UNIVERSITÀ
DEGLI STUDI
FIRENZE

PhD in
EARTH SCIENCES

CYCLE XXXIII

COORDINATOR Prof. Francalanci Lorella

# DEVELOPMENT AND APPLICATION OF BAYESIAN ALGORITHMS FOR RESERVOIR CHARACTERIZATION

Academic Discipline (SSD): GEO 11


**Doctoral Candidate**
Alessandro Salusti

**Supervisor**
Prof. Mazzotti Alfredo

Dr. Aleardi Mattia

**Coordinator**
Prof. Francalanci Lorella


Years 2017/2020

## ABSTRACT

Reservoir characterization jointly exploits seismic and well log data to infer the structural setting and the physical properties (i.e. elastic and/or petrophysical properties) of the subsurface reservoir, which constitute essential information for optimizing either hydrocarbon exploration and production or $CO_2$ sequestration. Frequently, reservoir characterization uses the changes of the seismic amplitudes with respect to the source-receiver offset (Amplitude *versus* Offset, AVO) to infer the elastic contrasts at the reflecting interfaces and to detect changes in the lithology and in the rock-saturating fluids.

In this thesis, we invert the AVO seismic responses to infer the elastic parameters and, when a calibrated rock-physics model is available, the associated petrophysical parameters. We focus on both target-oriented and interval-oriented AVO inversions of pre-stack seismic data. The former approach only inverts the AVO responses of the target reflection, whereas the interval-oriented method inverts the seismic amplitudes within a time interval including both the target layer and the surrounding geological formations. In all cases, we cast the inversion into a Bayesian framework that allows including prior information (e.g. derived from well logs) inside the inversion framework to mitigate the ill-conditioning of the inverse problem and to stabilize the final solution.

We developed four different inversion algorithms all based on numerical Markov chain Monte Carlo (McMC) approaches, that are used to numerically derive accurate uncertainty assessments in case of non-linear forward operators and/or non-Gaussian prior assumptions. The numerical solution of Bayesian inverse problems is computationally expensive and time-consuming, thus we present novel McMC strategies that not only use different strategies to mitigate the ill-conditioning of the AVO inversion but are also devoted to reduce the computational cost of the probabilistic sampling.

First, we present target- and interval-oriented McMC elastic inversion suited for non-parametrical and multimodal distributions for the joint estimation of elastic properties (P- and S-wave velocities and density) and litho-fluid facies from pre-stack data. It exploits geostatistical constraints to reduce the ill-conditioning of the elastic AVO inversion, whereas advanced meta-algorithms, as the Parallel Tempering and the Delayed Rejection Scheme, are used to speed up the convergence toward a stable posterior model. Then, differently from the first approach, the second McMC method we discuss assumes an uninformative prior model (i.e. a uniform prior distribution) and uses a transdimensional inversion framework to reduce the ill-conditioning of the inversion procedure. More in detail, the number of model parameters (i.e. the number of layers) is treated as an unknown, and a reversible jump Markov Chain Monte Carlo (rjMcMC) algorithm is used to sample the variable-dimension model space. This inversion scheme provides a parsimonious solution and

reliably quantifies the uncertainties affecting the estimated parameters. This approach is developed for the estimation of the elastic properties and also petrophysical parameters (i.e. porosity, clay content and water saturation) from the AVO responses, but this transdimensional framework is also employed to implement a data-driven inversion that automatically includes lateral constraints into the target-oriented elastic AVO inversion. It is known that the main drawback of McMC inversion is the considerable number of forward model evaluations needed to attain stable uncertainty estimations, although some strategies can be used to partially reduce this computational burden. To this end, we introduce a numerical Hamiltonian Monte Carlo method for both interval- and target-oriented AVO inversion that exploits the derivative information of the objective function to dramatically speed up the convergence of the sampling toward the stationary regime. Again, geostatistical constraints are used to mitigate the ill-conditioning of the inverse problem.

The curse-of-dimensionality issue usually hampers the application of the previously described inversion approaches to simultaneously invert a 2D seismic section or a 3D seismic volume. For example, the previously mentioned interval-oriented McMC inversions must be applied to each seismic gather separately to make the computational cost affordable. For this reason, the last method we present combines a discrete cosine transform (DCT) reparameterization of data and model spaces with a convolutional neural network (CNN) to simultaneously solve the elastic AVO inversion along a 2D section (but the method is also extendible to 3D volumes). On the one hand, the CNN is trained to predict the mapping between the DCT-transformed seismic data and the DCT-transformed 2-D elastic model. On the other hand, the DCT reduces the dimensionality of the input and output of the network, and also acts as a regularization operator in the model space that preserves the lateral and vertical continuity of the elastic properties in the recovered solution. Once trained, the network can estimate the elastic properties along a 2-D section from the observed seismic data in near real-time. In addition, a Monte Carlo simulation framework is used to propagate onto the estimated elastic model the uncertainties related to both noise contamination and network approximation.

All the presented approaches have been tested on synthetic data and benchmarked against analytical inversion algorithms, checking also the influence of possible errors in the propagation wavelet and of the presence of noise in the data. In some cases, the implemented inversion strategies have also been applied to actual seismic and well log data pertaining to a 3D land dataset.

## PREFACE

This dissertation is submitted for the degree of Doctor of Philosophy at the School of Graduate Studies of the University of Florence.

My PhD research focused on two parallel projects. The main research project, described here, concerns the development and application of inversion algorithms for the geophysical characterization of reservoirs. All these innovative algorithms have been developed and applied within the geophysical research group of the University of Pisa, and some of them have partly been published in international papers.

The minor research project, which is not discussed here, focuses on the implementation and testing of analytical inversion approaches for reservoir characterization. This work was supported by ENI, Exploration and Productions Division, and was conducted under the supervision of Professor Alfredo Mazzotti and Dr. Mattia Aleardi.

## ACKNOWLEDGEMENTS

## OBJECTIVES AND THESIS OUTLINES

The goal of this PhD thesis is to stably address highly non-linear, ill-conditioned inverse problems through McMC approaches. The theory of McMC approaches has been thoroughly developed in the literature, and it is well known that the non-linearity of the forward operator and the ill-conditioning are the main issues in inverse problems. The numerical treatment of these ill-conditioned systems is complicated and time-consuming; therefore, developing new strategies to make them computationally fast and still very accurate is essential.

This dissertation is organized as follow:

- Chapter 1 is composed of a brief introduction to the geophysical inverse problems, AVO analysis, rock-physics models, target- and interval-oriented approaches and Monte Carlo methods for reservoir characterization, all essential components of the successive chapters.

- Chapter 2 describes two inversion strategies for the joint estimation of elastic properties and litho-fluid facies from pre-stack seismic data in case of non-parametric mixture prior distributions and non-linear forward modelling. The introduction of geostatistical constraints reduces the inherent ill-conditioning of the problem, whereas the Parallel Tempering and the Delayed Rejection Scheme dramatically speed up the algorithm convergence rate. These approaches are then benchmarked against two analytical inversion approaches that assume Gaussian-mixture-distributed elastic parameters.

- Chapter 3 presents a Reversible-jump Markov-chain Monte Carlo (Rj-McMC) approach suitable for both the elastic and petrophysical inversion. This strategy samples the variable-dimension model space to provide parsimonious solutions, thus inherently reducing the ill-conditioning of the AVO inversion. The final outcomes provided by the presented algorithms are successively compared with the predictions from linear Bayesian elastic and petrophysical inversions.

- Chapter 4 discusses another Reversible-jump approach in which the adopted model discretization allows the introduction of an automatic, data-driven procedure to include lateral constraints inside the inversion framework. This method is again benchmarked against a standard Bayesian approach.

- Chapter 5 assesses the suitability of a Hamiltonian Monte Carlo (HMC) algorithm for non-linear target- and interval-oriented AVO inversions for the estimation of elastic properties and associated uncertainties from pre-stack seismic data. Compared to previous approaches, the Hamiltonian description of the system exploits the gradient information of the objective function to remarkably speed up the convergence rate.

- Chapter 6 proposes an alternative approach to include lateral constraints, potentially 3D, in the inversion framework through a machine learning approach that combines a discrete cosine transform (DCT) of both data and model spaces with a convolutional neural network (CNN).

- Chapter 7 draws the general conclusions and future developments of the presented approaches.

# INDEX

# CHAPTER 1

# An overview of probabilistic elasto-petrophysical seismic inversion.

## 1.1 Reservoir Characterization

Schatzinger and Jordan (1999) define reservoir characterization as the process of creating a high-resolution interdisciplinary geoscientific model that integrates different geological and engineering information to provide reliable predictions of reservoir performance, to reduce risks and to maximize profits. It is essential to effectively manage all disciplines to successfully conduct each operation, from the initial exploration that leads to a discovery to the development of the field to production.

This is clearly an interdisciplinary field; from geological and geophysical analyses, the overall oil or gas or water volume is estimated, which is then adopted by field engineers to generate a recovery production plan that is both economic and efficient. Geophysicists contribute to this process by providing measurements between the various drilled wells to extend geological information away from the wellbores. This quantitative information translates into the thickness of the reservoir, rock porosity and permeability, pressure distribution, distribution of fractures and saturation of fluids in the pores. These properties determine the model geometry and rheology to be used in production simulations. All these data are extremely critical and important since production volumes that exceed the estimated forecasts, pressure anomalies not foreseen in the drilling phase, or well test

results that do not comply with the predictions result in very large economic losses for the company that carries them out.

Although technology has made remarkable progress in recent years by improving the acquisition of data by digital telemetry, developing new data processing technologies and using supercomputers, the average cost of these operations, which is obviously a case-dependent variable, is estimated to be approximately several millions of dollars.

We remark the importance of reservoir characterization for hydrocarbon exploration, aquifer identification, $CO_2$ sequestration and groundwater pollution tracing. The $CO_2$ geo-sequestration is one of the most promising technologies for reducing atmospheric $CO_2$ emission. The carbon dioxide is captured and injected in deep porous and fractured geological formations for long-term storage; in this context the site characterization is critical for safe and effective storage since geological heterogeneity affects reservoir quality, fluid migration, and trapping. Reservoir characterization is critical also for groundwater pollution tracing because if chemical pollutants pour in the subsurface, due to spills or leakages, generate a fluid flow directed towards the porous rocks. The prior identification of the possible target reservoirs develops containment scenarios useful to reduce the pollution expansion.

The main tool at geophysicists' disposal is seismic data. When a potentially productive field is discovered, an extended seismic campaign is conducted, and several exploration wells are drilled. Artificially generated seismic waves are therefore generated through powerful mechanical sources and propagated in the subsurface.
When these waves encounter the discontinuities that define layer boundaries, they behave as described by Snell's law. Part of the energy carried by the waves is propagated through the layer, part is reflected back, and part is dissipated. Therefore, we can intuitively say that the waves interact with the subsurface rocks, and when they are reflected back, they carry information about the subsurface geometries and properties. The seismic inversion process is therefore aimed at tracing back these attributes from the seismic data.

The seismic data acquired in the field need to be carefully processed to be correctly used. We do not focus on this aspect since it is a very large and complex geophysical inversion; instead, we always assume that the seismic data we use have already been correctly processed. From pre-stack seismic data, we run elastic inversions to derive the volume distribution of elastic properties associated with

the area under investigation: P-wave velocity, S-wave velocity and density. Well logs are used as an alternative source of information through which we can derive both constraints that facilitate the inversion process and empirical/theoretical relations (rock-physics models) that link the elastic variables to the petrophysical ones we are interested in.

From the elastic attributes and a calibrated rock-physics model, we can therefore run petrophysical inversions to map the associated petrophysical variables or litho-fluid facies, as schematically reported in Figure (1).

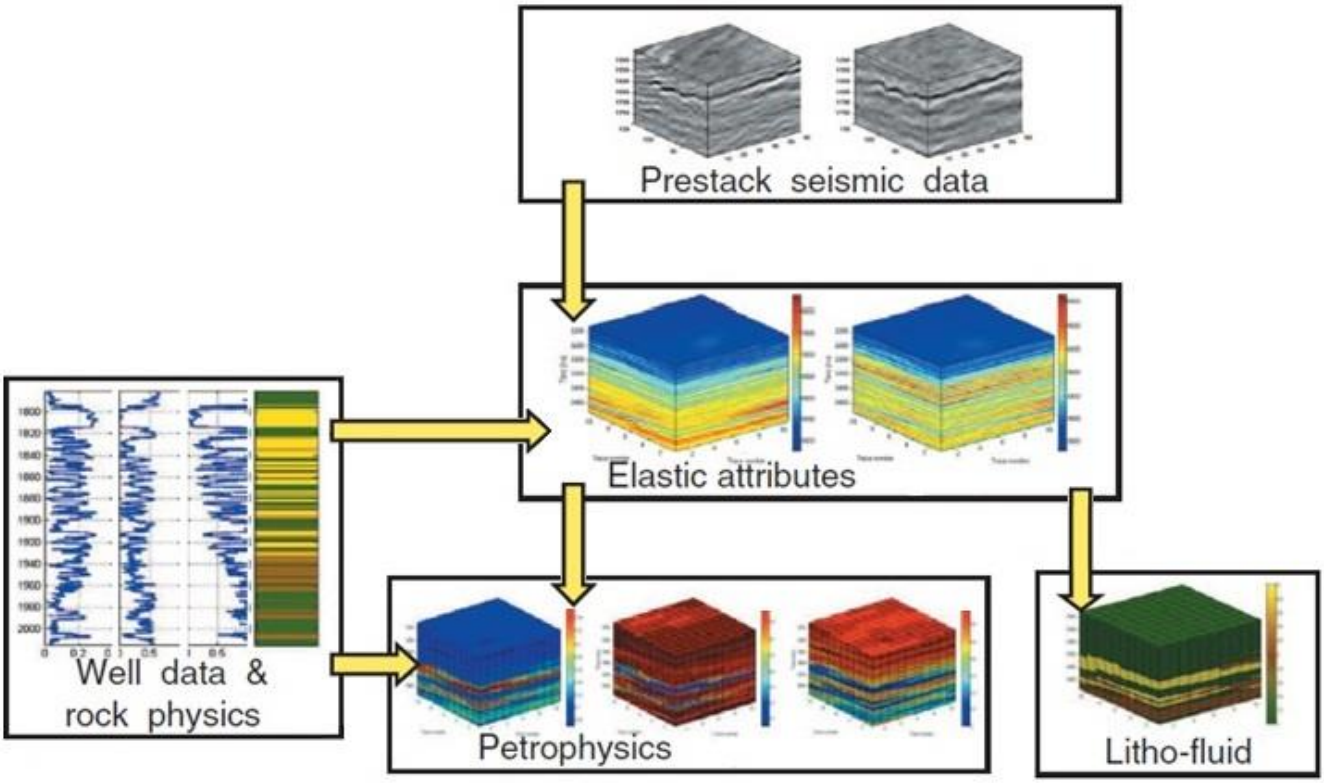


*Figure 1: Schematic flowchart of the elastic-petrophysical inversion. Modified from Grana et Della Rossa,2010 .*

## 1.2 Inverse problems

The problem of finding the cause of an observed effect is generally termed inversion (Tarantola, 1987). The objective of this process is to find a description of the model parameters from the recorded observations; in other words, we infer information about the Earth's interior from physical measurements at the surfaces.

Inverse theory is the opposite of forward theory. Menke (1989) defines forward theory as "the process of predicting the results of measurements on the basis of some general principle or model, and a set of specific conditions relevant to the problem available".
Through the inversion the model parameters can be estimated, starting with observed data and general knowledge of the model. Forward modelling consists of computing synthetic seismic data for a known Earth model. Therefore, the inverse problem consists of estimating the rock properties that describe the Earth model from the seismic data.

## 1.3 Seismic forward modelling

Forward modelling is the first step in model-based seismic inversion (Menke, 1989). The Earth model can be described by means of elastic parameters, such as P-wave velocity, S-wave velocity and density. Then, if the seismic pulse (wavelet) is known, the Earth model can be combined with a seismic source signature to create a synthetic seismogram. Typically, this signature is composed by offset-dependent wavelets whose central frequency changes to correctly address the natural loss of resolution at far offset. In the following, we always adopt as source signature an offset-independent single wavelet.

