepsilon_0 = 0.5$, derived respectively with equations (3.36) and (3.37). 76
3.10 Zoom on the S-wave velocity profile for the upper 1 000 m, extracted at $x = 1\,000$ m, of the reference fine-scale elastic medium, its effective medium and the homogenized FWI solution. 77

4.1 a) Elliptical fault space. The major-axis of the ellipse defines the fault line, while the minor-axis defines the thickness of the displacement field. b) Parametric profiles defining the displacement inside the elliptical fault space along the two directions. The displacement decreases smoothly from the center (where the displacement is $D_{\max}$) to the boundary of the ellipse (where the displacement vanishes). 84
4.2 a) Reference fine-layered elastic medium. b) Associated macro-layered medium. The elastic medium is illustrated here in terms of S-wave velocity V_s. The black rectangle delineates the region of the medium where downscaling is carried out to infer the target fault structure. 86
4.3 a) Region of the reference fine-layered medium, inside the black rectangle in Figure 4.2, including the target fault structure for which the downscaling inverse problem aims to reduce uncertainty. b) Initial macro-layered medium whose elastic properties (P-wave velocity, S-wave velocity and density) are treated as known in the inversion. The elastic medium is illustrated here in terms of S-wave velocity V_s. 87
4.4 The six independent components of the effective elastic tensor obtained by homogenizing the reference model (Fig.4.3a) with a scale-separation spatial wavelength $\lambda_0 = 200$ m. 88
4.5 The six components of the effective elastic tensor $\mathbf{c}^*_{\mathbf{obs}}$ representing the observation in the synthetic application. These are obtained by homogenizing the reference model (Fig. 4.3a) with a scale-separation spatial wavelength $\lambda_0 = 200$ m (Fig. 4.4) and by adding spatially correlated Gaussian noise $\boldsymbol{\eta}$. 89
4.6 Convergence diagnostics of the downscaling inversion with synthetic HFWI observations. Top graph: Negative log-likelihood functions of 4 parallel MCMC chains for number of iterations. The burn-in phase is considered to finish after 4 000 iterations (indicated by the red dotted line) when the functions reach a plateau and start sampling the posterior. Bottom graph: Acceptance rate of 4 parallel MCMC chains for number of iterations. As the number of iterations increases, the acceptance rates stabilize around 5 %. . 90

4.7 Inversion results of the downscaling inversion with synthetic HFWI observations: posterior densities of the fault-related parameters $\boldsymbol{\vartheta}$ from 4 MCMC chains (shown in different shades of blue) after burn-in. Black dotted lines indicate target values from the reference fault structure in Figure 4.3a. The minimum and maximum values of the uniform priors for each model parameter are indicated in bold. 91
4.8 Fault velocity models of the first downscaling application with synthetic HFWI observations: the 4 starting models, in which the fault structure is generated using fault-related variables randomly sampled from uniform prior distributions, and the mean model derived from post-burn-in posterior samples across the 4 MCMC chains. 92
4.9 a) Result of the elastic FWI for the S-wave velocity. b) Result of the elastic FWI for the P-wave velocity. The black rectangle delineates the region of the medium where the downscaling inverse problem is performed to characterize the target fault structure. The sources and receivers used in the synthetic FWI experiment are positioned in two vertical boreholes located at $x = 300$ m and $x = 3800$ m, respectively. . 93
4.10 First component of the effective elastic tensor $\mathbf{c}^*_{\mathbf{obs}}$, derived by homogenizing, with a scale-separation spatial wavelength $\lambda_0 = 40$ m, the elastic properties of the FWI solution in Figure 4.9. This represents the observations in the second application of the downscaling inversion. 94
4.11 Convergence diagnostics of the downscaling inversion with homogenized FWI observations. Top graph: Negative log-likelihood functions of 4 parallel MCMC chains for number of iterations. The burn-in phase is considered to finish after 1 000 iterations (indicated by the red dotted line) when the functions reach a plateau and start sampling the posterior. Bottom graph: Acceptance rate of 4 parallel MCMC chains for number of iterations. As the number of iterations increases, the acceptance rates stabilize around 47 %. 95

4.12 Inversion results of the downscaling inversion with homogenized FWI observations: posterior densities of the fault-related parameters $\boldsymbol{\vartheta}$ from 4 MCMC chains (shown in different shades of blue) after burn-in. Black dotted lines indicate target values from the reference fault structure in Figure 4.3a. The minimum and maximum values of the uniform priors for each model parameter are indicated in bold. 96
4.13 Posterior statistics of the downscaling inversion with homogenized FWI observations: mean and standard deviation model in which the fault structure is derived from posterior samples post-burn-in of 4 MCMC chains. 97

5.1 a) 1-D macro-layered velocity field, obtained by defining a constant velocity in each macro layer. b) Fine-layered velocity field, generated as described in Eq. (5.1). c) Corresponding vertical velocity profiles. . . . 102
5.2 a) Reference 1-D velocity model $\mathbf{V}_p$ and structural model $\boldsymbol{\vartheta}$. b) Fine-scale elastic medium $\mathbf{m}(\boldsymbol{\theta})$ used as reference medium for the synthetic application. The elastic medium is illustrated here in terms of P-wave velocity V_p. The elastic properties V_s and ρ, required for the elastic homogenization, are computed from V_p as explained in Section 5.4. c) First component of the effective elastic tensor $\mathbf{c}^* = \mathcal{H}(\mathbf{m}(\boldsymbol{\theta}))$ obtained by homogenizing the reference fine-scale medium with $\lambda_0 = 20$ m. 105
5.3 The six components of the effective elastic tensor $\mathbf{c}^*_{\mathbf{obs}}$ representing the observation in the synthetic application. These are obtained by homogenizing the reference medium with a scale-separation spatial wavelength $\lambda_0 = 20$ m and by adding spatially correlated Gaussian noise $\boldsymbol{\eta}$. 105
5.4 a) Top graph: Negative log-likelihood functions of 4 parallel MCMC chains for number of iterations. The burn-in phase is considered to finish after 50 000 iterations when the functions reach a plateau and start sampling the posterior. Bottom graph: Acceptance rate of 4 parallel MCMC chains for number of iterations. As the number of iterations increases, the acceptance rates stabilize around 10 %. b) Zoomed-in view of the two graphs post burn-in. 108

5.5 Inversion results of 4 MCMC chains. a) V_p velocity profiles extracted at x = 50 m. The blue line is the reference fine-scale velocity profile. The black and red lines show the mean of the prior and posterior distributions. Red-shaded region marks the 90 % confidence interval (5th and 95th percentile range), capturing the most probable range of velocity values. b) Posterior densities of the fault-related parameters $\boldsymbol{\vartheta}$ from 4 MCMC chains (shown in different shades of blue) after burn-in. Black dotted lines indicate target values from the reference fault structure in Figure 5.2. The minimum and maximum values of the uniform priors for each model parameter are indicated in bold. 109