A synthetic seismic trace is defined by the convolution of the wavelet with a reflectivity model. The synthetic trace is the result of forward modelling, obtained through the convolutional model as shown in Figure (2); in contrast the recorded seismic trace represents the input of the inversion. A seismic pulse or wavelet represents a packet of energy arriving from a seismic source (Barclay *et al*, 2008). It may be estimated from the seismic data in the target zone under certain assumptions, or from seismic data and well log data (Buland, 2002).

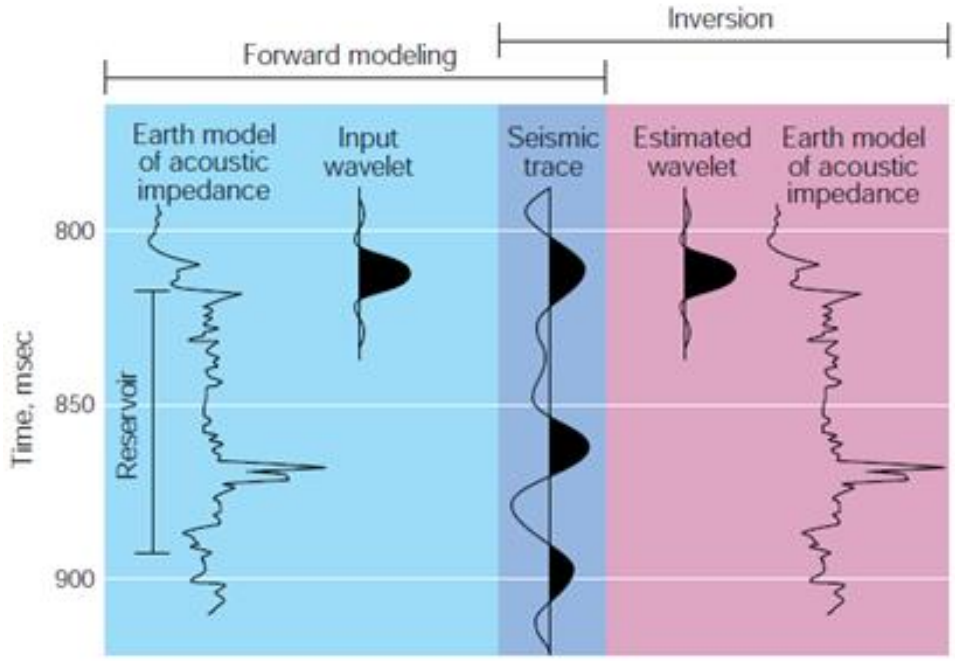


*Figure 2: Process of forward modelling based on the convolutional model and inversion. Modified from Barclay et al., 2008.*

### 1.3.1 Classic approach to seismic inversion

Menke (1989) asserts that the basic statement of an inverse problem is that the model parameters and the data are in some way related. In the classic approach, this problem is defined as a mathematical model: there are an Earth model **m** defined by some parameters, and seismic data **d**. Then, the forward modelling normally can be written as:

$$\mathbf{d} = G(\mathbf{m}) + \mathbf{n} \qquad (1)$$

where $G$ is a modelling operator and **n** is the error term. The modelling operator therefore describes the equations that link the model parameters to the observed data. It can be represented by an ordinary differential equation, by a partial differential equation or by a system of algebraic equations. As Buland (2002) affirmed in his work, the error term **n** represents both random noise and systematic errors related to effects that are not included in the modelling operator (Scales and Snieder, 1998).

According to Aster *et al.*, (2018) the goal is to recover **m** from noisy data **d**. When the problem is a well-conditioned linear problem the maximum likelihood principle yields the least square solution, optimizing an objective function error **E**:

$$\mathbf{E} = \|\mathbf{d} - G(\mathbf{m})\|^2 \qquad (2)$$

Some misfit is expected in the data prediction of the forward model and the observed data because of the noise in the data. However, an optimal solution may be the **m** which minimizes **E**. Although the least square solution works satisfactorily in the case of well-conditioned problems, when the problem is ill-posed, this solution leads to many physically unreasonable models; thus, the problem has to be regularized.

Well-known problems in inversion strategies involve (Aster *et al.* 2018):

- Absence of solutions: the absence of a model that perfectly describes the data can derive either from the noise affecting the data, which changes the observations, or from an overly approximated mathematical model.
- Uniqueness of solution: the existence of a solution does not mean that is the only one able to describe our observations. There is a potentially infinite number of models that perfectly describe our observations with as much accuracy as the true model.
- Instability of the inversion approach: small fluctuations in the observed data may result in large variations in the estimated models. Inverse problems characterized by such problematics are named ill-conditioned; in this case regularization processes are needed.

In practice, all these problems can be overcome by using prior information about the parameters. The most fundamental and straightforward method is the so-called Bayesian approach to inversion.

### 1.3.2 Bayesian approach to inverse problems

Regarding what constitutes a solution to an inverse problem; there are many points of view. Buland (2003) asserts that the statistical approach addresses the solution of an inverse problem by a probability density function in the model space instead of limiting the solution to a single set of predicted parameters. Duijndam (1998) states that the solution of an inverse problem is obtained through the combination of information concerning the data, theoretical relations and prior

information about the parameters in a suitable form. Therefore the mathematical tools to be used are those of probability theory knowing that the uncertainty in the information plays an important role. In all proposed McMC algorithms, we use the Bayesian approach because through this methodology it is possible to characterize the uncertainty of the inversion results and not merely to limit the solution to a single best-fitting set of the model parameters.

The two main differences between the Bayesian and the classic approaches are therefore as follows:

- The probabilistic nature of the derived solution described here as a probability distribution with respect to the model parameters.
- The inclusion of prior information in the inversion process. This information is derived from general knowledge about the investigated area, from well logs or from other information obtained independently from the observed data. This is again described by a probability distribution with respect to the model parameters.

The interest in the characterization and determination of uncertainties associated with measurements derives from the numerous approximations that are made in both the seismic processing and the inversion: the seismic data are processed to obtain images of the Earth; similarly, well log measurements are processed to calculate several average properties; seismic modelling is then based on an approximated wave propagation model within a medium considered isotropic and elastic.
To understand and analyse uncertainties therefore helps us to better estimate the risks, to make the best decisions aimed at avoiding these risks and to assign appropriate weights to the various data before combining them in the interpretative model.

Following this Bayesian approach, the model is a vector of random variables, and the solution is therefore a probability distribution for the model parameters. The Bayes theorem can be expressed as:

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m}) \cdot p(\mathbf{m})}{p(\mathbf{d})} \qquad (3)$$

where both **d** and **m** are vectors of variables. The vector **m** represents the model's parameters and **d** represents the data. The term $p(\mathbf{m})$ denotes the prior probability and it is the marginal probability

density function (PDF) for **m**. The denominator $p(\mathbf{d})$ can be considered as a constant factor in the inverse problem since it basically normalizes the conditional distribution $p(\mathbf{m}|\mathbf{d})$. The function $p(\mathbf{m}|\mathbf{d})$ is the so-called posterior PDF, which is the conditional PDF of **m** after measurements of the data vector **d** are obtained (Duijndam, 1998). In addition to the theoretical relation between parameters and data, it takes noise into account. The posterior distribution $p(\mathbf{m}|\mathbf{d})$contains the complete information for **m**. In cases where a single model is required as an answer it may be suitable to use the model with the largest value of $p(\mathbf{m}|\mathbf{d})$ which is referred to as the maximum a posteriori (MAP) model. When the posterior distribution is normal the mean denotes the MAP.

## 1.4 Amplitude versus Offset Analysis

All proposed approaches use the inversion process to derive the subsurface properties of interest from the amplitude *versus* offset (AVO) response. AVO analysis is based on the reflectivity dependence of the offset increase on the interface between two materials with contrasting elastic properties. Diverse lithologies can exhibit different values of the elastic modulus, making it possible to use the AVO as a tool for lithological discrimination (Castagna and Backus, 1993, Avseth *et al.* 2003).

### 1.4.1 AVO Principles

When seismic waves travel into the Earth and encounter layer boundaries with property contrasts (velocity and density), the energy of the incident wave is divided at each boundary. Part of the incident energy associated with a compressional source is converted to a shear wave (Bancroft, 2006). This conversion occurs when a wave is not incident with a normal angle (zero degrees from the perpendicular to the surface) and its energy is divided into reflected and converted S- and P-waves (Burianyk and Pickford, 2000). Then, both the compressional and shear wave energy are partly reflected from and partly transmitted through each of these layer boundaries.

The fraction of the incident energy that is reflected therefore depends upon the angle of incidence, unknown, and that is typically estimated through a low-frequency model. In the following, we always assume that the low-frequency model is available and this conversion is allowed.

The AVO technique allows studying the variation in the amplitude of the seismic response based on the source-receiver distance (offset or angle).

- Zoeppritz equations (1919).

Zoeppritz (1919) published a theoretical work that is the foundation of the AVO technique. This work is summarized in a set of equations to calculate the amplitude of the reflected and transmitted plane wave at an interface as a function of the angle of incidence knowing the basic elastic constants of the media involved (P-wave velocity, S-wave velocity and density). The solutions of these equations are expressed in terms of the angle of reflection for the P- and S-waves.

$$\begin{bmatrix} Rpp \\ Rss \\ Tpp \\ Tss \end{bmatrix} = \begin{bmatrix} -\sin\theta_1 & -\cos\varphi_1 & \sin\theta_2 & \cos\varphi_2 \\ \cos\theta_1 & -\sin\varphi_1 & \cos\theta_2 & -\sin\varphi_2 \\ \sin 2\theta_1 & \frac{V_{p1}}{V_{s1}}\cos 2\varphi_1 & \frac{\rho_2 V_{s2}^2 V_{p1}}{\rho_1 V_{s1}^2 V_{p2}}\sin 2\theta_2 & \frac{\rho_2 V_{s2} V_{p1}}{\rho_1 V_{s1}^2}\cos 2\varphi_2 \\ -\cos 2\varphi_1 & \frac{V_{s1}}{V_{p1}}\sin 2\varphi_1 & \frac{\rho_2 V_{p2}}{\rho_1 V_{p1}}\cos 2\varphi_2 & -\frac{\rho_2 V_{p2}}{\rho_1 V_{p1}}\sin 2\varphi_2 \end{bmatrix}^{-1} \quad (4)$$

$$\cdot \begin{bmatrix} sin\theta_1 \\ cos\theta_1 \\ sin2\theta_1 \\ cos2\varphi_1 \end{bmatrix}$$

where $Rpp$, $Rss$, $Tpp$, and $Tss$ are the reflected P-, reflected S-, transmitted P- and transmitted S-wave amplitude coefficients, respectively; $\theta_1$ is the angle of incidence of the P-wave; $\theta_2$, angle of the transmitted P-wave, $\varphi_1$, angle of the reflected S-wave, and $\varphi_2$, angle of the transmitted S-wave.

Inverting the four Zoeppritz equations provides the coefficients as functions of angle. Many people have worked on developing simplifications of the Zoeppritz equations. The most commonly used approximations are those derived by Aki and Richards (1980) for three-term inversions and Shuey (1985) for two-term inversions.

- Aki and Richards approximation (1980)

In 1980, Aki and Richard postulated estimated values for the reflection coefficients for an incident P-wave. The approximation of Aki and Richards (1980) applies when there are small changes in density and velocity between two layers. This approximation is known as a weak contrast, and the equation is given in terms of average elastic properties.

The Aki and Richards linear approximation is appealing because it is written with three terms. The first term involves $Vp$, the second involves $Vs$, and the third involves density $\rho$. The equation can be written as follows:

$$Rpp(\theta) = \alpha_{Vp}(\theta)\frac{\Delta Vp}{\overline{Vp}} + \alpha_{Vs}(\theta)\frac{\Delta Vs}{\overline{Vs}} + \alpha_{\rho}(\theta)\frac{\Delta\rho}{\bar{\rho}} \tag{5}$$

$$= \frac{1}{2}\left(1 + tan^2(\theta)\right)\frac{\Delta Vp}{\overline{Vp}} - 4\frac{\overline{Vs}^2}{\overline{Vp}^2}sin^2(\theta)\frac{\Delta Vs}{\overline{Vs}}$$

$$+ \frac{1}{2}\left(1 - 4\frac{\overline{Vs}^2}{\overline{Vp}^2}sin^2(\theta)\right)\frac{\Delta\rho}{\bar{\rho}}$$

where *Rpp* is the P-P wave reflection coefficient, $\theta$ is the incidence angle, $\Delta x$ indicates the contrast in the elastic property $x$ across the reflecting interface and $\bar{x}$ is the average value of the property $x$ over the reflecting interface.

- Shuey approximation (1985)

Shuey (1985) subsequently published a closed form approximation of the Zoeppritz equations which is simplified in the following equation

$$Rpp(\theta) = R_0 + G\,sin^2(\theta) \tag{6}$$

The term $R_0$ is called the intercept and represents the reflection amplitude at normal incidence (θ=0), while the term $G$ is called the gradient and indicates the reflectivity variations as a function of the angle of incidence ($\theta$). This equation is accurate for angles of incidence as large as 25° to 30° (Castagna and Backus, 1993).

Plotting the observed reflection coefficients with a common mid-point (CMP) as a function of the incidence angle $\theta$ makes it possible to graphically observe the behaviour of the famous four AVO classes, as illustrated in Figure (3).

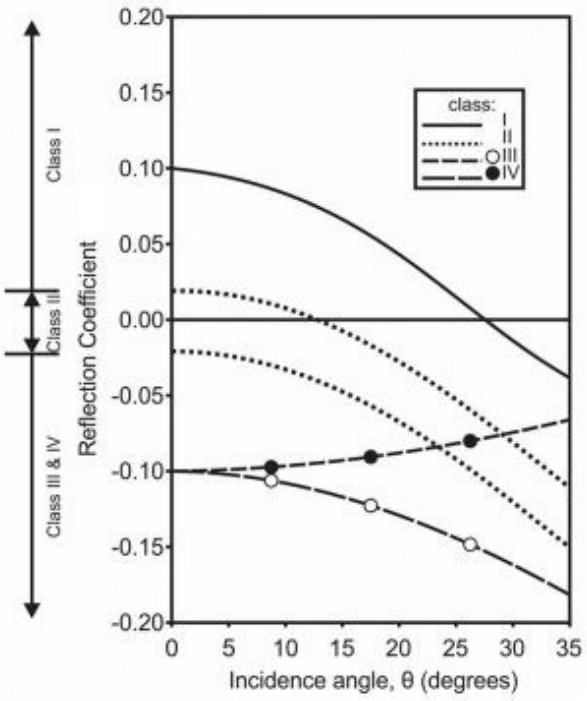


*Figure 3: The four different well-known AVO classes, divided by their gradient and intercept values.*

The four classes are identified by the different acoustic contrasts between shale and sand: Class I is generated by sands with high acoustic impedance, class II represents sands surrounded by shale with moderate acoustic impedance and class III represents sands with low acoustic impedance. In class IV, the normal reflection coefficient is negative and decreases as offset increases.

### 1.4.2 AVO processing

It is important to emphasize and remark that AVO processing should preserve the relative trace amplitudes within common mid-point (CMP) gathers. Seismic reflections must therefore be correctly positioned in the subsurface and data quality should be sufficient to guarantee that reflection amplitudes contain information about reflection coefficients (Avseth *et al.* 2005).
Yilmaz (2001) addressed the detailed processing sequence in his book. There are three important aspects of a processing sequence tailored for AVO analysis:

- The relative amplitudes of the seismic data must be preserved throughout the analysis in order to recognize amplitude variations with offset.
- The processing sequence must retain the broadest possible signal band in the data with a flat spectrum within the passband.
- Pre-stack amplitude inversion to derive the AVO attributes must be applied.

To take full advantage of the AVO technique, seismic data must be carefully processed so that any noticeable variation in amplitude is solely the result of reflection coefficient changes and is not a result of some processing artefact (Castagna and Chopra, 2014).
In processing, appropriate noise-removal is required in the sequence, as seismic data are analyzed pre-stack. Hence, there is not a unique sequence of processing to follow, but the adopted procedure must be determined by the data quality (Castagna and Chopra, 2014).

In summary, based on the fact that the reflection strength from the subsurface interface depends on the reflection angles and on the material properties where the reflection takes place, AVO inversion is a strong seismic pre-stack inversion technique for estimating elastic subsurface parameters (Buland and Omre 2003).

## 1.5 Rock-Physics Model

If we are interested in jointly estimating both the petrophysical and the elastic properties of the subsurface, an additional crucial component of the inversion is the so-called rock-physics Model (RPM). This is a link between the elastic variables, extracted by the AVO inversion, and the petrophysical variables and can be generally termed as:

$$[Elastic\ variables] = f_{RPM}(Petrophysical\ variables) \tag{7}$$

In particular we are interested in estimating the petrophysical variables of porosity ($\varphi$), water saturation ($Sw$) and shaliness ($Sh$) of an investigated area.

The RPM can be defined using empirical approaches, calibrated on data extracted from all available wellbores, or theoretical and general equations, formulated independently from the data (Avseth *et al.* 2010, Grana et Della Rossa 2010). The estimation of elastic properties with the various rock-physics Models represents an essential key, a classic procedure in the characterization of reservoirs. Recent developments then include rock-physics relationships as constraints inside the inversion (Avseth, P., T. Mukerji, and G. Mavko, 2005).

The adopted RPM is successively explained in detail.

## 1.6 Target-Oriented and Interval-Oriented Inversion

In all presented works, we investigate two different AVO inversion approaches: target-oriented and interval-oriented inversion.

The former extracts the AVO response of a single target reflection along a previously interpreted 2D stratigraphic horizon (Mazzotti and Zamboni 2003), while the latter inverts all seismic responses in a given time interval, as reported in Figure (4).

The target-oriented approach relies on two main assumptions: the data are plane waves, and all effects produced during the wave propagation in the subsurface are adequately compensated by a careful seismic processing procedure (Mazzotti and Ravagnan 1995).

Compared to the interval-oriented approach, the target-oriented approach has the main advantage of extremely fast computational cost of the inversion; on the other hand, it requires an accurate geological interpretation to map and identify the horizon of interest through the 3D volume of data, and a very important step is the extraction of the amplitude coefficients from the seismic data. Despite these limitations, this approach has been successfully applied in several exploration projects (Zhu and McMechan 2012, Adriansyah and McMechan 2001).

In all presented works, when the target-oriented approach is applied, we are interested in the reflection generated by a shale/sand interface. Shale acts as an impermeable cap rock that seals a sandy reservoir and prevents upward fluid migration. In this configuration, we are interested in deriving only the elastic-petrophysical properties of the sand below the interface.

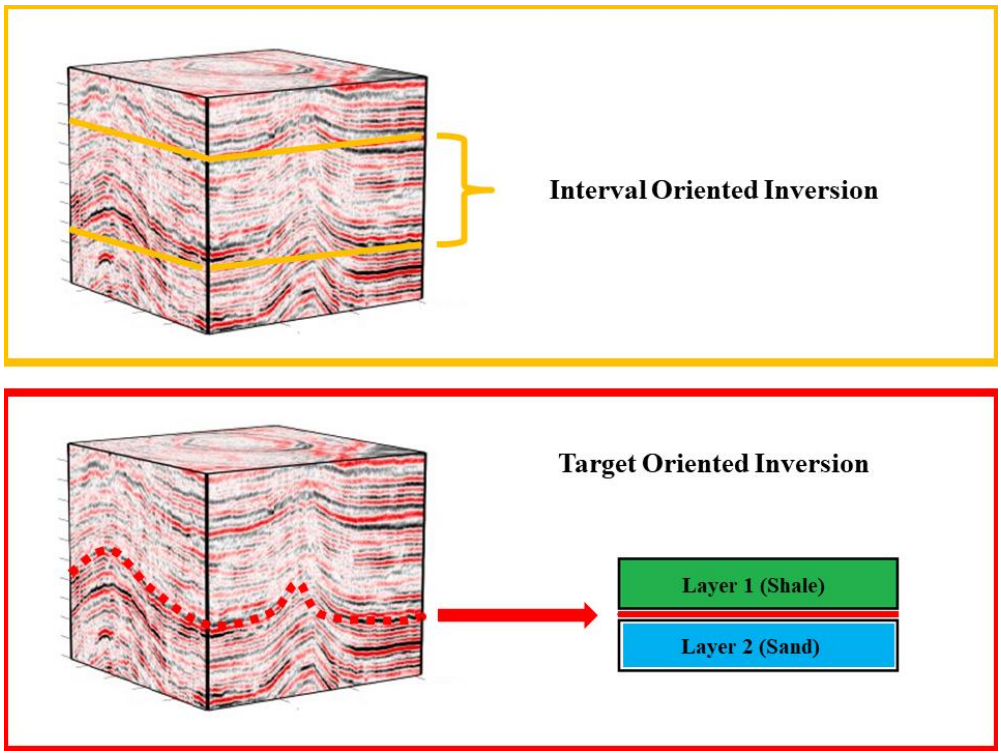


*Figure 4 : The two different AVO inversion strategies adopted.*

## 1.7 Monte Carlo Methods

To address every inversion problem, two main different approaches may be taken: analytical and numerical. The former types, computationally quicker, are valid under two assumptions:

1. Linear forward models.
2. Gaussian or Gaussian-mixture distribution of properties.

When these assumptions are not met, the result is a misfit function that is arbitrarily complex, with several local minima representing possible solutions and absolute minima representing the most likely solutions. The use of classic analytical gradient-based inversion approaches is here generally not recommended because they are easily trapped in local minima. To correctly address these problems, numerical global approach search methods are needed.

There are numerous strategies for global inversions; the simplest and most direct strategy is a systematic model search in all discretized model spaces. Through this method, all models are singularly investigated until the best one is found. Clearly, this approach is feasible in low-

dimensional spaces, i.e., those with few parameters, while it becomes computationally unsustainable as they increase. When we are dealing with highly non-linear, multi-dimensional inverse problems, it is therefore necessary to adopt a different strategy; the most widely used are the numerical Markov chain Monte Carlo methods (McMC).

McMC methods date back to the mid-1940s within the "*Manhattan Project",* but the method was formalized by Nicholas Constantine Metropolis and Stanisław Marcin Ulam (1949). The use of techniques based on the selection of random numbers had, however, already been mentioned in a paper by Lord Kelvin (1901) and in some studies by William Sealy Gosset (1909). In 1953, Nicholas Metropolis was the first to show how to sample a spatial domain using simple probabilistic rules. Since then, the method has been extensively developed and has been applied to a wide range of scientific and mathematical problems. Topic research has placed the McMC among the top ten algorithms with the greatest influence on scientific and engineering development in the twentieth century (Beichl & Sullivan, 2000). Currently, the statistical theory that underlies them constitutes a wide and active research area (Flournoy & Tsutakawa, 1991), and the numbers of geophysical problems to which they have been applied have increased steadily (Jeong *et al.* 2011, Lange *et al.* 2012, Zunino *et al.* 2015, Rimstad & Omre 2010).

McMC algorithms involve an iterative procedure in which a model is generated through a random walk in the model space. This random path satisfies the Markov property of "absence of memory": each generated model depends only on the previous model in the chain. The new model is then accepted or rejected based on some acceptance rule.
Iteration by iteration, the chain collects models spread all over the model space and reaches the stationary regime, as shown in Figure (5). The ensemble of all collected models during the random walk defines the posterior PDF (left term of equation 3). These methods therefore convert the inverse problem into a sampling problem in which the sampling density is proportional to the posterior density function so that the sampled models can be used to approximate the statistical properties of the PDF.

This iterative random walk begins with a starting model that can be arbitrarily set. During the first iterations, the algorithm therefore samples models that cannot be representative of the true posterior distribution (since they are very distant from the true model). To avoid such problems, an accurate choice for burn-in is essential to prevent biased posterior PDFs.

For all global inversions, a careful choice of the starting model is theoretically not required since during its random walk, the algorithm will explore all model space in mathematically infinite iterations. In practice, however, especially in the case of high-dimensional problems, a "good" starting model is extremely important to avoid excessive computational effort.

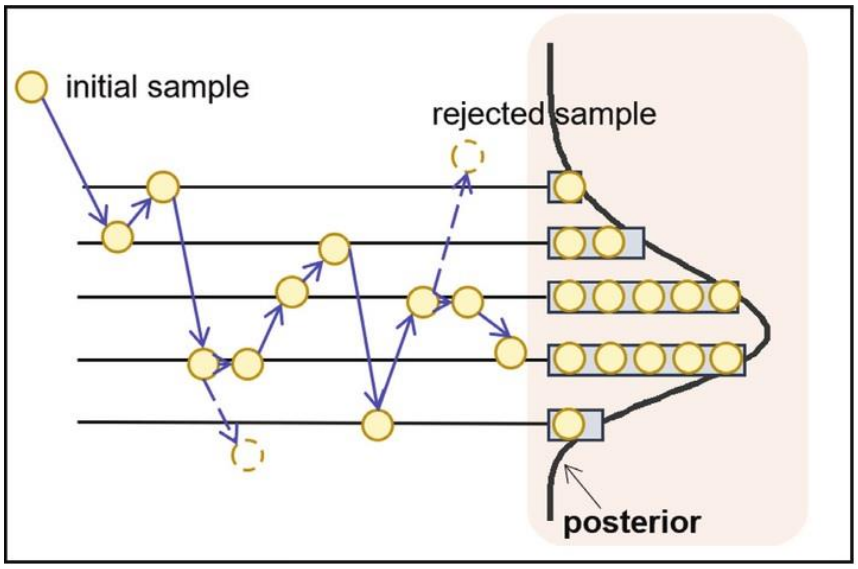


*Figure 5: Graphical description of Monte Carlo sampling. From the starting model, the algorithm proposes subsequent models, which can be accepted or rejected on the basis of the acceptance probability rule. All accepted models sampled at the end of this iterative procedure compose the posterior density function, i.e., the aim of the inversion.*

The main downside of all Markov-chain Monte Carlo algorithms is that they require considerable computational efforts with respect to classic analytical inversion approaches. Even if McMC methods are powerful and effective, they nevertheless involve important issues related to complex posterior distribution estimations involving numerous local optima, long tails and multi-modality. Another crucial problem of all McMC methods is the correlation among adjacent sampled models; it is indeed known that with highly-correlated samples the algorithm convergence is strongly slowed and biased PDFs can be derived (MacKay, 2003).

Several methods have been deployed and applied to overcome this problem. Sen and Biswas (2017) hybridized a transdimensional McMC with a faster Hamiltonian Monte Carlo algorithm, and Lan proposed modifying the Riemannian geometric properties of the target distribution to create wormholes connecting high probability density regions ( Lan, Streets and Shahbaba, 2014). For these reasons, a specific McMC procedure, tailored to the problem at hand, is usually needed to speed up the convergence rate and efficiently sample the PDF.

# CHAPTER 2

# A numerical approach with non-linear forward modelling and non-parametric priors.

In Chapter 2, we face the AVO inversion problem in the case of non-parametric distributions, non-linear forward models and mixed continuous-discrete model properties; under these assumptions, an analytical approach is no longer available, and we therefore adopt a McMC approach. To reduce the ill-conditioning of the problem, we here make use of geostatistical constraints in the form of a variogram model, and to hasten the algorithm convergence, we hybridize the standard McMC approach with the differential evolution Markov chain method and introduce meta-algorithms such as parallel tempering and the delayed rejection scheme.

## 2.1 Introduction

The goal of the reservoir characterization study is to exploit the acquired seismic and well- log data to infer the distribution of the elastic parameters and the litho-fluid facies around an investigated area.

One challenge in this process concerns the simultaneous estimation of discrete (litho-fluid facies) properties and continuous (elastic) properties from seismic data (Gunning and Sams 2018, Larsen *et al*. 2006, Kolbjørnsen *et al*. 2016, De Figueiredo *et al.* 2017, Grana *et al.* 2012). Another important problem is related to the complexity of the property distributions, which is often multi-modal due to the presence of multiple litho-fluid facies. In this context, the assumption of Gaussian-distributed properties is extremely limiting and may not be sufficient to adequately explain the existing relations between the elastic attributes and litho-fluid facies. Furthermore, the linear forward model

might not be sufficiently accurate to describe the relation between seismic data and elastic parameters in cases of strong elastic contrasts at the reflecting interface and far source–receiver offsets (Madsen and Hansen 2018).
Since an analytical approach is no longer applicable under these assumptions, we propose a numerical McMC approach to correctly address these problems and derive the PDF for the elastic variables and litho-fluid facies sought.

To reduce the ill-conditioning inherent the AVO inversion (Avseth *et al.* 2010), we make use of geostatistical constraints in the form of a variogram model. The use of non-parametric distributions often complicates the inclusion of geostatistics in the inversion framework, and for this reason, the use of non-parametric models is not common in geophysical inversions (Sabeti *et al.* 2017, Grana 2018).
Furthermore, classic McMC methods are known to become slow in case of multi-modal target distributions. For this reason, we employ multiple chains and adopt meta-algorithms such as parallel tempering and delayed rejection scheme to exhaustively explore the high-probability regions of the model space.

We therefore present a target- and interval-oriented Bayesian McMC algorithm to solve the mixed continuous-discrete inversion problem for non-linear forward models and non-parametric distributions. To compare the results obtained with those from a classic analytical approach, we benchmark our algorithm with analytical methods that assume Gaussian-mixture distribution of properties and linearized forward models.

## 2.2 The Method

As previously mentioned, McMC algorithms perform a random walk in the model space by applying a two-step procedure: in the first step, the current model starts from a random point and moves towards high-probability regions; in the second step, the small misfit fluctuations related to the sampled models indicate that the stationary regime has been reached.

The acceptance or rejection of each model during this walk is given by a probability that depends on its fit with the observed data. For all presented works, this fit is measured by the Metropolis-

Hastings rule, where the acceptance probability for the perturbed model $\mathbf{m}'$ over the current model $\mathbf{m}$ is given by:

$$\alpha = p(\mathbf{m}'|\mathbf{m}) = \min[1, \text{prior ratio} \cdot \text{likelihood ratio} \cdot \text{proposal ratio}] \quad (8)$$
$$= \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \cdot \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \cdot \frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})}\right]$$

where $q()$ is the proposal distribution that defines the new model $\mathbf{m}'$ as a random perturbation of the current model $\mathbf{m}$. The $p()$ term is the prior information derived from well log data, and $p(\ \ |\ \ )$ is the likelihood distribution for the models. Note that if symmetric Gaussian distributions are employed, the proposal distribution of equation (8) disappears.

Under the assumption of Gaussian uncorrelated white noise contaminating the data, the likelihood function can be expressed as:

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^2|\mathbf{C_d}|}} \exp\left(\frac{-\Phi(\mathbf{m})}{2}\right) \propto \exp\left(\frac{-\Phi(\mathbf{m})}{2}\right) \quad (9)$$

where L is the number of data points, $\mathbf{C_d}$ is the covariance matrix and $\Phi$ is the L2 norm between the observed and predicted data, weighted with the covariance matrix.

To improve the model space exploration and correctly derive posterior distributions, several chains are run starting from different points defined on the basis of the prior distribution. In addition, to ensure good property mixing between the chains, we apply a parallel tempering (PT) simulation method. Each chain is initialized with a temperature *T*, and after a fixed number of iterations, we allow a temperature swap between 2 chains, chosen with equal probability among all running chains.

The Metropolis-Hastings rule becomes:

$$p(\mathbf{m}'|\mathbf{m}) = \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}')^{1/T}}{p(\mathbf{d}|\mathbf{m})^{1/T}} \frac{p(\mathbf{m}')}{p(\mathbf{m})}\right] \quad (10)$$

where the likelihood for a given chain is:

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left\{-\frac{\Phi(\mathbf{m})}{2\,T}\right\}. \qquad (11)$$

According to Sambridge (2013), the probability $p$ to swap the temperature of two running chains identified with models $\mathrm{m}_i$ and $\mathrm{m}_j$ and temperatures $T_\mathrm{i}$ and $T_\mathrm{j}$, respectively, is equal to:

$$p(i,j) = \min\left[1, \frac{\exp\left(-\frac{\Phi(\mathrm{m}_j)}{T_i} - \frac{\Phi(\mathrm{m}_i)}{T_j}\right)}{\exp\left(-\frac{\Phi(\mathrm{m}_i)}{T_i} - \frac{\Phi(\mathrm{m}_j)}{T_j}\right)}\right]. \qquad (12)$$

The temperature acts as a normalizing factor, modifying the shape of the likelihood function (Figure 6). Higher temperatures produce flatter shapes, enabling the chains to freely explore the model space; lower temperatures force the chains to exploit their current model regions. We collect models from chains at temperature = 1 because all other models are representative of biased posterior PDFs (Dosso *et al.* 2014).