5.6 a) Starting models for different chains. The initial fault parameters $\boldsymbol{\vartheta}$ are randomly sampled from the uniform prior distributions, whereas the initial velocity $\mathbf{V}_p$ corresponds to the macro-scale velocity field. b) Example of posterior models for different chains. These models show a fault that matches the target structure and a fine-layering velocity $\mathbf{V}_p$ that closely resembles the reference one (Fig. 5.2a). 110

5.7 Posterior statistics post burn-in: mean, standard deviation and relative error of the small-scale elastic medium $\mathbf{m}(\boldsymbol{\theta})$ obtained from posterior samples of 4 MCMC chains. The fault structure has been successfully recovered and well-constrained, whereas higher uncertainty is observed for some fine-scale velocity layers. 111

c.1 Theoretical variogram model (solid blue line), with sill= 1 and correlation range= 5, used to generate the random fields with conditional simulations and the mean empirical variogram (solid orange line) computed from multiple simulations (grey dotted lines) conditioned to the pilot points values. Top graph: Empirical variograms of multiple simulations conditioned by the pilot points values independently perturbed. In this case all the perturbations are accepted because no MCMC acceptance-rejection step is applied. Bottom graph: Empirical variograms of multiple posterior simulations conditioned by the pilot points values from the last MCMC iteration of our experiment. 122

Bibliography

Aarnes, I., Skorstad, A., & Kallekleiv, H., 2014. Acknowledging structural uncertainties in the reservoir models–a consistent approach to improve subsurface predictions, *Production Geoscience, Stavanger, Norway*, pp. 4–5. 28

Afanasiev, M., Boehm, C., May, D., & Fichtner, A., 2016. Using effective medium theory to better constrain full waveform inversion, in *78th EAGE Conference and Exhibition 2016*, vol. 2016, pp. 1–3, European Association of Geoscientists & Engineers. 29, 54

Aissat, R., Pryet, A., Saltel, M., & Dupuy, A., 2023. Comparison of different pilot point parameterization strategies when measurements are unevenly distributed in space, *Hydrogeology Journal*, **31**(8), 2381–2400. 112

Aki, K. & Lee, W., 1976. Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. a homogeneous initial model, *Journal of Geophysical research*, **81**(23), 4381–4399. 34

Aki, K., Christoffersson, A., & Husebye, E. S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *Journal of Geophysical Research*, **82**(2), 277–296. 34

Alcalde, J., Bond, C. E., Johnson, G., Ellis, J. F., & Butler, R. W., 2017. Impact of seismic image quality on fault interpretation uncertainty, *GSA Today*. 27

Alcolea, A., Carrera, J., & Medina, A., 2006. Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Advances in water resources*, **29**(11), 1678–1689. 112

Alder, C., Bodin, T., Ricard, Y., Capdeville, Y., Debayle, E., & Montagner, J., 2017. Quantifying seismic anisotropy induced by small-scale chemical heterogeneities, *Geophysical Journal International*, **211**(3), 1585–1600. 54

Alford, R., Kelly, K., & Boore, D. M., 1974. Accuracy of finite-difference modeling of the acoustic wave equation, *Geophysics*, **39**(6), 834–842. 72

Allaire, G., 1992. Homogenization and two-scale convergence, *SIAM Journal on Mathematical Analysis*, **23**(6), 1482–1518. 56

Aster, R. C., Borchers, B., & Thurber, C. H., 2018. *Parameter estimation and inverse problems*, Elsevier. 23

Aydin, O. & Caers, J. K., 2017. Quantifying structural uncertainty on fault networks using a marked point process within a Bayesian framework, *Tectonophysics*, **712**, 101–124. 97

Babuška, I., 1976. Homogenization and its application. Mathematical and computational problems, in *Numerical solution of partial differential equations–III*, pp. 89–116, Elsevier. 54

Bacigalupo, A. & Gambarotta, L., 2014. Second-gradient homogenized model for wave propagation in heterogeneous periodic media, *International Journal of Solids and Structures*, **51**(5), 1052–1065. 54

Backus, G. & Gilbert, F., 1968. The resolving power of gross earth data, *Geophysical Journal International*, **16**(2), 169–205. 34

Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **266**(1173), 123–192. 34

Backus, G. E., 1962. Long-wave elastic anisotropy produced by horizontal layering, *Journal of Geophysical Research*, **67**(11), 4427–4440. 53, 64

Backus, G. E. & Gilbert, J. F., 1967. Numerical applications of a formalism for geophysical inverse problems, *Geophysical Journal International*, **13**(1-3), 247–276. 23, 34

Bárbara, C. P., Cabello, P., Bouche, A., Aarnes, I., Gordillo, C., Ferrer, O., Roma, M., & Arbués, P., 2019. Quantifying the impact of the structural uncertainty on the gross rock volume in the Lubina and Montanazo oil fields (Western Mediterranean), *Solid Earth*, **10**(5), 1597–1619. 28

Barnier, G. C. M., 2022. *Full waveform inversion by model extension: A robust method to estimate the seismic propagation velocity in the subsurface from recorded seismograms*, Ph.D. thesis, Stanford University. 25, 124

Baysal, E., Kosloff, D. D., & Sherwood, J. W., 1983. Reverse time migration, *Geophysics*, **48**(11), 1514–1524. 24

Bensoussan, A., Lions, J.-L., & Papanicolaou, G., 1978. *Asymptotic analysis for periodic structures*, North-Holland. 54

Besag, J., 2001. Markov chain Monte Carlo for statistical inference, *Center for Statistics and the Social Sciences*, **9**(24-25), 118. 52

Bocquet, M., Pires, C. A., & Wu, L., 2010. Beyond Gaussian statistical modeling in geophysical data assimilation, *Monthly Weather Review*, **138**(8), 2997–3023. 48

Bodin, T., Capdeville, Y., Romanowicz, B., & Montagner, J.-P., 2015. Interpreting radial anisotropy in global and regional tomographic models, *The Earth's Heterogeneous Mantle: A Geophysical, Geodynamical, and Geochemical Perspective*, pp. 105–144. 1, 2, 11, 12, 22, 29, 54

Bond, C. E., 2015. Uncertainty in structural interpretation: Lessons to be learnt, *Journal of Structural Geology*, **74**, 185–200. 27, 124

Bond, C. E., Gibbs, A. D., Shipton, Z. K., Jones, S., et al., 2007. What do you think this is "Conceptual uncertainty" in geoscience interpretation, *GSA today*, **17**(11), 4. 27