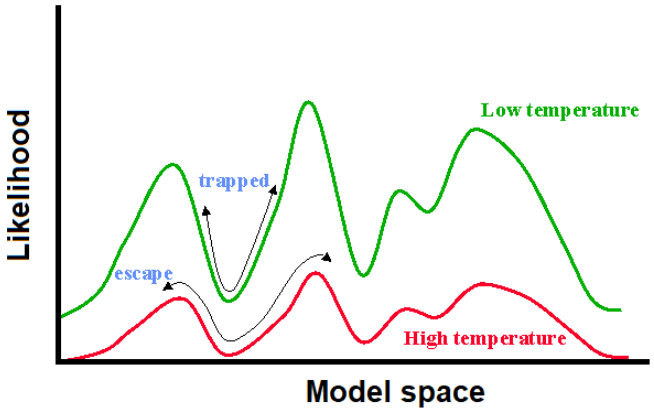


*Figure 6: Effect of parallel tempering on the likelihood function. Higher temperatures allow the chains to easily explore the model space, whereas lower temperatures force the chains to explore those model regions where they become trapped.*

Note that very high temperature values make the likelihood a non-informative parameter (i.e., at T = ∞, the likelihood tends to 0), so the inversion simply follows the prior information (see equation 3).

In addition to the PT method, a delayed rejection scheme (DRS) is adopted here.

Theoretically, McMC convergence is guaranteed for an infinite number of iterations. However, to decrease the computational effort related to these approaches, attention should be paid to the characteristics of the proposal distribution. Persistent model rejection during the random walk, and therefore slower convergence rate, may indeed indicate a badly calibrated proposal distribution. In particular, it is advisable to change the variance of the proposal distribution when we expect the target distribution to have different spreads along the model dimensions (Bodin and Sambridge 2009); in the case of AVO inversion, we expect the PDF to have different spreads along the Vp, Vs and density directions due to the different resolvability of these parameters.

As an alternative to manually tuning the solution, the algorithm is therefore modified so that when a first step is rejected, we propose a second model $\mathbf{m}''$drawn form a second-stage Gaussian proposal $q_2(\mathbf{m}''|\mathbf{m}')$ centred at $\mathbf{m}$ but with a reduced variance. In this case the second proposal is independent on the rejected model $\mathbf{m}'$ and the acceptance term for the second try is simplified as (see appendix of Bodin and Sambridge, 2009):

$$p(\mathbf{m}''|\mathbf{m}) = \min\left[1, \frac{p(\mathbf{m}''|\mathbf{d})}{p(\mathbf{m}|\mathbf{d})} \frac{q_1(\mathbf{m}'|\mathbf{m}'')}{q_1(\mathbf{m}'|\mathbf{m}')} \frac{(1-p(\mathbf{m}'|\mathbf{m}''))}{(1-p(\mathbf{m}'|\mathbf{m}))}\right]. \tag{13}$$

Recalling the classic Bayes formulation (equation 3), identifying the facies as $\boldsymbol{\pi}$ and the elastic variables as $\boldsymbol{e}$, the posterior PDF can be expressed as:

$$p(\boldsymbol{\pi}, \mathbf{e}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{e}, \boldsymbol{\pi})\boldsymbol{p}(\mathbf{e}|\boldsymbol{\pi})\boldsymbol{p}(\boldsymbol{\pi})}{p(\mathbf{d})} = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})} \tag{14}$$

where $\mathbf{m} = [\mathbf{e}, \boldsymbol{\pi}]$.

From all available well log data and geological information about the investigated area, we define the prior distribution for the discrete and continuous properties associated with each considered facies, $p(\boldsymbol{\pi})$ and $p(\mathbf{e}|\boldsymbol{\pi})$, respectively. Since $p(\mathbf{e}|\boldsymbol{\pi})$ is a complex non-parametric mixture distribution in our implementation, we make use of the kernel density estimation algorithm to derive it directly from the available well log data.

Given the ill-conditioning of the AVO inversion under investigation, we make use of geostatistics to introduce a further constraint into the inversion in the form of a variogram model.

The normal score transformation is adopted here to convert the non-parametric distributions into Gaussian models so that geostatistical information about the variability of the elastic parameters can easily be included in the inversion procedure. The prior non-parametric distribution $p(\mathbf{e}|\boldsymbol{\pi})$ is converted to a Gaussian distribution, deriving $p(\mathbf{z}|\boldsymbol{\pi})$, where $\boldsymbol{z}$ represents the normal score transformed elastic variables.

From this Gaussian-mixture model, we can easily calculate the mean, the covariance matrix for all components of the prior distribution and the variogram model that expresses the lateral (for the target-oriented approach) and temporal (for the interval-oriented approach) variability of the elastic parameters.

The variogram is derived by the Kroeneker product between a general correlation function and the prior elastic covariance matrix. The correlation function must be a positive definite function, take values in the interval [-1,1] and have the property that $\nu(0) = 1$. One such correlation function is the second-order exponential function:

$$\nu(\tau) = \exp\left[-\left(\frac{\tau}{d}\right)^2\right] \tag{15}$$

where $\tau$ is the lateral or temporal axis of the autocorrelation function and $d$ is the parameter that defines the lateral/temporal dependency. The lateral amplitude variability of the seismic data, integrated with the available well log information and the geological knowledge about the investigated area, can be used to define the spatial constraints. In interval-oriented inversion, the autocorrelation of available well log data is usually employed to properly set the temporal constraints. Note that the inclusion of lateral constraints finally results in dimensionally increased matrices to invert. This reflects unfeasible computational efforts when 2D or 3D constraints are introduced in the McMC approach.

The adopted approach first helps us preserve the geostatistical constraints in each model sampled during the McMC iterations and second allows us to generate starting models using computationally fast geostatistical simulation techniques such as fast Fourier transform moving averages for continuous properties (Le Ravalec *et al.* 2000) and truncated Gaussian simulations for discrete properties (Matheron *et al.* 1987).

Our McMC steps are as follow:

1. A given CMP position (for target-oriented inversion) or time position (for interval-oriented inversion) is selected.
2. Even iterations: The $\xi$ parameter over [0,1] is defined, and a random number $\kappa$ uniformly distributed over [0,1] is drawn.

2.1 $\kappa > \xi$. We perform an elastic perturbation. The current facies at the selected position is preserved, and the elastic perturbation follows a Gaussian distribution with zero mean and a previously defined covariance matrix.

2.2 $\kappa \leq \xi$. We perform a facies perturbation, thus deriving $\boldsymbol{\pi}$' from $\boldsymbol{\pi}$. The probability of replacing the current facies with the proposed facies at the selected position depends on the lateral (target-oriented) and vertical (interval-oriented) transition matrices (described below). The elastic variables are then realizations of the Gaussian distribution $p(\mathbf{z}|\boldsymbol{\pi}')$.

3. Odd iterations: Differential Evolution Markov Chain *(DEMC)* perturbation. On the current chain *p*, the *i*th time sample or spatial position to be perturbed is selected. Two different chains *a* and *b* that share with the current one the facies on the *i*th position are selected, and the new elastic model is defined as:

$$\mathbf{z}'_{p,i} = \mathbf{z}_{p,i} + \beta\left(\mathbf{z}_{a,i} - \mathbf{z}_{b,i}\right) + \delta N(\mathbf{z}; 0, \boldsymbol{\sigma}_{\mathbf{o}}) \quad with\ a \neq b \neq p \tag{16}$$

where $\beta$ is the jump rate and $\delta$ is a random number <1 (see Vrugt 2016). If no chains share the current facies, we return to step 2.1.

4. To guarantee that all sampled models follow the geostatistical prior constraints, the perturbation of the elastic properties at the *i*th position is propagated to the neighbouring positions (Doyen 2007) through the variogram model.
5. The current transformed model $\boldsymbol{z'}$ is converted to $\boldsymbol{e'}$ through the inverse normal score transformation.
6. Through the Zoeppritz equations, the predicted data on $\boldsymbol{e'}$ are calculated.
7. The Metropolis-Hastings acceptance rule is computed over the current perturbed model. This changes depending on what perturbation has been applied; for the facies perturbation (2.3), the acceptance probability is:

$$p(\mathbf{e}'|\mathbf{e}) = \min\left[1, \frac{p(\mathbf{d}|\mathbf{e}')^{1/T_p}}{p(\mathbf{d}|\mathbf{e})^{1/T_p}}\right] \tag{17}$$

where the *p*th chain has a temperature of *Tp*. For the DEMC perturbation (3) and elastic move (2.2), this becomes:

$$p(\mathbf{e}'|\mathbf{e}) = \min\left[1, \frac{p(\mathbf{d}|\mathbf{e}')^{1/T_p}}{p(\mathbf{d}|\mathbf{e})^{1/T_p}} \frac{p(\mathbf{e}')}{p(\mathbf{e})}\right] \tag{18}$$

8. If the proposed perturbed elastic model is not accepted, we return to the current model ***e*** and apply the DRS to derive ***e''***.
9. If the current perturbed model is not accepted, we set $\boldsymbol{\pi} = \boldsymbol{\pi}$ and **e** = **e**; otherwise, we let $\boldsymbol{\pi} = \boldsymbol{\pi}'$ and **e** = **e'** (or ***e''***=***e***).
10. From all chains at $T_p$=1, the facies and elastic variables generated are collected.
11. The TP method is applied to two randomly chosen chains *p* and *q*.
12. The current elastic model **e** is converted to **z** through the normal score transformation.
13. Steps from 1 to 11 are repeated until a previously defined fixed number of iterations is reached.
14. Discarding all models sampled before burn-in, we use the ensemble of the remaining collected models to perform statistical analysis and to compute the statistical properties of the posterior PDF.

All user-defined hyper-parameters such as $\beta$, $\xi$ or the temperature limits are set to ensure an acceptance ratio of approximately 0.2 - 0.4, which, according to Sambridge and Moseggard (2002), is the optimal rate for McMC methods.

When samples are correlated, not only is the convergence typically slower, but also there is a risk of deriving biased PDF estimations. For this reason, not all the samples collected after the burn-in period are usually used to numerically estimate the posterior, but several iterations of the algorithm are allowed to elapse in between successive samples: this number of iterations is the so-called lag value that can be set from the analysis of the autocorrelation function of the sampled models.

To further reduce the correlation among sampled models, a frequently used approach is to consecutively perturb the elastic properties for different time (interval-oriented inversion) or CMP (target-oriented inversion) positions (Aleardi *et al.* 2018a), repeating steps from 1 to 4 for each chain. This strongly increases the likelihood variation between the current and proposed models.

During sampling, we also promote the lateral continuity of the discrete properties through a first-order Markov model simulation. For example, in the target-oriented approach, the facies transition probability at the spatial position with coordinates (*x*,*y*) with respect to the neighbouring four CMPs with coordinates ($x + 1$, $y$), ($x - 1$, $y$), ($x$, $y + 1$), ($x$, $y - 1$) and the actual facies configuration $h$ is given by:

$$p\left(\boldsymbol{\pi}_{x,y}^{a}\middle|\boldsymbol{\pi}_{x-1,y}^{h},\boldsymbol{\pi}_{x+1,y}^{h},\boldsymbol{\pi}_{x,y+1}^{h},\boldsymbol{\pi}_{x,y-1}^{h}\right) = p\left(\boldsymbol{\pi}_{x-1,y}^{h}\middle|\boldsymbol{\pi}_{x,y}^{a}\right)p\left(\boldsymbol{\pi}_{x+1,y}^{h}\middle|\boldsymbol{\pi}_{x,y}^{a}\right)p\left(\boldsymbol{\pi}_{x,y-1}^{h}\middle|\boldsymbol{\pi}_{x,y}^{a}\right)p\left(\boldsymbol{\pi}_{x,y+1}^{h}\middle|\boldsymbol{\pi}_{x,y}^{a}\right) \quad (19)$$

where each term in the right part of equation (19) is derived by making use of the lateral transition matrix, derived by prior information. The probability of moving from facies $b$ at one neighbouring position to facies $a$ at position ($x$, $y$) can be derived from the $b$th row and $a$th column of the lateral transition matrix $\mathbf{T}l$. For example:

$$p\left(\boldsymbol{\pi}_{x-1,y}^{b}\middle|\boldsymbol{\pi}_{x,y}^{a}\right) = \mathbf{T}_{b,a}^{l}. \quad (20)$$

Similarly, the 1D interval-oriented inversion includes a first-order Markov model to define the probability of a transition from facies $b$ at time sample $i - 1$ to facies $a$ at the $i$th time position.

## 2.3 Analytical Approaches

### 2.3.1 Target-oriented Inversion

For the target-oriented analytical approach, the forward model is represented by the linear Aki-Richards equation (5). Following matrix formalism, this can be rewritten as:

$$\mathbf{d} = \mathbf{Ge} \quad (21)$$

where:

$$\mathbf{d} = [Rpp(\theta_1), \dots, Rpp(\theta_N)]^T \tag{22}$$

$$\mathbf{e} = [R_{Vp}, R_{Vs}, R_\rho]^T \tag{23}$$

$$\mathbf{G} = \begin{bmatrix} \frac{1}{2}(1 + tan^2(\theta_1)) & -4\frac{\overline{Vs}^2}{\overline{Vp}^2} sin^2(\theta_1) & \frac{1}{2}\left(1 - 4\frac{\overline{Vs}^2}{\overline{Vp}^2} sin^2(\theta_1)\right) \\ \frac{1}{2}(1 + tan^2(\theta_2)) & -4\frac{\overline{Vs}^2}{\overline{Vp}^2} sin^2(\theta_2) & \frac{1}{2}\left(1 - 4\frac{\overline{Vs}^2}{\overline{Vp}^2} sin^2(\theta_2)\right) \\ \vdots & \vdots & \vdots \\ \frac{1}{2}(1 + tan^2(\theta_N)) & -4\frac{\overline{Vs}^2}{\overline{Vp}^2} sin^2(\theta_N) & \frac{1}{2}\left(1 - 4\frac{\overline{Vs}^2}{\overline{Vp}^2} sin^2(\theta_N)\right) \end{bmatrix} \tag{24}$$

where *N is* the number of data points, and the $\frac{\overline{Vs}^2}{\overline{Vp}^2}$ values are extracted from all available well log information.

As previously highlighted, the analytical approaches require Gaussian-distributed properties. Under the assumption of a Gaussian mixture, following Grana and Della Rossa (2010), the prior information is formulated as:

$$p(\mathbf{e}) = \sum_{k=1}^{K} \omega_i N\left(\mathbf{e}; \boldsymbol{\mu}_{\mathbf{e}}^k, \boldsymbol{\Sigma}_{\mathbf{e}}^k\right), \quad (6) \tag{25}$$

where *k* is the Gaussian component of the mixture, $\omega_i$ is the prior weight associated with the *i*th component (the prior probability for the *i*th facies) and *N* is the Gaussian distribution characterized by mean and a covariance matrix calculated on each facies; we therefore assume that all properties are Gaussian-distributed within each facies. From well logs and all other prior information, we derive the total number of considered facies and the associated statistical properties.