Bousserez, N., Henze, D. K., Perkins, A., Bowman, K. W., Lee, M., Liu, J., Deng, F., & Jones, D. B. A., 2015. Improved analysis-error covariance matrix for high-dimensional variational inversions: application to source estimation using a 3D atmospheric transport model, *Quarterly Journal of the Royal Meteorological Society*, **141**(690). 69

Boutin, C. & Auriault, J., 1993. Rayleigh scattering in elastic composite materials, *International journal of engineering science*, **31**(12), 1669–1689. 54

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N., & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, **207**(3), 1739–1766. 1, 11, 22

Brocher, T. M., 2005. Empirical relations between elastic wavespeeds and density in the earth's crust, *Bulletin of the seismological Society of America*, **95**(6), 2081–2092. 110

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L., 2011. *Handbook of Markov chain Monte Carlo*, CRC press. 29

Brooks, S. P. & Roberts, G. O., 1998. Convergence assessment techniques for Markov chain Monte Carlo, *Statistics and Computing*, **8**(4), 319–335. 51

Browaeys, J. T. & Chevrot, S., 2004. Decomposition of the elastic tensor and geophysical applications, *Geophysical Journal International*, **159**(2), 667–678. 63

Broyden, C. G., 1970. The convergence of a Class of Double-rank Minimization Algorithms 1. general Considerations, *IMA Journal of Applied Mathematics*, **6**(1), 76–90. 41

Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., & Wilcox, L. C., 2012. Extreme-scale UQ for Bayesian inverse problems governed by PDEs, in *2012 International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–11, IEEE, Salt Lake City, UT. 45, 69, 70

Burgers, G., Jan van Leeuwen, P., & Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter, *Monthly weather review*, **126**(6), 1719–1724. 47

Caers, J., 2011. *Modeling uncertainty in the earth sciences*, John Wiley & Sons. 28

Capdeville, Y. & Marigo, J.-J., 2007. Second order homogenization of the elastic wave equation for non-periodic layered media, *Geophysical Journal International*, **170**(2), 823–838. 29, 54

Capdeville, Y. & Marigo, J.-J., 2008. Shallow layer correction for spectral element like methods, *Geophysical Journal International*, **172**(3), 1135–1150. 63

Capdeville, Y. & Marigo, J.-J., 2013. A non-periodic two scale asymptotic method to take account of rough topographies for 2-D elastic wave propagation, *Geophysical Journal International*, **192**(1), 163–189. 63, 79

Capdeville, Y. & Métivier, L., 2018. Elastic full waveform inversion based on the homogenization method: theoretical framework and 2-D numerical illustrations, *Geophysical Journal International*, **213**(2), 1093–1112. 2, 4, 5, 12, 15, 29, 31, 54, 55, 71, 80, 98

Capdeville, Y., Guillot, L., & Marigo, J.-J., 2010a. 1-D non-periodic homogenization for the seismic wave equation, *Geophysical Journal International*, **181**(2), 897–910. 57, 60

Capdeville, Y., Guillot, L., & Marigo, J.-J., 2010b. 2-D non-periodic homogenization to upscale elastic media for P–SV waves, *Geophysical Journal International*, **182**(2), 903–922. 29, 53, 54, 57, 62, 63, 65

Capdeville, Y., Guillot, L., & Marigo, J.-J., 2010c. 1-D non-periodic homogenization for the seismic wave equation, *Geophysical Journal International*. 54, 57, 58, 59

Capdeville, Y., Stutzmann, E., Wang, N., & Montagner, J.-P., 2013. Residual homogenization for seismic forward and inverse problems in layered media, *Geophysical Journal International*, **194**(1), 470–487. 29, 30, 54

Capdeville, Y., Zhao, M., & Cupillard, P., 2015a. Fast Fourier homogenization for elastic wave propagation in complex media, *Wave Motion*, **54**, 170–186. 30, 125

Capdeville, Y., Zhao, M., & Cupillard, P., 2015b. Fast Fourier homogenization for elastic wave propagation in complex media, *Wave Motion*, **54**, 170–186. 54, 63, 67, 79

Capdeville, Y., Cupillard, P., & Singh, S., 2020. An introduction to the two-scale homogenization method for seismology, in *Advances in Geophysics*, vol. 61, pp. 217–306, Elsevier. 55, 61, 63, 64

Carr, J. R., 2002. *Data visualization in the geosciences*, Prentice-Hall Upper Saddle River, NJ. 100

Castagna, J. P., Batzle, M. L., & Eastwood, R. L., 1985. Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks, *Geophysics*, **50**(4), 571–581. 110

Cauchy, A. et al., 1847. Méthode générale pour la résolution des systemes d'équations simultanées, *Comp. Rend. Sci. Paris*, **25**(1847), 536–538. 41

Caumon, G., 2010. Towards stochastic time-varying geological modeling, *Mathematical Geosciences*, **42**(5), 555–569. 24

Caumon, G., Collon-Drouaillet, P., Le Carlier de Veslud, C., Viseur, S., & Sausse, J., 2009. Surface-based 3D modeling of geological structures, *Mathematical geosciences*, **41**(8), 927–945. 24

Chaljub, E., Komatitsch, D., Vilotte, J.-P., Capdeville, Y., Valette, B., & Festa, G., 2007. Spectral-element analysis in seismology, *Advances in geophysics*, **48**, 365–419. 66

Cherpeau, N. & Caumon, G., 2015. Stochastic structural modelling in sparse data situations, *Petroleum Geoscience*, **21**(4), 233–247. 97

Cherpeau, N., Caumon, G., & Lévy, B., 2010. Stochastic simulations of fault networks in 3D structural modeling, *Comptes Rendus. Géoscience*, **342**(9), 687–694. 96

Cherpeau, N., Caumon, G., Caers, J., & Lévy, B., 2012. Method for stochastic inverse modeling of fault geometry and connectivity using flow data, *Mathematical Geosciences*, **44**(2), 147–168. 28

Claerbout, J. F., 1971. Toward a unified theory of reflector mapping, *Geophysics*, **36**(3), 467–481. 24

Claerbout, J. F. & Muir, F., 1973. Robust modeling with erratic data, *Geophysics*, **38**(5), 826–844. 38

Colletta, B., Letouzey, J., Pinedo, R., Ballard, J., & Balé, P., 1991. Computerized X-ray tomography analysis of sandbox models: examples of thin-skinned thrust systems, *Geology*, **19**, 1063–1067. 63

Colli, L., Fichtner, A., & Bunge, H.-P., 2013. Full waveform tomography of the upper mantle in the South Atlantic region: Imaging a westward fluxing shallow asthenosphere?, *Tectonophysics*, **604**, 26–40. 1, 11, 22