Since we are making use of a linear forward operator, the posterior distribution is again a Gaussian mixture defined as:

$$p(\mathbf{e}|\mathbf{d}) = \sum_{k=1}^{K} \lambda_i N\left(\mathbf{e}; \boldsymbol{\mu}_{\mathbf{e}|\mathbf{d}}^k, \boldsymbol{\Sigma}_{\mathbf{e}|\mathbf{d}}^k\right) \tag{26}$$

where $\lambda_i$ is the posterior weights, and the posterior mean ( $\boldsymbol{\mu}_{\mathrm{e|d}}^{k}$) and covariance matrices associated with each facies ($\boldsymbol{\Sigma}_{\mathrm{e|d}}^{k}$) are derived as follows:

$$\boldsymbol{\mu}_{\mathbf{e|d}}^{k} = \boldsymbol{\mu}_{\mathbf{e}}^{k} + \left(\mathbf{G}^{T}\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\mathbf{G} + \left(\boldsymbol{\Sigma}_{\mathbf{e}}^{k}\right)^{-1}\right)^{-1}\mathbf{G}^{T}\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\left(\mathbf{d} - \mathbf{G}\boldsymbol{\mu}_{\mathbf{e}}^{k}\right) \tag{27}$$

$$\boldsymbol{\Sigma}_{\mathbf{e|d}}^{\boldsymbol{k}} = \left(\mathbf{G}^{\boldsymbol{T}}\boldsymbol{\Sigma}_{\mathbf{d}}^{-\mathbf{1}}\mathbf{G} + \left(\boldsymbol{\Sigma}_{\mathbf{e}}^{\boldsymbol{k}}\right)^{-\mathbf{1}}\right)^{-\mathbf{1}} \tag{28}$$

$$\lambda_i = \frac{\omega_i N(\mathbf{d}; \boldsymbol{\mu}_{\mathbf{d}}^{k}, \boldsymbol{\Sigma}_{\mathbf{d}}^{k})}{\sum_{i=1}^{K} \omega_i N(\mathbf{d}; \boldsymbol{\mu}_{\mathbf{d}}^{k}, \boldsymbol{\Sigma}_{\mathbf{d}}^{k})} \tag{29}$$

Finally, following De Figueiredo *et al.* (2017), the posterior mean for the elastic model can be derived as a weighted sum of all the Gaussian components:

$$\mathbf{e}_{est} = \sum_{k=1}^{K} \lambda_i \boldsymbol{\mu}_{\mathbf{e|d}}^{k} \tag{30}$$

Since the inclusion of a 2D geostatistical model to constrain the inversion would greatly increase the computational effort (Aleardi *et al.* 2018b), each AVO extracted by CMP gathers is inverted independently.

### 2.3.2 Interval-oriented Inversion

Following Buland and Omre (2003), for the interval-oriented approach, the forward modelling is the temporal extension of equation (5), where the convolutional model that generates the observed data **d** can be reformulated as:

$$\mathbf{d} = \mathbf{SADe} = \mathbf{Ge} \tag{31}$$

where the matrix **S** contains the wavelet for each incidence angle, **A** contains the coefficients $\alpha Vp(t)$, $\alpha Vs(t)$ and $\alpha\rho(t)$ of equation (5), **D** is the first-order derivative operator and **e** represents the logarithm of the elastic properties. Again, we assume that all properties have Gaussian distributions within all facies.

The adopted approach is derived by De Figueiredo *et al.* (2018), to which the interested reader is referred to further details. Following this approach, the posterior mean, covariance matrix, and posterior weights are obtained as:

$$\boldsymbol{\mu}_{\mathbf{e}|\mathbf{d},\mathbf{f}}^{k} = \boldsymbol{\mu}_{\mathbf{e}}^{k} + \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} \\ \mathbf{G}\boldsymbol{\Sigma}_{\mathbf{e}}^{k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} + \boldsymbol{\Sigma}_{\mathrm{f}} & \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G}^{T} \\ \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G} & \mathbf{G}^{T}\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\mathbf{G} \end{bmatrix}^{-1} \left( \begin{bmatrix} \mathbf{f} \\ \mathbf{d} \end{bmatrix} - \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{e}}^{k} \\ \mathbf{G}\boldsymbol{\mu}_{\mathbf{e}}^{k} \end{bmatrix} \right) \tag{32}$$

$$\boldsymbol{\Sigma}_{\mathbf{e}|\mathbf{d},\mathbf{f}}^{k} = \boldsymbol{\Sigma}_{\mathbf{e}}^{k} - \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} \\ \mathbf{G}\boldsymbol{\Sigma}_{\mathbf{e}}^{k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} + \boldsymbol{\Sigma}_{\mathrm{f}} & \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G}^{T} \\ \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G} & \mathbf{G}^{T}\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\mathbf{G} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} \\ \mathbf{G}\boldsymbol{\Sigma}_{\mathbf{e}}^{k} \end{bmatrix}^{T} \tag{33}$$

$$\lambda_i = \frac{\omega_i N\left(\begin{bmatrix} \mathbf{f} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{e}}^{k} \\ \mathbf{G}\boldsymbol{\mu}_{\mathbf{e}}^{k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} + \boldsymbol{\Sigma}_{\mathrm{f}} & \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G}^{T} \\ \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G} & \mathbf{G}^{T}\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\mathbf{G} \end{bmatrix}\right)}{\sum_{i=1}^{K} \omega_i N\left(\begin{bmatrix} \mathbf{f} \\ \mathbf{d} \end{bmatrix}; \begin{bmatrix} \boldsymbol{\mu}_{\mathbf{e}}^{k} \\ \mathbf{G}\boldsymbol{\mu}_{\mathbf{e}}^{k} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}}^{k} + \boldsymbol{\Sigma}_{\mathrm{f}} & \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G}^{T} \\ \boldsymbol{\Sigma}_{\mathbf{e}}^{k}\mathbf{G} & \mathbf{G}^{T}\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\mathbf{G} \end{bmatrix}\right)} \tag{34}$$

where the vector **f** represents the low-frequency elastic model and the matrix $\Sigma_{\mathrm{f}}$ represents the relative uncertainty. The prior covariance matrix is here obtained by a Kroeneker product between the stationary covariance matrix, expressing the correlation among the elastic properties, and a first-order exponential function (Buland and Omre 2003).

Similar to the target-oriented approach, this approach derives the posterior elastic distribution as a weighted sum of all components, which are again assumed Gaussian based on the linear forward operator.

## 2.4 Inversion Results

### 2.4.1 Target-oriented Inversions

The 3D synthetic discrete facies model is derived using the Gaussian simulation algorithm (Matheron *et al.* 1987), while the Fast Fourier Transform moving average (Le Ravalec *et al.* 2000) is used to define the continuous elastic properties associated with each facies from the prior distribution.

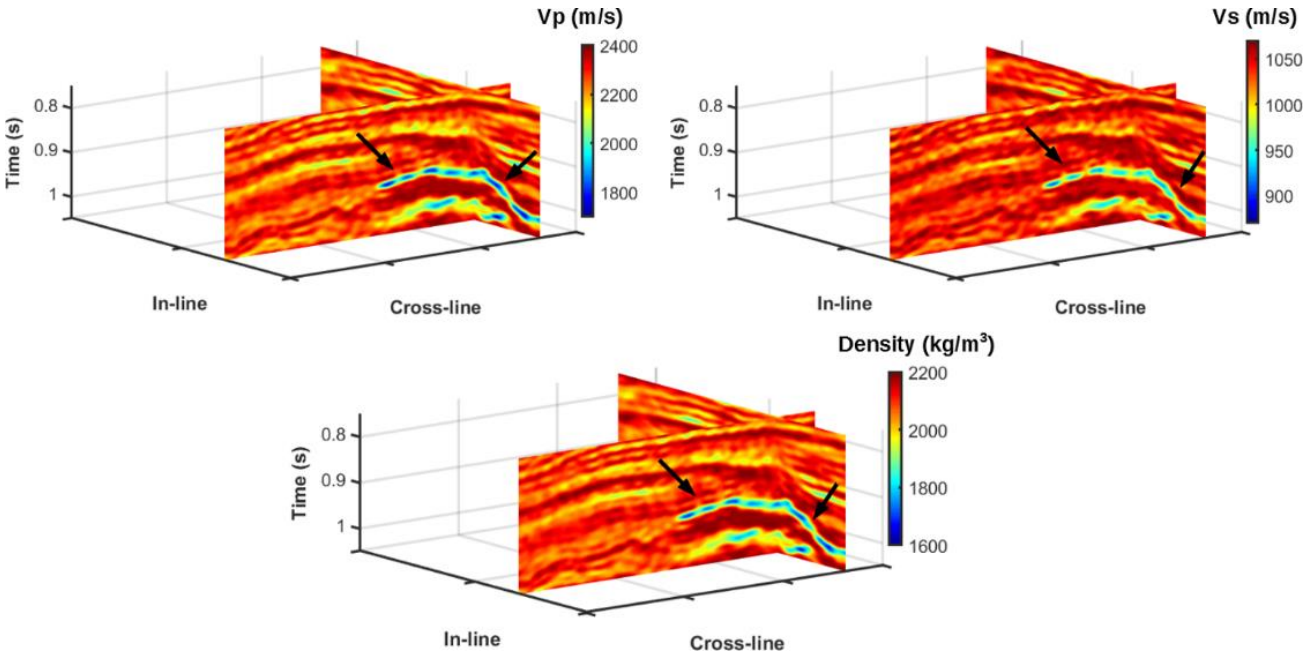


*Figure 7: Example of in-line and cross-line sections extracted from the simulated 3D models of Vp, Vs and density. From this 3D model, we extract the 2D reference model for the target-oriented inversion, which represents the elastic reflectivity contrasts at the top of the target, gas-saturated layer (indicated by the black arrows).*

From this reference 3D model, we extract a 2D time slice, reported in Figure 7, referring to the top of a clastic reservoir located in the porous portions of clayey-sandy sequences (black arrows in Figure 7).

In this complex reservoir model, 3 facies are considered: gas-sand, brine-sand and shales. The sandy portions of interest are characterized by strong decreases in Vp, Vs and density reflectivities (respectively $R_{Vp}, R_{Vs}, R_{\rho}$) with respect to the encasing shales, as illustrated in Figure 8.

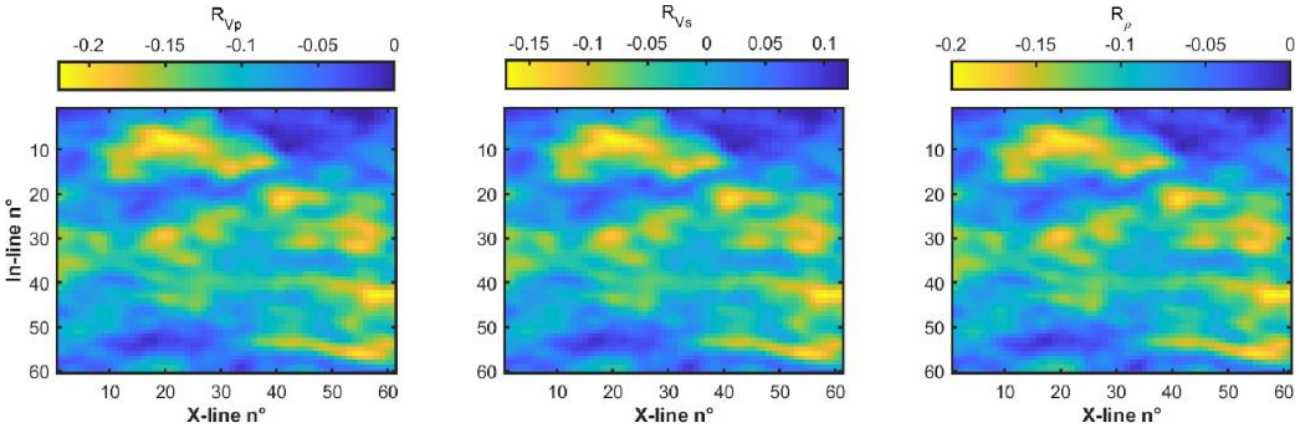


*Figure 8: The reference model for the target-oriented inversion that represents the elastic reflectivity contrasts at the top of the considered reservoir interval (see also Figure 7). From left to right we represent* $\boldsymbol{R_{Vp}, R_{Vs}, R_{\rho}}$.

These elastic contrasts are then feedforwarded to the forward model and the relative AVO responses, representing the observed data, are calculated analytically. To better simulate the field data, these AVOs are subsequently contaminated with white Gaussian-distributed noise. To allow a comparison between the analytical and numerical approaches, both algorithms consider an incidence angle range of 0-30 [°].

Figure 9 shows the prior marginal distributions for the elastic reflectivities within each litho-fluid facies extracted by well logs. Even through a qualitative analysis, we note high similarities between the non-parametric marginal distributions derived with the kernel density estimator algorithm and the Gaussian-mixture distributions, as presented in Figure 9.

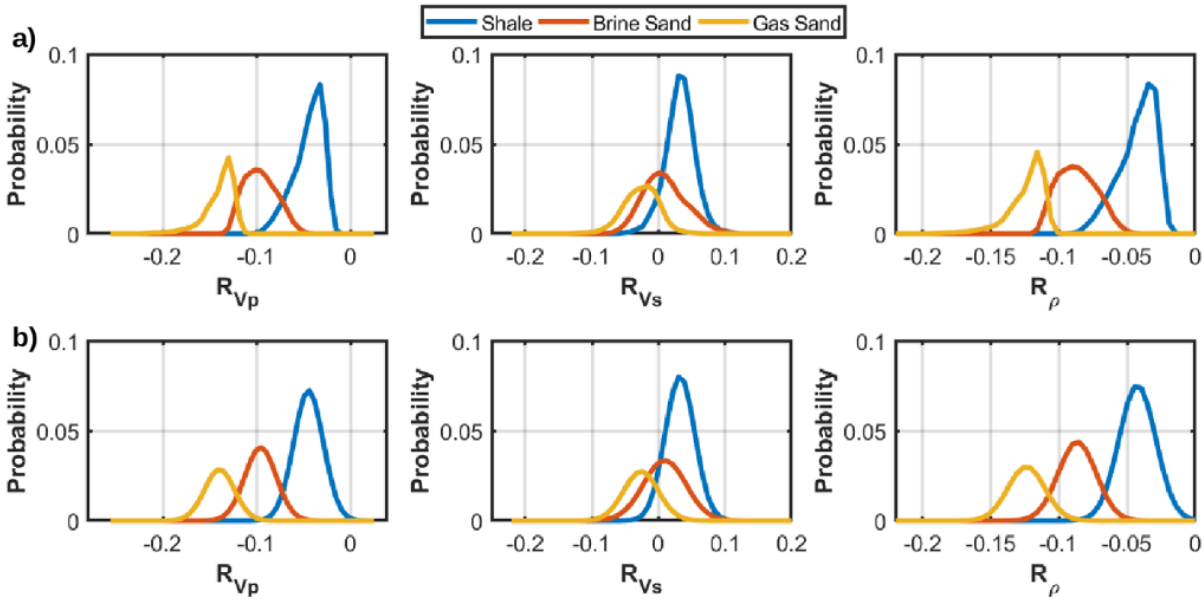


*Figure 9: Prior marginal distributions for the elastic reflectivities within each litho-fluid facies derived from the available well log information. a) The prior non-parametric distributions derived through the kernel density estimation algorithm. b) The Gaussian-mixture prior model.*

The normal probability plot calculated before and after the normal score transformation ( Figure 10), indicate that the Gaussian-mixture assumption for the elastic and litho-fluid facies under investigation is definitively acceptable.

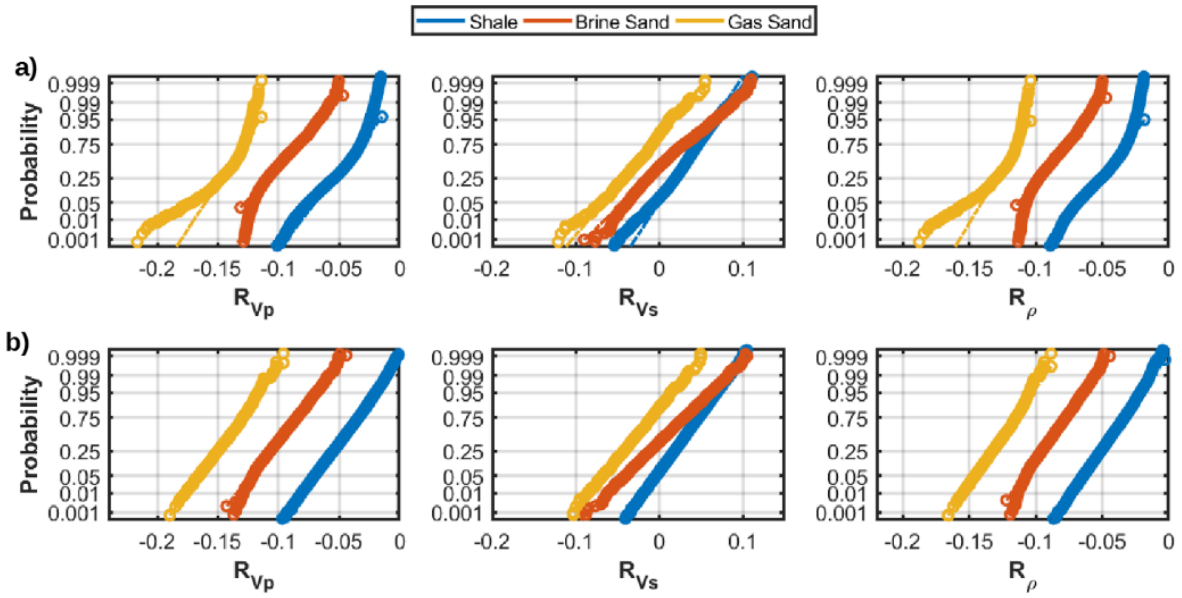


*Figure 10: a) Normal probability plots for each elastic reflectivity and litho-fluid facies derived from the available well log information around the target. b) Normal probability plots derived after the normal-score transformation of a). In these plots, the dotted lines represent the theoretical Gaussian distribution, whereas the circles represent the actual data.*

For the target oriented approach, we use the autocorrelations of seismic stack amplitudes to derive the spatial constraints expressing the lateral variability of the elastic properties. The inclusion of the variogram model, derived from the autocorrelation functions, in the inversion process allows inserting lateral constraints on the elastic models sampled during the McMC random walk. For this reason, the aim of the following inversion tests is twofold: to demonstrate the suitability of the implemented McMC approach for target-oriented AVO inversion and to prove that the lateral constraints included in the MCMC inversion provide more realistic and stable predictions in the case of low signal-to-noise (S/R) ratios.

We employ 40 different chains, 20 identified by temperature T=1 and the remaining 20 by logarithmically spaced temperature values ranging from 10 to 500, to ensure an optimal acceptance ratio (Dosso *et al.* 2012). To decrease the correlation among sampled models, we perturb the continuous-discrete properties at 20 CMP gather positions before evaluating the likelihood function. The McMC algorithm runs for 400,000 iterations, and the burn-in is set to 200,000 with a lag of 30. All these hyper-parameters are determined after a cautious analysis of the likelihood evolution for all running chains and the autocorrelation function for all model parameters (see next section).

The lateral transition matrix, used for both the in-line and cross-line directions, employed by the McMC inversion algorithm is:

$$\mathbf{T}^{l} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix} \tag{35}$$

This describes the transition probabilities for all facies configurations. Along rows, from top to bottom, there are shale, brine-sand and gas-sand, while along columns, from left to right, there are shale, brine-sand and gas-sand. Note the higher values along the main diagonal of the matrix that tend to preserve the current litho-fluid facies for the sampled models.

To explore the algorithm performances and predictions, we run two inversions with different S/N ratios affecting the data.

In the first inversion run, we simulate an ideally optimum S/N ratio of 50. In this case, both analytical and numerical approaches derive similar predictions of the elastic and litho-fluid facies. Figure (11) shows comparable posterior mean models for the elastic variables for both methods, while figures demonstrate that both approaches also achieve similar results for the maximum posterior (MAP) solutions and the PDFs for facies classification.

Nevertheless, we note that in this ideal scenario, with almost Gaussian-distributed elastic reflectivities and low data noise, the analytical approach erroneously predicts brine sand in the portion highlighted by green circles in Figure (12). The McMC is able instead to correctly identify these circular small areas of clayey portions, as shown in Figure (13).

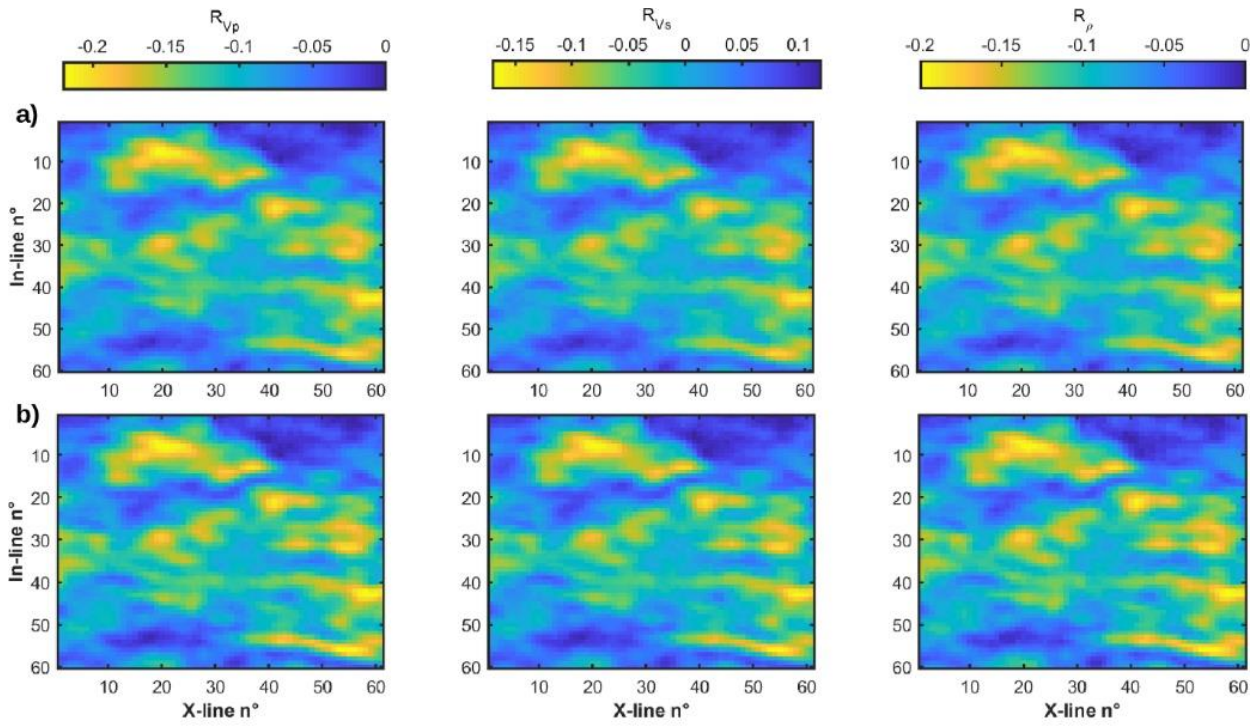


*Figure 11: Results for a high S/N ratio equal to 50. a) Mean posterior models predicted by the analytical inversion. b) Mean posterior models predicted by the MCMC inversion. In a) and b)* $\boldsymbol{R_{Vp}, R_{Vs}, R_{\rho}}$ *are represented from left to right.*

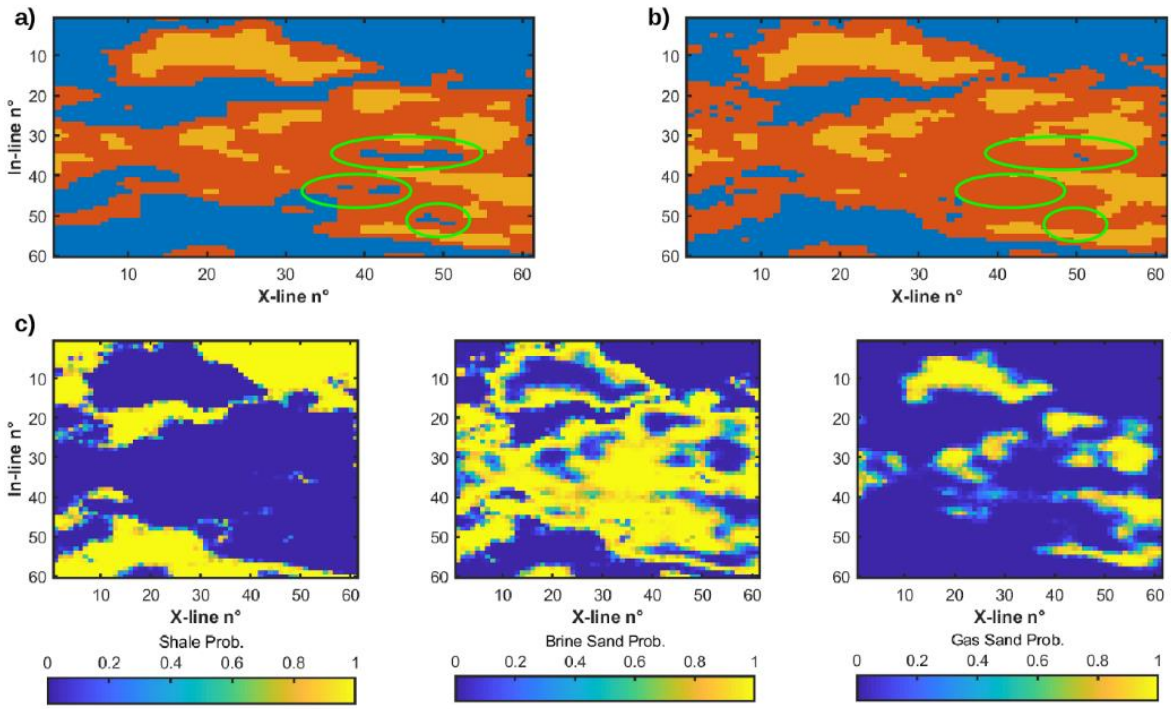


*Figure 12: Facies classification results provided by the analytical inversion for a high S/N ratio equal to 50. a) True facies model. b) MAP facies solution. c) Posterior probability for shale (left), brine sand (centre) and gas sand (right). The circles in (a) and (b) highlight some examples of misclassifications.*

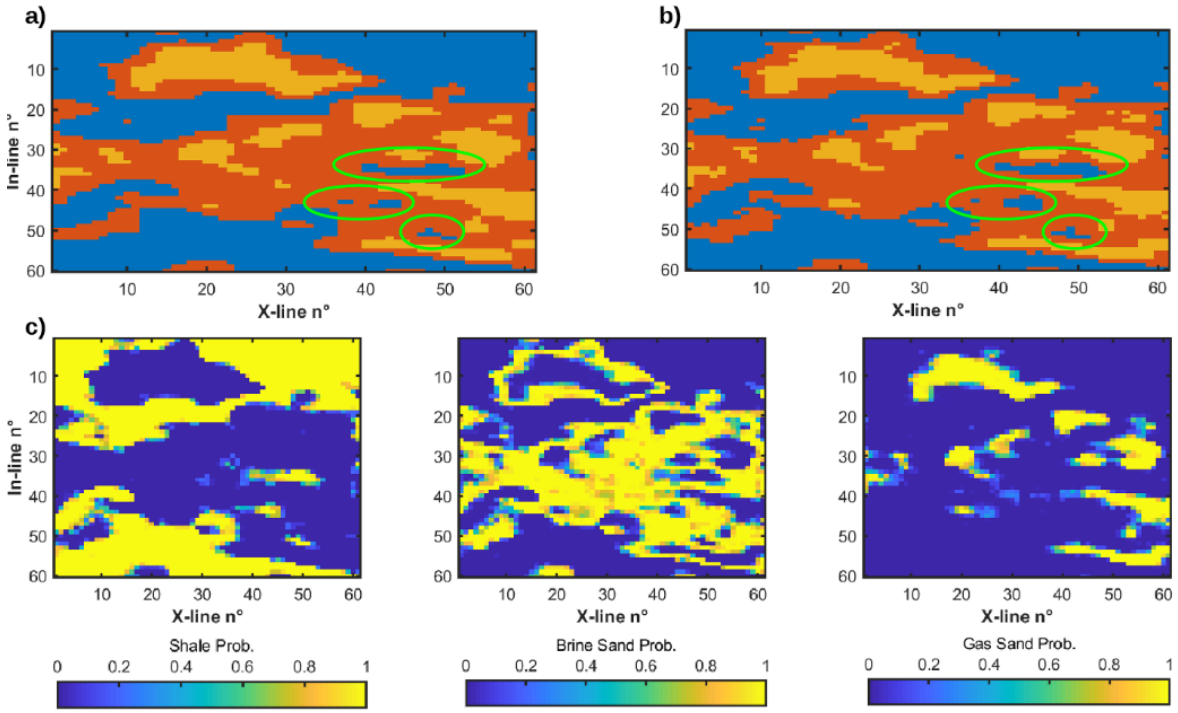


*Figure 13: Facies classification results provided by the McMC inversion for a high S/N ratio equal to 50. a) True facies model. b) MAP facies solution. c) Posterior probability for shale (left), brine sand (centre) and gas sand (right). The circles in (a) and (b) highlight the same areas marked in Figure 12. Here, note the superior classification results provided by the McMC method.*

In the second inversion run, we contaminate the AVO data with higher Gaussian-distributed noise, decreasing the S/N ratio to 2. In this noisy scenario, we expect superior performance from the numerical approach since the inclusion of lateral constraints has been neglected for the analytical method (each CMP is hence independently inverted).

Figure (14) indeed shows that the latter approach produces a much more scattered estimation of the elastic variables, and the lateral boundaries of the considered facies are difficult to identify. In contrast, the McMC approach produces elastic properties with more preserved lateral continuity.

Figure (15) shows that the match between the predicted and true facies models is decreased, but analysing Figure (16), we again note that the analytical approach generates much less stable and much more scattered results.

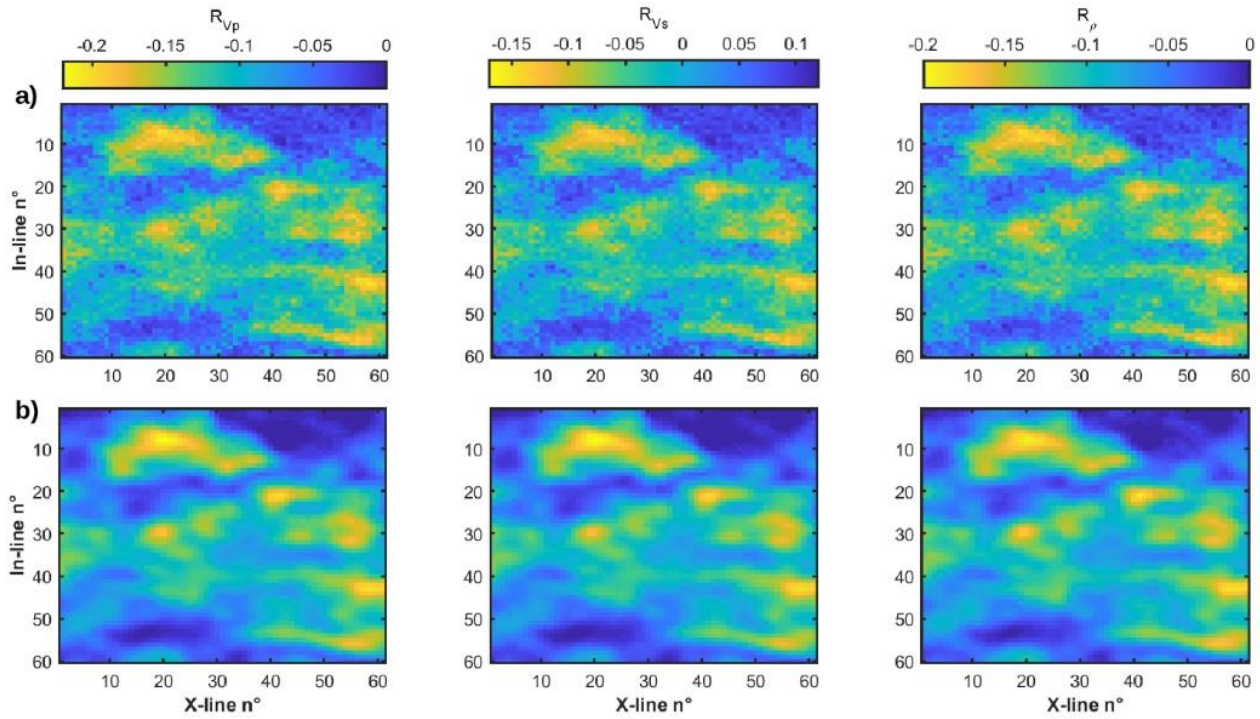


*Figure 14: As in Figure 6 but for a S/N ratio equal to 2. a) and b) refer to the analytical and McMC estimated mean models, respectively. The* $\boldsymbol{R_{Vp}}, \boldsymbol{R_{Vs}}, \boldsymbol{R_{\rho}}$ *are represented from left to right.*