Cupillard, P. & Capdeville, Y., 2018. Non-periodic homogenization of 3-D elastic media for the seismic wave equation, *Geophysical Journal International*, **213**(2), 983–1001. 54, 58, 63, 67, 126

Cupillard, P., Mulder, W. A., Irakarama, M., Mazuyer, A., & Anquez, P., 2019. Small-scale-induced anisotropy of a 3D subsurface model: quantitative analysis and numerical simulations of waves within, in *AGU Fall Meeting Abstracts*, vol. 2019, pp. S41A–07. 29, 36, 125

Curtis, A. & Lomax, A., 2001. Prior information, sampling distributions, and the curse of dimensionality, *Geophysics*, **66**(2), 372–378. 98

Dahlke, T., Biondi, B., & Clapp, R., 2020. Applied 3D salt body reconstruction using shape optimization with level sets, *Geophysics*, **85**(5), R437–R446. 28

de la Varga, M. & Wellmann, J. F., 2016. Structural geologic modeling as an inference problem: A Bayesian perspective, *Interpretation*, **4**(3), SM1–SM16. 28

de Marsily, G., Lavedan, G., Boucher, M., & Fasamino, G., 1984. Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for natural resources characterization. NATO advanced Study Institute*, pp. 831–849. 102

Deutsch, C. V., Journel, A. G., et al., 1992. Geostatistical software library and user's guide, *New York*, **119**(147), 578. 100

Dovera, L. & Della Rossa, E., 2011. Multimodal ensemble Kalman filtering using Gaussian mixture models, *Computational Geosciences*, **15**(2), 307–323. 48

Dramsch, J. S. & Lüthje, M., 2018. Deep-learning seismic facies on state-of-the-art CNN architectures, in *SEG International Exposition and Annual Meeting*, pp. SEG–2018, SEG. 24

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D., 1987. Hybrid Monte Carlo, *Physics letters B*, **195**(2), 216–222. 47, 52

Dziewonski, A. M., Hager, B. H., & O'Connell, R. J., 1977. Large-scale heterogeneities in the lower mantle, *Journal of Geophysical Research*, **82**(2), 239–255. 34

Engl, H. W. & Ramlau, R., 2015. Regularization of inverse problems, in *Encyclopedia of applied and computational mathematics*, pp. 1233–1241, Springer. 38

Engquist, B. & Souganidis, P. E., 2008. Asymptotic and numerical homogenization, *Acta Numerica*, **17**, 147–190. 53

Evensen, G., 2009. *Data assimilation: the ensemble Kalman filter*, Springer. 47

Evensen, G. & Van Leeuwen, P. J., 2000. An ensemble Kalman smoother for nonlinear dynamics, *Monthly Weather Review*, **128**(6), 1852–1867. 47

Faccenda, M., Ferreira, A. M., Tisato, N., Lithgow-Bertelloni, C., Stixrude, L., & Pennacchioni, G., 2019. Extrinsic elastic anisotropy in a compositionally heterogeneous earth's mantle, *Journal of Geophysical Research: Solid Earth*, **124**(2), 1671–1687. 54

Fernandes, F. J. D., Neto, E. R. O., Teixeira, L., & Menezes, A. F., 2023. Seismic impedance inversion using closed-loop convolutional neural network: An application in turbidite reservoirs, in *18th International Congress of the Brazilian Geophysical Society*, SBGf. 28

Fernández-Martínez, J. L., Fernandez Muniz, M. Z., & Tompkins, M. J., 2012. On the topography of the cost functional in linear and nonlinear inverse problems, *Geophysics*, **77**(1), W1–W15. 46

Fernández-Martínez, J. L., Fernández-Muñiz, Z., Pallero, J., & Pedruelo-González, L. M., 2013. From Bayes to Tarantola: new insights to understand uncertainty in inverse problems, *Journal of Applied Geophysics*, **98**, 62–72. 35

Festa, G. & Vilotte, J.-P., 2005. The newmark scheme as velocity–stress time-staggering: an efficient pml implementation for spectral element simulations of elastodynamics, *Geophysical Journal International*, **161**(3), 789–812. 73

Fichtner, A. & Hanasoge, S. M., 2017. Discrete wave equation upscaling, *Geophysical Journal International*, **209**(1), 353–357. 54

Fichtner, A. & Leeuwen, T. v., 2015. Resolution analysis by random probing, *Journal of Geophysical Research: Solid Earth*, **120**(8), 5549–5573. 69

Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophysical Journal International*, **187**(3), 1604–1624. 69, 70

Fichtner, A., Kennett, B. L., & Trampert, J., 2013a. Separating intrinsic and apparent anisotropy, *Physics of the Earth and Planetary Interiors*, **219**, 11–20. 54

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., & Villasenor, A., 2013b. Multiscale full waveform inversion, *Geophysical Journal International*, **194**(1), 534–556. 29

Fish, J. & Chen, W., 2001. Higher-order homogenization of initial/boundary-value problem, *Journal of engineering mechanics*, **127**(12), 1223–1230. 54

Fish, J. & Chen, W., 2004. Space–time multiscale model for wave propagation in heterogeneous media, *Computer Methods in applied mechanics and engineering*, **193**(45-47), 4837–4856. 54

Fletcher, R., 1970. A new approach to variable metric algorithms, *The computer journal*, **13**(3), 317–322. 41

Fletcher, R., 2000. *Practical methods of optimization*, John Wiley & Sons. 38

Fletcher, R. & Reeves, C. M., 1964. Function minimization by conjugate gradients, *The computer journal*, **7**(2), 149–154. 41

French, S., Lekic, V., & Romanowicz, B., 2013. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere, *Science*, **342**(6155), 227–230. 1, 11, 22

Freulon, X. & de Fouquet, C., 1993. Conditioning a gaussian model with inequalities, *Geostatistics Tróia'92: Volume 1*, pp. 201–212. 112

Gardner, G., Gardner, L., & Gregory, A., 1974. Formation velocity and density—The diagnostic basics for stratigraphic traps, *Geophysics*, **39**(6), 770–780. 86

Gejadze, I. Y., Shutyaevb, V., & Dimetc, F.-X. L., 2013. Analysis error covariance versus posterior covariance in variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, **139**(676), 1826–1841. 69

Gelman, A. & Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences, *Statistical science*, **7**(4), 457–472. 52, 88

Geman, S. & Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721–741. 47

Gercek, H., 2007. Poisson's ratio values for rocks, *International Journal of Rock Mechanics and Mining Sciences*, **44**(1), 1–13. 110

Gholami, A., 2015. Nonlinear multichannel impedance inversion by total-variation regularization, *Geophysics*, **80**(5), R217–R224. 38

Gilks, W. R., Richardson, S., & Spiegelhalter, D., 1995. *Markov chain Monte Carlo in practice*, CRC press. 29

Godefroy, G., Caumon, G., Ford, M., Laurent, G., & Jackson, C. A.-L., 2018. A parametric fault displacement model to introduce kinematic control into modeling faults from sparse data, *Interpretation*, **6**(2), B1–B13. 7, 16, 31, 82, 83, 96

Gold, N., Shapiro, S., Bojinski, S., & Müller, T., 2000. An approach to upscaling for seismic waves in statistically isotropic heterogeneous elastic media, *Geophysics*, **65**(6), 1837–1850. 53

Goldfarb, D., 1970. A family of variable-metric methods derived by variational means, *Mathematics of computation*, **24**(109), 23–26. 41

Gómez-Hernánez, J. J., Sahuquillo, A., & Capilla, J., 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—I. theory, *Journal of Hydrology*, **203**(1-4), 162–174. 112

Goovaerts, P., 1997. *Geostatistics for natural resources evaluation*, Oxford university press. 31, 100

Grana, D., Mukerji, T., & Doyen, P., 2021. *Seismic reservoir modeling: Theory, examples, and algorithms*, John Wiley & Sons. 100

Grana, D., de Figueiredo, L., & Mosegaard, K., 2023. Markov chain Monte Carlo for seismic facies classification, *Geophysics*, **88**(3), M131–M143. 100

Grechka, V., 2003. Effective media: A forward modeling view, *Geophysics*, **68**(6), 2055–2062. 53

Green, P., 1995. Reversible jump MCMC computation and Bayesian model determination, *Biometrika*, **57**, 97–109. 47

Grenander, U. & Miller, M. I., 1994. Representations of knowledge in complex systems, *Journal of the Royal Statistical Society: Series B (Methodological)*, **56**(4), 549–581. 52

gstlearn, 2023. Geostatistics and machine learning toolbox, Free download from: https://gstlearn.org. 104

Guillot, L., Capdeville, Y., & Marigo, J.-J., 2010. 2-D non-periodic homogenization of the elastic wave equation: SH case, *Geophysical Journal International*, **182**(3), 1438–1454. 53, 54, 62

Haario, H., Saksman, E., & Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm, *Computational statistics*, **14**(3), 375–395. 51

Haario, H., Saksman, E., & Tamminen, J., 2001. An adaptive metropolis algorithm, *Bernoulli*, **7**(2), 223–242. 51

Haario, H., Laine, M., Mira, A., & Saksman, E., 2006. DRAM: efficient adaptive MCMC, *Statistics and computing*, **16**(4), 339–354. 51

Hadamard, J., 1902. Sur les problèmes aux dérivées partielles et leur signification physique, *Princeton university bulletin*, pp. 49–52. 34, 37

Halko, N., Martinsson, P.-G., & Tropp, J. A., 2011. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM review*, **53**(2), 217–288. 46, 70, 118

Hansen, T. M., Journel, A. G., Tarantola, A., & Mosegaard, K., 2006. Linear inverse Gaussian theory and geostatistics, *Geophysics*, **71**(6), R101–R111. 100

Hedjazian, N., Capdeville, Y., & Bodin, T., 2021. Multiscale seismic imaging with inverse homogenization, *Geophysical Journal International*, **226**(1), 676–691. 2, 12, 29, 30, 31, 54, 71, 81, 98

Herrero, J., Caumon, G., & Bodin, T., 2024. Transdimensional reservoir data inversion for a 2D geological layered model with horizon slopes and horizontal permeability trends, in *International Geostatistics Congress*, pp. 105–119, Springer. 116

Hestenes, M. R., Stiefel, E., et al., 1952. Methods of conjugate gradients for solving linear systems, *Journal of research of the National Bureau of Standards*, **49**(6), 409–436. 41

Hofmann, B., 2013. *Regularization for applied inverse and ill-posed problems: a numerical approach*, vol. 85, Springer-Verlag. 38

Holden, L., Mostad, P., Nielsen, B. F., Gjerde, J., Townsend, C., & Ottesen, S., 2003. Stochastic structural modeling, *Mathematical geology*, **35**, 899–914. 97

Hollund, K., Mostad, P., Nielsen, B. F., Holden, L., Gjerde, J., Contursi, M. G., McCann, A. J., Townsend, C., & Sverdrup, E., 2002. Havana—a fault modeling tool, in *Norwegian petroleum society special publications*, vol. 11, pp. 157–171, Elsevier. 28

Hoteit, I., Luo, X., & Pham, D.-T., 2012. Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters, *Monthly weather review*, **140**(2), 528–542. 48

Howley, E. & Meyer, R. S., 2015. Developing an integrated structural modelling workflow, *First Break*, **33**(3). 28

Huang, L., Dong, X., & Clee, T. E., 2017. A scalable deep learning platform for identifying geologic features from seismic attributes, *The Leading Edge*, **36**(3), 249–256. 24

Huang, R., Zhang, Z., Wu, Z., Wei, Z., Mei, J., & Wang, P., 2021. Full-waveform inversion for full-wavefield imaging: Decades in the making, *The Leading Edge*, **40**(5), 324–334. 24

Irakarama, M., Cupillard, P., Caumon, G., Sava, P., & Edwards, J., 2019. Appraising structural interpretations using seismic data — Theoretical elements, *Geophysics*, **84**(2), N29–N40. 26, 27, 63, 86, 124

Jeong, W., Pyun, S., Son, W., & Min, D.-J., 2013. A numerical study of simultaneous-source full waveform inversion with l 1-norm, *Geophysical Journal International*, **194**(3), 1727–1737. 38

Jessell, M., 1981. Noddy: an interactive map creation package. Unpublished MSc Thesis, *University of London*. 82

Jessell, M., Aillères, L., De Kemp, E., Lindsay, M., Wellmann, F., Hillier, M., Laurent, G., Carmichael, T., & Martin, R., 2014. Next generation Three-dimensional geologic modeling and inversion, *Building Exploration Capability for the 21st Century*, **18**. 24

Jessell, M. W., Ailleres, L., & De Kemp, E. A., 2010. Towards an integrated inversion of geoscientific data: What price of geology?, *Tectonophysics*, **490**(3-4), 294–306. 28

Jordan, T. H., 2015. An effective medium theory for three-dimensional elastic heterogeneities, *Geophysical Journal International*, **203**(2), 1343–1354. 54

Kadu, A., Van Leeuwen, T., & Mulder, W., 2017. Parametric level-set full-waveform inversion in the presence of salt bodies, in *SEG Technical Program Expanded Abstracts 2017*, pp. 1518–1522, Society of Exploration Geophysicists. 28

Kaipio, J. P. & Somersalo, E., 2005. *Statistical and computational inverse problems*, Springer. 48

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems, *Trans. ASME, D*, **82**, 35–44. 47