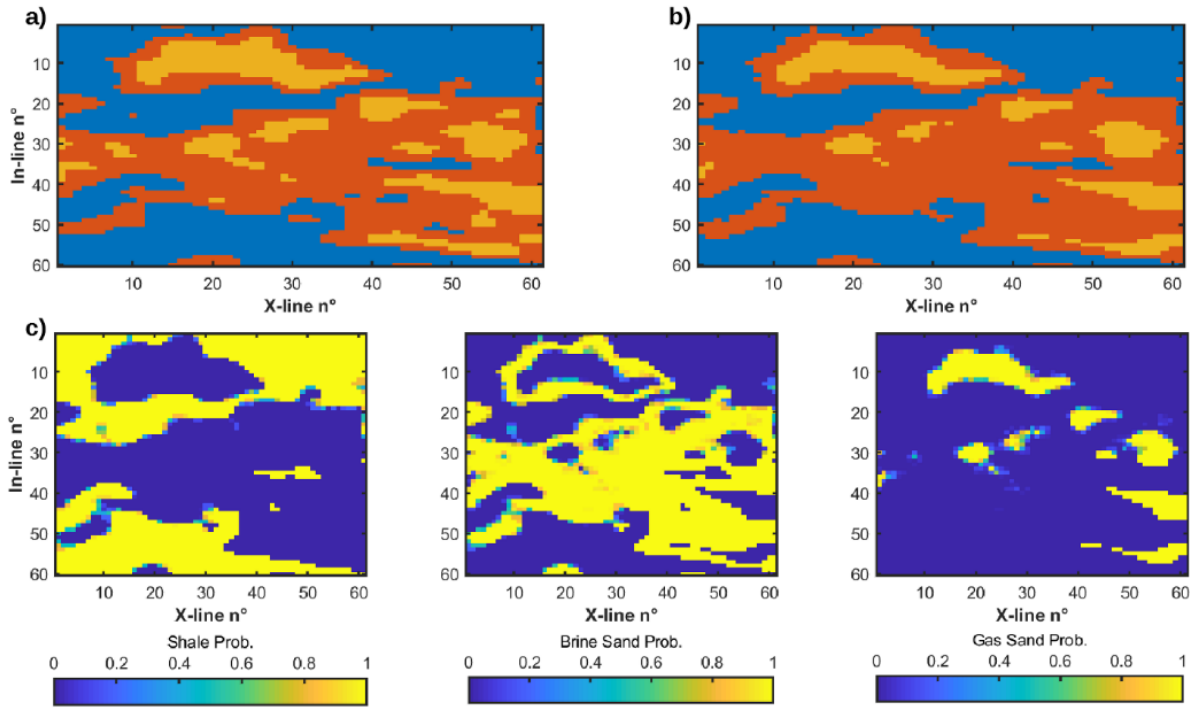


*Figure 15: Facies classification results provided by the McMC inversion for a S/N ratio equal to 2. a) True facies model. b) MAP facies solution. c) Posterior probability for shale (left), brine sand (centre) and gas sand (right).*

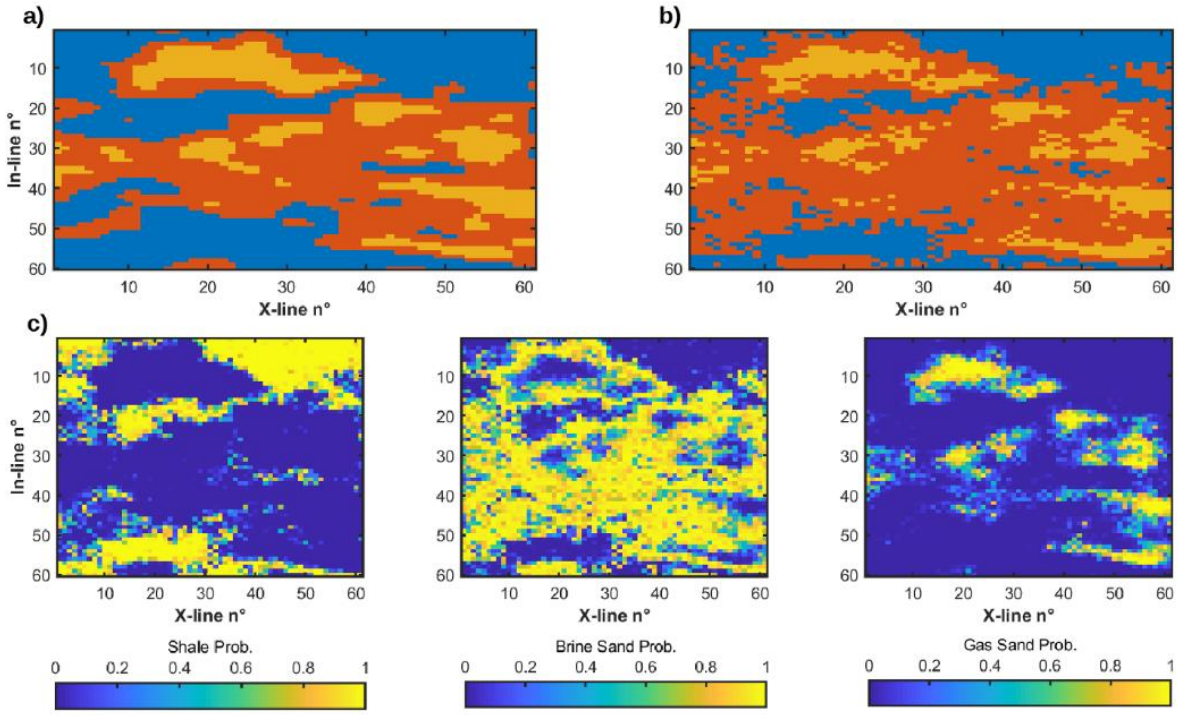


*Figure 16: Facies classification results provided by the analytical inversion for a S/N ratio equal to 2. a) True facies model. b) MAP facies solution. c) Posterior probability for shale (left), brine sand (centre) and gas sand (right).*

Figure (17) shows two AVOs extracted from different CMP gathers, and despite the severe noise contamination affecting the AVO responses, both approaches match the noisy observed data.

By computing the L2 norm data misfit for both approaches over the same CMP gather, we derive values of 0.0662 (bottom left of Figure 17) and 0.0741 (bottom right of Figure 17) for the analytical method, whereas we obtain 0.0668 (top left of Figure 17) and 0.0757 (top right of Figure 17) for the McMC algorithm. The slightly higher misfit associated with the numerical approach can be ascribed to the lateral constraints nested in the numerical approach, which act as regularization terms, stabilizing the inversion and slightly increasing the data misfit.

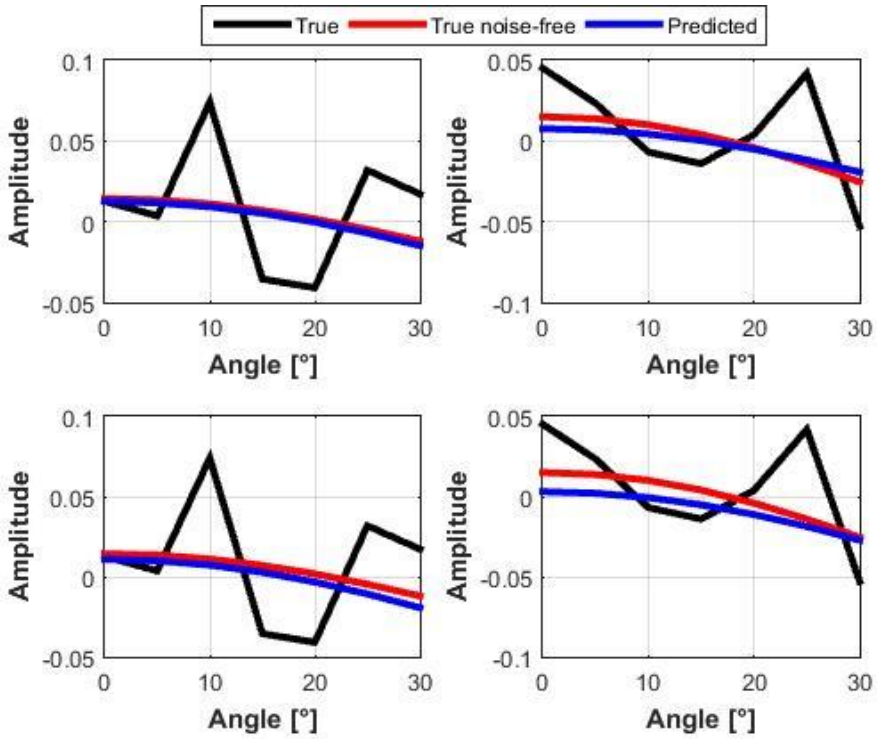


*Figure 17: Examples of observed noisy AVO responses (black lines), noise-free AVO responses (red lines) and predicted AVO responses (blue lines) computed for the MAP solutions provided by the analytical inversion (down) and the MCMC inversion (up) in the case of S/N=2.*

Figure (18) shows the negative log-likelihood evolution for 5 of 40 running chains and indicates that all chains attain a stable likelihood after approximately 200,000 iterations; therefore, all models collected before this value are not considered in the estimation of the posterior distributions. The beneficial effects of the PT and DRS approaches to reach the stationary regime are highlighted by the black arrows, where a sharp decrease in the likelihood values associated with the running chains is more evident. In these portions, how the temperature swap between chains allows better and faster model exploration is clearer.

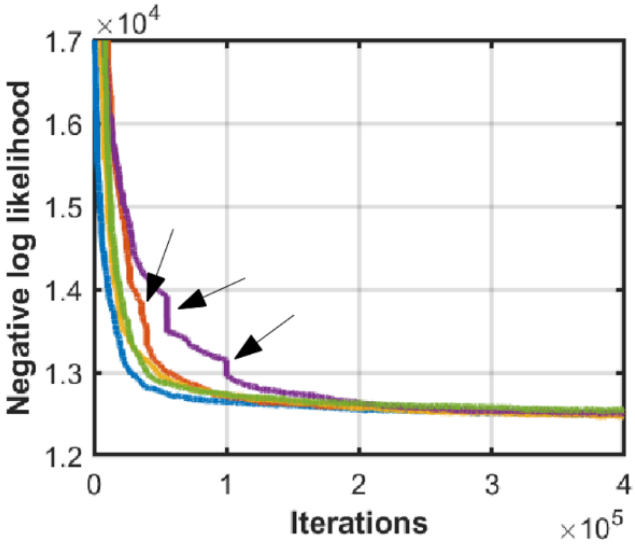


*Figure 18: Examples of evolutions of the negative log-likelihood for 5 of 40 interactive chains. The black arrows highlight the beneficial effect of the PT strategy. Different colours represent different chains.*

To quantitatively assess the algorithm convergence performance and stability, we monitor the evolution of the potential scale reduction factor (PSRF) for different model parameters (elastic reflectivity estimated at different CMP locations). Examples of PSRF values for 6 CMP positions are presented in Figure (19). We note that only approximately 20,000 iterations are needed to attain reliable posterior PDF estimations for $R_{Vp}$, whereas 200,000 iterations are usually required to accurately estimate the posterior PDFs for $R_{Vs}$ and $R_{\rho}$. These different convergence rates are obviously related to the different influences that the elastic parameters exert on the observed data. Indeed, $R_{Vp}$ plays a major role in determining the observed P-wave reflection coefficients, while $R_{Vs}$ and $R_{\rho}$ exert weaker influences.

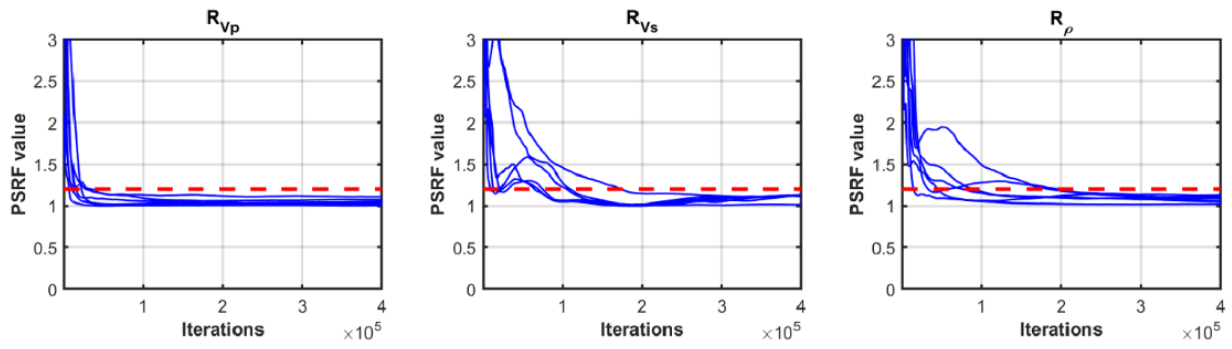


*Figure 19: Examples of the evolution of the PSRF value for the elastic reflectivities pertaining to 6 CMP gathers. The dotted red lines show the desired PSRF value of 1.2. In each plot, the blue lines indicate model parameters estimated at different CMP gather positions.*

Finally, Figure (20) presents some examples of normalized autocorrelation functions computed for the sampled $R_{Vp}$, $R_{Vs}$ and $R_{\rho}$ models. To set the lag value of 30 used in the previous inversion tests, we simply count the number of consecutive models requested to obtain autocorrelation values below 0.5 – 0.4. This threshold value is assumed to represent a good compromise between reliable posterior assessments and the overall computational cost of the inversion procedure. A lower threshold value would guarantee more accurate PDF estimations but at the expense of extra computational cost required to sample a number of models sufficient to reliably evaluate the posterior. The analysis of the evolution of the negative log-likelihood value, of the PSRF and of the autocorrelation function is crucial to determine the number of McMC iterations necessary to attain accurate posterior PDF estimations.

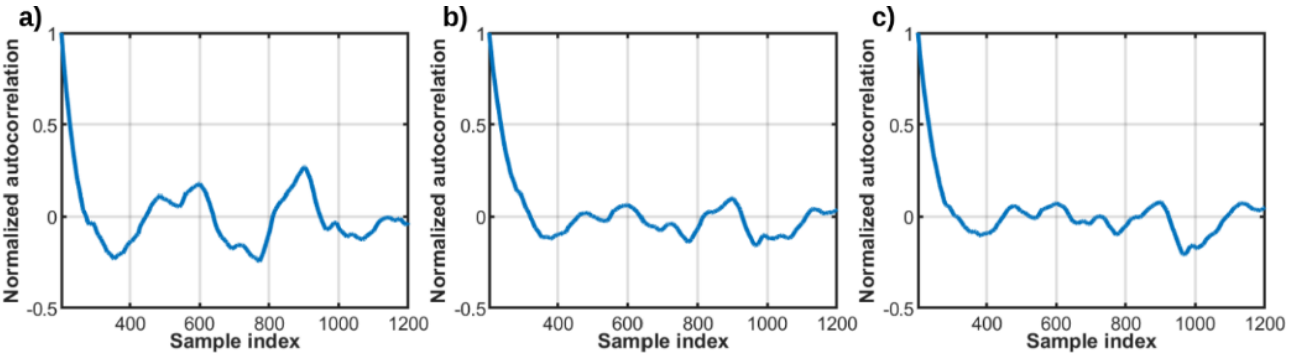


*Figure 20: Close-up of the normalized autocorrelation functions derived from the $\boldsymbol{R_{Vp}}$ (a), $\boldsymbol{R_{Vs}}$ (b), and $\boldsymbol{R_{\rho}}$(c) vales sampled at a given CMP gather position.*

Let us now briefly discuss the computational efforts; the analytical inversion takes less than 5 [s] if we consider a serial code running on an Intel i7-7700HQ processor at 2.8 GHz with 16 Gb RAM. The McMC algorithm takes approximately 40 minutes with a parallel code running on two computer nodes equipped with two deca-core Intel E5-2630 processor at 2.2 GHz (128 Gb RAM). As expected, the McMC method clearly retrieves highly accurate uncertainty appraisals but at the expense of increased computational effort.