Kirkpatrick, S., Gelatt Jr, C. D., & Vecchi, M. P., 1983. Optimization by simulated annealing, *Science*, **220**(4598), 671–680. 52

Köhn, D., 2011. *Time domain 2D elastic full waveform tomography*, Ph.D. thesis, Kiel. 72

Kolbjørnsen, O., Buland, A., Hauge, R., Røe, P., Ndingwan, A. O., & Aker, E., 2020. Bayesian seismic inversion for stratigraphic horizon, lithology, and fluid prediction, *Geophysics*, **85**(3), R207–R221. 28

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophysical Journal International*, **139**(3), 806–822. 66, 110

Komatitsch, D. & Vilotte, J.-P., 1998. The spectral element method: an efficient tool to simulate the seismic response of 2d and 3d geological structures, *Bulletin of the seismological society of America*, **88**(2), 368–392. 66

Leahy, G. M. & Skorstad, A., 2013. Uncertainty in subsurface interpretation: a new workflow, *First Break*, **31**(9). 27

Lecour, M., Cognot, R., Duvinage, I., Thore, P., & Dulac, J.-C., 2001. Modelling of stochastic faults and fault networks in a structural uncertainty study, *Petroleum Geoscience*, **7**(S), S31–S42. 28

Li, Y., Alkhalifah, T., & Guo, Q., 2021. Target-oriented time-lapse waveform inversion using deep learning-assisted regularization, *Geophysics*, **86**(4), R485–R495. 28

Lines, L. R. & Newrick, R. T., 2004. *Fundamentals of geophysical interpretation*, Society of Exploration Geophysicists. 23

Liu, Q. & Peter, D., 2019. Square-root variable metric based elastic full-waveform inversion—part 2: Uncertainty estimation, *Geophysical Journal International*, **218**(2), 1100–1120. 69, 81

Liu, Q., Beller, S., Lei, W., Peter, D., & Tromp, J., 2021. Preconditioned BFGS-based Uncertainty Quantification in elastic Full Waveform Inversion, *Geophysical Journal International*, **228**(2), 796–815. 70, 76

Martin, J., Wilcox, L. C., Burstedde, C., & Ghattas, O., 2012a. A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM Journal on Scientific Computing*, **34**(3), A1460–A1487. 44, 45, 46

Martin, J., Wilcox, L. C., Burstedde, C., & Ghattas, O., 2012b. A Stochastic Newton MCMC Method for Large-Scale Statistical Inverse Problems with Application to Seismic Inversion, *SIAM Journal on Scientific Computing*, **34**(3), A1460–A1487. 69

Matheron, G., 1973. The intrinsic random functions and their applications, *Advances in applied probability*, **5**(3), 439–468. 100

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., & Virieux, J., 2014. Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures, *Geophysical Prospecting*, **62**(6), 1353–1375. 70

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E., 1953. Equation of state calculations by fast computing machines, *The journal of chemical physics*, **21**(6), 1087–1092. 47

Michel, J.-C., Moulinec, H., & Suquet, P., 1999. Effective properties of composite materials with periodic microstructure: a computational approach, *Computer methods in applied mechanics and engineering*, **172**(1-4), 109–143. 63

Mitchell, M., 1998. *An introduction to genetic algorithms*, MIT press. 47

Mitchell, T. & Faulkner, D., 2012. Towards quantifying the matrix permeability of fault damage zones in low porosity rocks, *Earth and Planetary Science Letters*, **339**, 24–31. 31

Mosegaard, K. & Sambridge, M., 2002. Monte Carlo analysis of inverse problems, *Inverse problems*, **18**(3), R29. 48

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research: Solid Earth*, **100**(B7), 12431–12447. 35

Mosser, L., Dubrule, O., & Blunt, M. J., 2020. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *Mathematical Geosciences*, **52**(1), 53–79. 28

Moulinec, H. & Suquet, P., 1998. A numerical method for computing the overall response of nonlinear composites with complex microstructure, *Computer methods in applied mechanics and engineering*, **157**(1-2), 69–94. 63

Muir, J. B. & Tsai, V. C., 2020. Geometric and level set tomography using ensemble Kalman inversion, *Geophysical Journal International*, **220**(2), 967–980. 28

Muir, J. B., Clayton, R. W., Tsai, V. C., & Brissaud, Q., 2022. Parsimonious velocity inversion applied to the Los Angeles Basin, CA, *Journal of Geophysical Research: Solid Earth*, **127**(2), e2021JB023103. 28

Mukumoto, K., Capdeville, Y., Singh, S., & Tsuji, T., 2024. Accounting for subwavelength heterogeneities in full waveform inversion based on wavefield gradient measurements, *Geophysical Journal International*, **238**(1), 235–256. 54

Neal, R. M., 1993. Probabilistic inference using Markov chain Monte Carlo methods, *Technical Report CRG-TR-93-1. Department of Computer Science, University of Toronto Toronto, ON, Canada*. 29

Nguetseng, G., 1989. A general convergence result for a functional related to the theory of homogenization, *SIAM Journal on Mathematical Analysis*, **20**(3), 608–623. 56

Nocedal, J. & Wright, S. J., 2006. *Numerical optimization*, Springer. 38, 39, 40, 41, 42, 43

Papanicolaou, G. C., 1979. Boundary value problems with rapidly oscillating random coefficients, in *Colloquia Math. Soc., Janos Bolyai*, vol. 27, pp. 853–873. 62

Parnell, W. & Abrahams, I., 2008. Homogenization for wave propagation in periodic fibre-reinforced media with complex microstructure. I—theory, *Journal of the Mechanics and Physics of Solids*, **56**(7), 2521–2540. 54

Patel, R. G. & Jain, T., 2019. Performance of ensemble Kalman filter and Markov chain Monte Carlo under uncertainty in forecast model, *Journal of Petroleum Science and Engineering*, **177**, 415–431. 47

Peter, D., Komatitsch, D., Luo, Y., Martin, R., Le Goff, N., Casarotti, E., Le Loher, P., Magnoni, F., Liu, Q., Blitz, C., et al., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophysical Journal International*, **186**(2), 721–739. 66

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503. 69

Plessix, R.-E. & Mulder, W., 2004. Frequency-domain finite-difference amplitude-preserving migration, *Geophysical Journal International*, **157**(3), 975–987. 73

Polak, E. & Ribiere, G., 1969. Note sur la convergence de méthodes de directions conjuguées, *Revue française d'informatique et de recherche opérationnelle. Série rouge*, **3**(16), 35–43. 41

Pyrcz, M. J. & Deutsch, C. V., 2014. *Geostatistical reservoir modeling*, Oxford university press. 31, 106

Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M. K., 2014. Seismic tomography and the assessment of uncertainty, *Advances in geophysics*, **55**, 1–76. 46

Raydan, M., 1997. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM Journal on Optimization*, **7**(1), 26–33. 41

Rudin, L. I., Osher, S., & Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms, *Physica D: nonlinear phenomena*, **60**(1-4), 259–268. 38

Sambridge, M., 2014. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International*, **196**(1), 357–374. 52

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Reviews of Geophysics*, **40**(3), 3–1. 35

Sambridge, M., Gallagher, K., Jackson, A., & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophysical Journal International*, **167**(2), 528–542. 97

Sánchez-Palencia, E., 1980. Non-homogeneous media and vibration theory, *Lecture Note in Physics, Springer-Verlag*, **320**, 57–65. 54

Santos, T., Bodin, T., Soulez, F., Ricard, Y., & Capdeville, Y., 2024. Refining tomography with generative neural networks trained from geodynamics, *Geophysical Journal International*, **238**(3), 1676–1695. 31, 54

Santosa, F. & Symes, W. W., 1988. High-frequency perturbational analysis of the surface point-source response of a layered fluid, *Journal of Computational Physics*, **74**(2), 318–381. 70

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., & Lenzen, F., 2009. *Variational methods in imaging*, vol. 167, Springer. 48

Schoenberg, M. & Sayers, C. M., 1995. Seismic anisotropy of fractured rock, *Geophysics*, **60**(1), 204–211. 53

Shanno, D. F., 1970. Conditioning of quasi-Newton methods for function minimization, *Mathematics of computation*, **24**(111), 647–656. 41

Sheriff, R. E. & Geldart, L. P., 1995. *Exploration seismology*, Cambridge university press. 110

Shewchuk, J. R. et al., 1994. *An introduction to the conjugate gradient method without the agonizing pain*, Carnegie-Mellon University. Department of Computer Science Pittsburgh. 41

Singh, S., Tsvankin, I., & Naeini, E. Z., 2018. Bayesian framework for elastic full-waveform inversion with facies information, *The Leading Edge*, **37**(12), 924–931. 28

Snieder, R., 2001. Inverse problems in geophysics, in *Signal recovery and synthesis*, p. SMA2, Optica Publishing Group. 35

Snieder, R. & Trampert, J., 2003. Linear and nonlinear inverse problems, in *Geomatic method for the analysis of data in the earth sciences*, pp. 93–164, Springer. 34

Srivastava, R., 1992. Reservoir characterization with probability field simulation, in *SPE Annual Technical Conference and Exhibition*, pp. SPE–24753, SPE. 100

Stacey, R., 1988. Improved transparent boundary formulations for the elastic-wave equation, *Bulletin of the Seismological Society of America*, **78**(6), 2089–2097. 66

Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., & Vallès, B., 2011. Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter, *Computational Geosciences*, **15**(2), 293–305. 48

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical geology*, **34**(1), 1–21. 100

Stuart, A. M., 2010. Inverse problems: a Bayesian perspective, *Acta numerica*, **19**, 451–559. 48

Suzuki, S., Caumon, G., & Caers, J., 2008. Dynamic data integration for structural modeling: model screening approach using a distance-based model parameterization, *Computational Geosciences*, **12**(1), 105–119. 28

Tape, C., Liu, Q., Maggi, A., & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophysical Journal International*, **180**(1), 433–462. 1, 11, 22

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, SIAM. 34, 69, 70

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Reviews of Geophysics*, **20**(2), 219–232. 34, 35

Taty Moukati, F., Stoica, R. S., Bonneau, F., Wu, X., & Caumon, G., 2025. From fault likelihood to fault networks: stochastic seismic interpretation through a marked point process with interactions, *Mathematical Geosciences*, **57**(1), 115–151. 97

Thore, P., Shtuka, A., Lecour, M., Ait-Ettajer, T., & Cognot, R., 2002. Structural uncertainties: Determination, management, and applications, *Geophysics*, **67**(3), 840–852. 28

Tikhonov, A. N., 1963. On the solution of incorrectly stated problems and a method of regularization, in *Dokl. Acad. Nauk. SSSR*, vol. 151, pp. 501–504. 23

Tompson, A. F., Ababou, R., & Gelhar, L. W., 1989. Implementation of the three-dimensional turning bands random field generator, *Water Resources Research*, **25**(10), 2227–2243. 120

Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophysical Journal International*, **160**(1), 195–216. 68

Tromp, J., Komatitsch, D., Liu, Q., et al., 2008. Spectral-element and adjoint methods in seismology, *Communications in Computational Physics*, **3**(1), 1–32. 66

Verschuur, D. J., Berkhout, A., & Wapenaar, C., 1992. Adaptive surface-related multiple elimination, *Geophysics*, **57**(9), 1166–1177. 24

Versteeg, R., 1994. The Marmousi experience: Velocity model determination on a synthetic complex data set, *The Leading Edge*, **13**(9), 927–936. 26, 124

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method, *Geophysics*, **51**(4), 889–901. 72

Visser, G., Guo, P., & Saygin, E., 2019. Bayesian transdimensional seismic full-waveform inversion with a dipping layer parameterization, *Geophysics*, **84**(6), R845–R858. 28

Walsh, J., Watterson, J., & Yielding, G., 1991. The importance of small-scale faulting in regional extension, *Nature*, **351**(6325), 391–393. 31

Wang, F. & Alkhalifah, T., 2024. Geological and Well Prior Assisted Full Waveform Inversion Using Conditional Diffusion Models, in *85th EAGE Annual Conference & Exhibition (including the Workshop Programme)*, vol. 2024, pp. 1–5, European Association of Geoscientists & Engineers. 28

Wang, H., Wellmann, J. F., Li, Z., Wang, X., & Liang, R. Y., 2017. A segmentation approach for stochastic geological modeling using hidden Markov random fields, *Mathematical Geosciences*, **49**(2), 145–177. 28

Wang, N., Montagner, J.-P., Fichtner, A., & Capdeville, Y., 2013a. Intrinsic versus extrinsic seismic anisotropy: The radial anisotropy in reference earth models, *Geophysical Research Letters*, **40**(16), 4284–4288. 54

Wang, P., Ray, S., Peng, C., Li, Y., & Poole, G., 2013b. Premigration deghosting for marine streamer data using a bootstrap approach in tau-p domain, in *SEG Technical Program Expanded Abstracts 2013*, pp. 4221–4225, Society of Exploration Geophysicists. 24

Wei, Z., Mei, J., Wu, Z., Zhang, Z., Huang, R., & Wang, P., 2023. Pushing seismic resolution to the limit with FWI imaging, *The Leading Edge*, **42**(1), 24–32. 24, 25

Wellmann, J. F., Horowitz, F. G., Schill, E., & Regenauer-Lieb, K., 2010. Towards incorporating uncertainty of structural data in 3D geological inversion, *Tectonophysics*, **490**(3-4), 141–151. 28

Wellmann, J. F., Lindsay, M., Poh, J., & Jessell, M., 2014. Validating 3-D structural models with geological knowledge for improved uncertainty evaluations, *Energy Procedia*, **59**, 374–381. 28

Wellmann, J. F., Thiele, S. T., Lindsay, M. D., & Jessell, M. W., 2016. pynoddy 1.0: an experimental platform for automated 3-D kinematic and potential field modelling, *Geoscientific Model Development*, **9**(3), 1019–1035. 7, 16, 82

Wiggins, R. A., 1972. The general linear inverse problem: Implication of surface waves and free oscillations for earth structure, *Reviews of Geophysics*, **10**(1), 251–285. 34

Wu, X., Shi, Y., Fomel, S., & Liang, L., 2018. Convolutional neural networks for fault interpretation in seismic images, in *SEG International Exposition and Annual Meeting*, pp. SEG–2018, SEG. 24

Zhang, J., Shan, X., Huo, S., Huang, L., Zheng, W., Zhou, X., & Liu, E., 2025. Deep learning-driven multi-frequency seismic inversion for enhanced thin-layer stratigraphic characterization, *Journal of Applied Geophysics*, p. 105749. 28

Zhang, Z. & Alkhalifah, T., 2022. Regularized elastic full-waveform inversion using deep learning, in *Advances in subsurface data analytics*, pp. 219–250, Elsevier. 28

Zhang, Z., Wu, Z., Wei, Z., Mei, J., Huang, R., & Wang, P., 2020. FWI imaging: Full-wavefield imaging through full-waveform inversion, in *SEG International Exposition and Annual Meeting*, p. D031S027R004, SEG. 24

Zhao, X. & Curtis, A., 2024. Physically Structured Variational Inference for Bayesian Full Waveform Inversion, *Journal of Geophysical Research: Solid Earth*, **129**(11), e2024JB029557. 28

Zhou, H., Gómez-Hernández, J. J., & Li, L., 2014. Inverse methods in hydrogeology: Evolution and recent trends, *Advances in Water Resources*, **63**, 22–37. 48

Zhu, H., Li, S., Fomel, S., Stadler, G., & Ghattas, O., 2016. A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration, *Geophysics*, **81**(5), R307–R323. 69, 70, 81

Stochastic inversion of FWI images for reducing structural uncertainty

Abstract: Despite ongoing improvements in the accuracy and efficiency of seismic interpretation, resolving small-scale geological structures remains challenging due to inherent limitations. Indeed, seismic imaging techniques are based on frequency band-limited data, and therefore they provide smooth representations of the real complex geology. This means that, within the resolution limit of seismic data, there is room for alternative valid interpretations at the small-scale. To make sure that such small-scale geological features are appropriately considered in subsurface uncertainty studies, this thesis proposes the use of a downscaling strategy. In particular, from a smooth representation of the real subsurface geology, obtained through the well-known Full Waveform Inversion (FWI) technique, the downscaling inversion aims to recover all the finer scale models compatible with the FWI solution. Downscaling is formulated here as an inverse homogenization problem, where the non-periodic homogenization serves as forward operator to construct smooth long-wavelength equivalent models of the subsurface. This thesis presents applications on synthetic models, where downscaling is applied to the two-dimensional case of a fault characterization problem in a finely stratified medium in order to properly detect fault geometric features and estimate small-scale elastic property variations. For this purpose, the inverse problem is cast into a Bayesian probabilistic framework that incorporates stochastic structural modeling, wherein fault displacement is produced by using a direct kinematic modeling approach and fine-scale stratigraphy is modeled with a geostatistical approach. The results show that downscaling inversion provides a reliable estimate of both small-scale stratigraphic and fault structures while effectively quantifying uncertainties. This work also highlights the additional challenges to be considered for real-case studies and the importance of the chosen parametrization to represent complex geological scenarios. In spite of these extra challenges, the proposed methodology appears promising as a target-oriented and cost-effective strategy to reduce FWI interpretation uncertainty.

Keywords: Inverse problems, Homogenization, Downscaling inversion, Structural uncertainty, Seismic imaging

---

Inversion bayésienne d'images tomographiques pour réduire les incertitudes structurales

Résumé : Malgré les améliorations concernant la précision et l'efficacité de l'interprétation sismique, il est difficile de cartographier les structures géologiques à petite échelle. Les techniques d'imagerie sismique sont en effet basées sur des données à bande de fréquences limitée, ce qui fournit des représentations lissées de la géologie complexe réelle. Il existe donc plusieurs interprétations valables à petite échelle, dans les limites de résolution des données sismiques. Afin de garantir que ces objets géologiques à petite échelle sont correctement pris en compte dans les évaluations d'incertitude souterraine, cette thèse propose d'utiliser une stratégie de réduction d'échelle (*downscaling*). À partir d'une représentation lisse de la géologie réelle du sous-sol obtenue grâce à la technique bien connue d'inversion de forme d'onde complète (FWI), l'inversion par réduction d'échelle vise à récupérer tous les modèles à échelle plus fine compatibles avec la solution FWI. Le downscaling est formulé comme un problème d'homogénéisation inverse, dans lequel l'homogénéisation non périodique sert d'opérateur direct pour obtenir des modèles équivalents lisses à longue longueur d'onde du sous-sol. Cette thèse présente des applications sur des modèles synthétiques, où la réduction d'échelle est appliquée au cas bidimensionnel d'un problème de caractérisation de faille dans un milieu finement stratifié afin de détecter correctement les caractéristiques géométriques de la faille (position, forme et rejet) et d'estimer les variations des propriétés élastiques à petite échelle. À cette fin, le problème inverse est intégré dans un cadre probabiliste Bayésien qui intègre une modélisation stochastique des paramètres de faille. Le déplacement de la faille est produit à l'aide d'une approche de modélisation cinématique directe et la stratigraphie à petite échelle est modélisée à l'aide d'une approche géostatistique inspirée des points pilotes. Les résultats montrent que l'inversion avec le downscaling fournit une estimation fiable des structures stratigraphiques et faillées à petite échelle tout en quantifiant efficacement les incertitudes. Ce travail met également en évidence les défis supplémentaires à prendre en compte pour les études de cas réels et l'importance du choix de la paramétrisation pour représenter des scénarios géologiques complexes. Malgré ces nouveaux défis, la méthodologie proposée semble prometteuse en tant que stratégie ciblée et abordable pour quantifier et réduire l'incertitude liée à l'interprétation géologique de la FWI.

Mots-clés : Problèmes inverses, Homogénéisation, Inversion par réduction d'échelle, Incertitude structurale, Imagerie sismique