### 2.4.2 Interval-oriented Inversions

For this approach, 5 of 7 available well log data define the prior elastic and litho-facies distributions. The elastic information extracted from the remaining 2 well logs (used for blind well tests) is used in the interval-oriented extended Aki-Richards equations (Buland and Omre, 2003) to

generate the synthetic CMP gathers over which the analytical and numerical approaches are benchmarked. Unlike the previous inversion tests, these inversions have characteristics that make the process more challenging and realistic, since the prior distributions are derived from a subset of all available information and not from all information used to define the true model.

As in the target-oriented approach, to better simulate the field data, all CMP gathers are contaminated with Gaussian-distributed white noise. The convolutional process is simulated using a 55 Hz Ricker wavelet and an angle range of 0-30 [°].

Analogous to the previous approach, this approach, illustrated in Figure (21), shows the non-parametric and Gaussian-mixture marginal distributions for the elastic properties associated with each litho-fluid facies. In contrast to the previous approach, in this case, we note that despite the high similarities associated with the clayey sequences, the actual non-parametric brine sand and gas sand facies show strong skewness and differ significantly from the Gaussian-mixture distributions. This means that the Gaussian-distribution assumption is no longer verified; therefore, we expect the analytical approach to show significantly lower performances than the McMC.

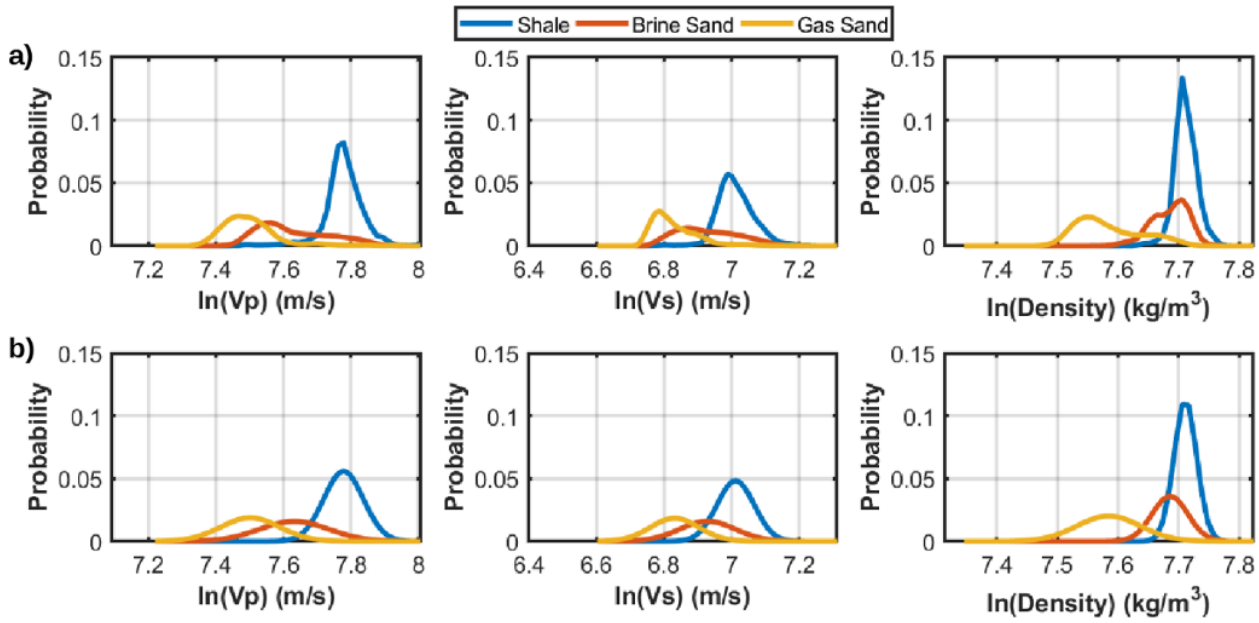


*Figure 21: Prior models for the elastic parameters and for each litho-fluid facies derived from 5 of 7 available wells. a) The prior non-parametric model derived through the kernel density estimation algorithm. b) The Gaussian-mixture prior model. In both a) and b), we consider the natural logarithm of the elastic properties.*

This strong skewness can also be observed in the normal probability plot of Figure (22), where the major differences appear in the Vp and Vs distributions associated with the gas and brine sands. Despite this factor, note how these deviations disappear after the application of the normal score transformation to all properties.

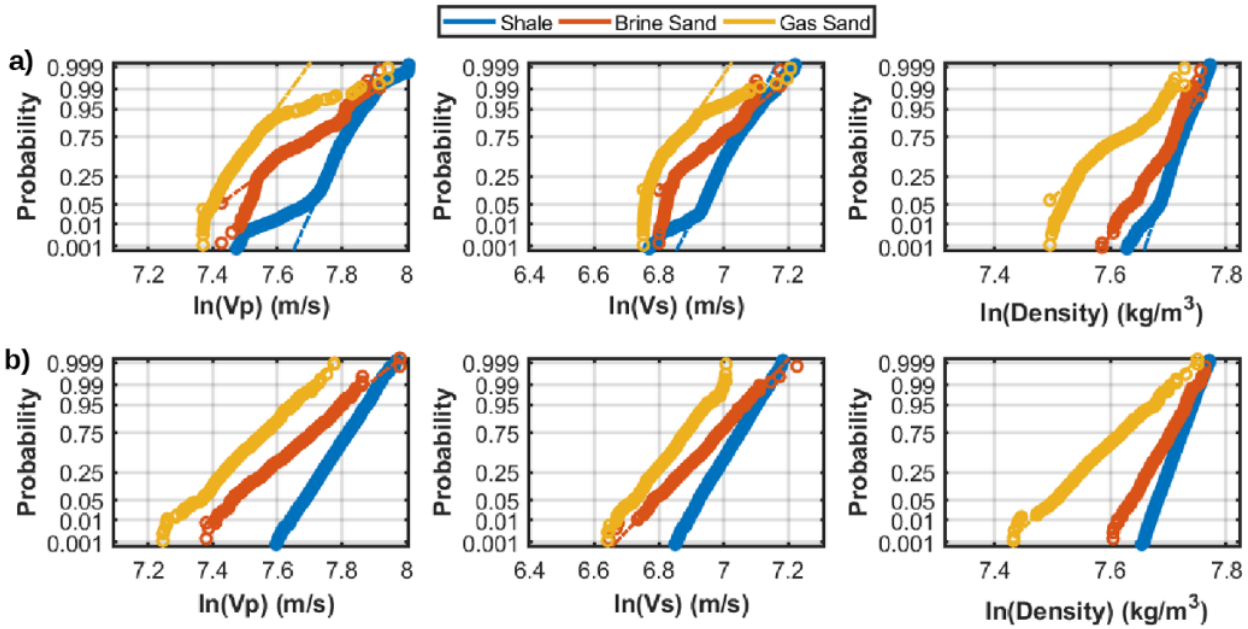


*Figure 22: a) Normal probability plot derived from the actual well log data pertaining to 7 available wells. b) Normal probability plot derived on the normal score transformed from actual well log data. In both a) and b), we consider the natural logarithm of the elastic properties. In these plots, the dotted lines represent the theoretical Gaussian distribution, whereas the circles represent the actual well log data.*

For the interval-oriented approach, we run 40 different chains for 10,000 iterations with a burn-in value set to 5,000. Twenty chains have temperature $T = 1$, and the remaining chains have associated temperature values ranging in a logarithmically spaced interval. To decrease the correlation between the current and the perturbed models, the elastic properties are perturbed at 10 different time positions before the evaluation of the likelihood function.

From the analysis of the available well log data, the vertical transition probability matrix can be derived, describing the transition probabilities among the available facies:

$$\mathbf{T}^{v} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.3 & 0.7 & 0 \\ 0.1 & 0.4 & 0.5 \end{bmatrix} \tag{36}$$

Note that since gas has smaller density compared to brine, unphysical transition probabilities (brine-sand/gas-sand configuration) are avoided with null values.

Figures (23) and (24) show the inversion results obtained on the first blind test with both approaches, where the synthetic CMP gather has been contaminated with a low S/N ratio of 10. In this case, both approaches seem able to correctly identify the lateral elastic variabilities, producing final litho-fluid facies in good agreement with the true distribution and the posterior model correctly reproduces the transition matrix with the related facies proportion classification. Note that at approximately 936 [ms], the analytical approach describes an unphysical transition (Figure 25), while the introduction of the Markov model simulator in the numerical approach avoids such a transition (Figure 26).

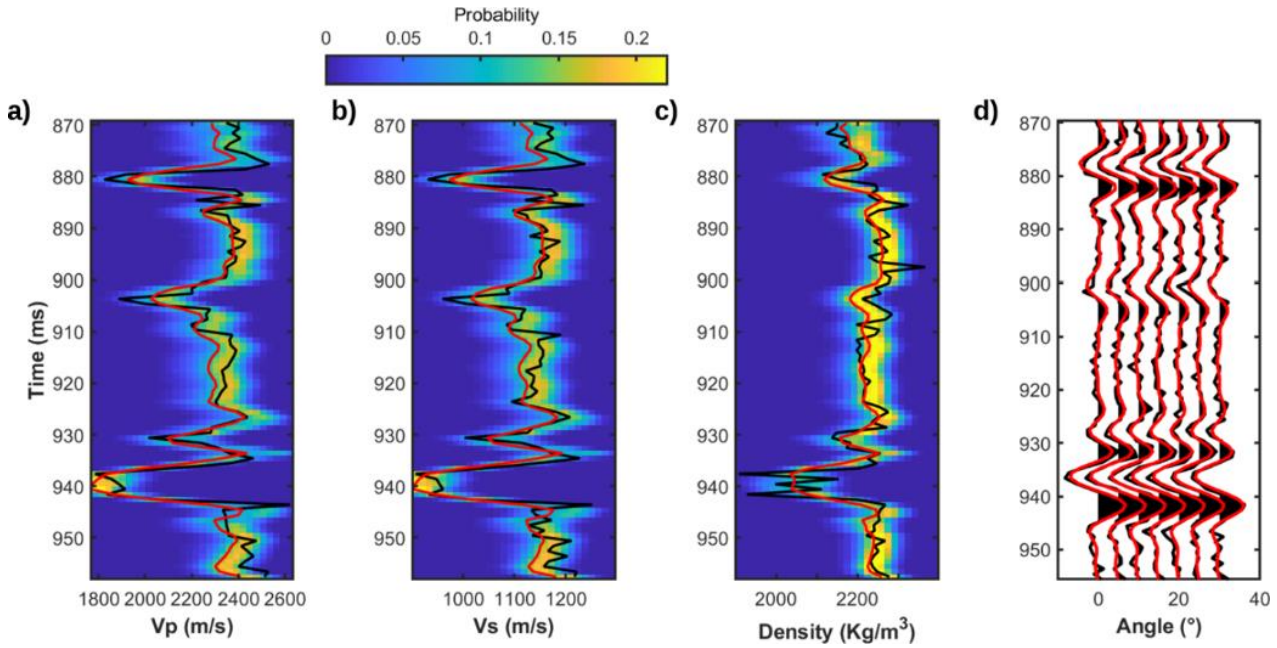


*Figure 23: Results provided by the analytical inversion for the first blind well test. In a), b) and c), the black lines represent the true property values, the red lines are the estimated mean models, and the colourmap codes the estimated posterior PDF. d) Comparison of observed (black) and predicted (red) seismic data computed on the posterior mean model.*

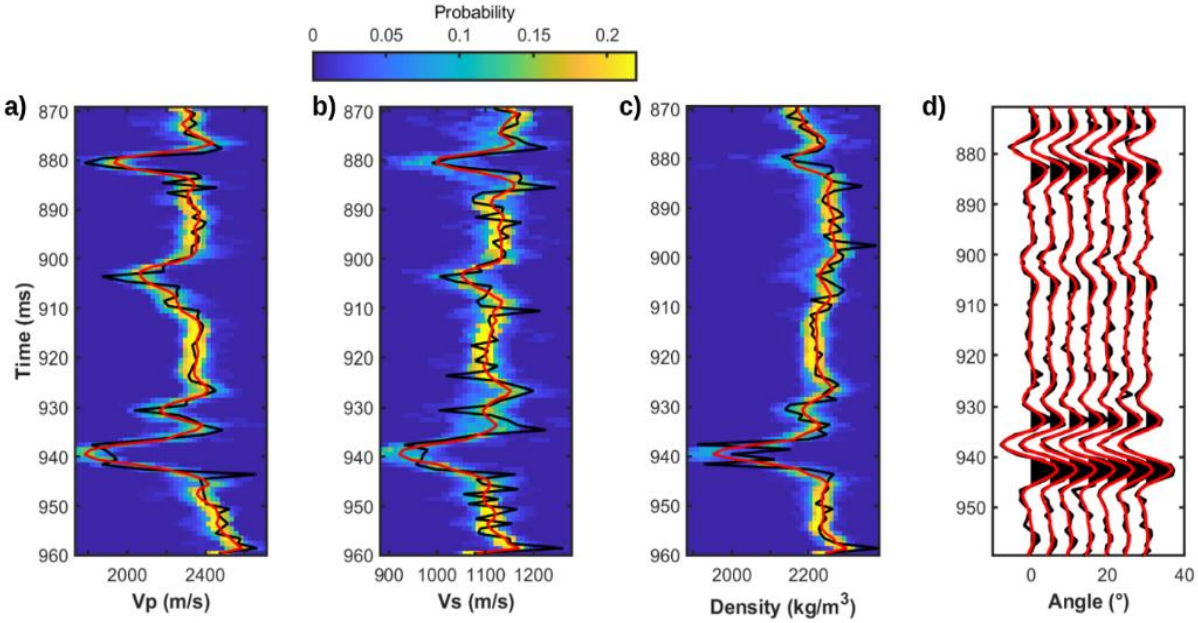


*Figure 24: As in Figure 23 but for the McMC inversion.*

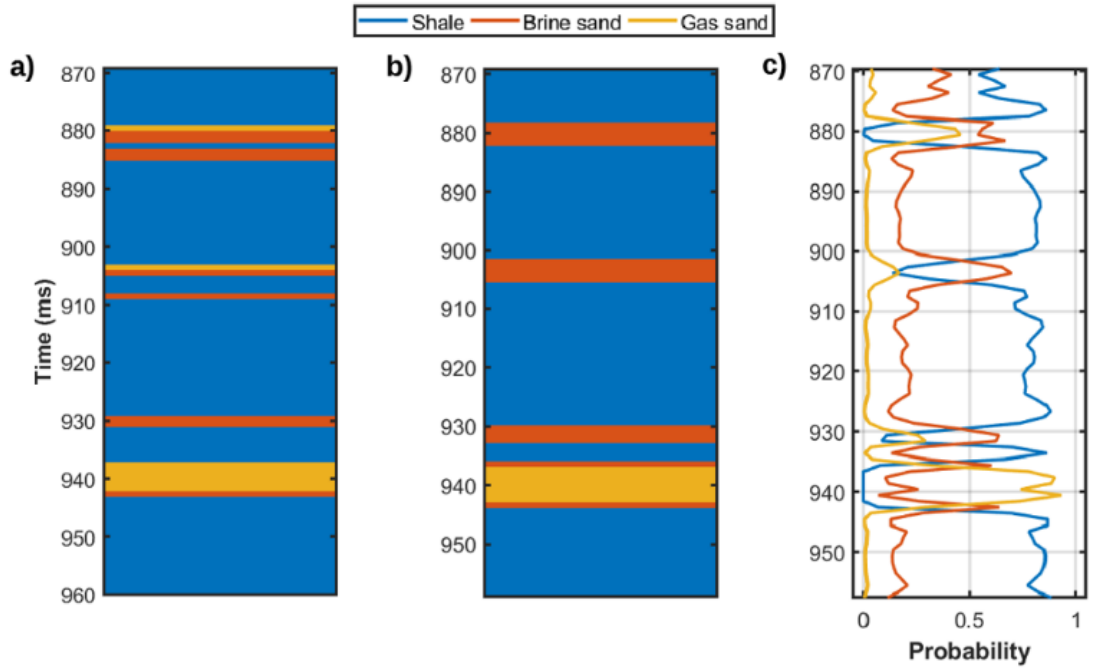


*Figure 25: Facies classification results provided by the analytical inversion for the first blind test. a) True facies profile. b) MAP facies solution. c) Estimated posterior PDF of facies.*

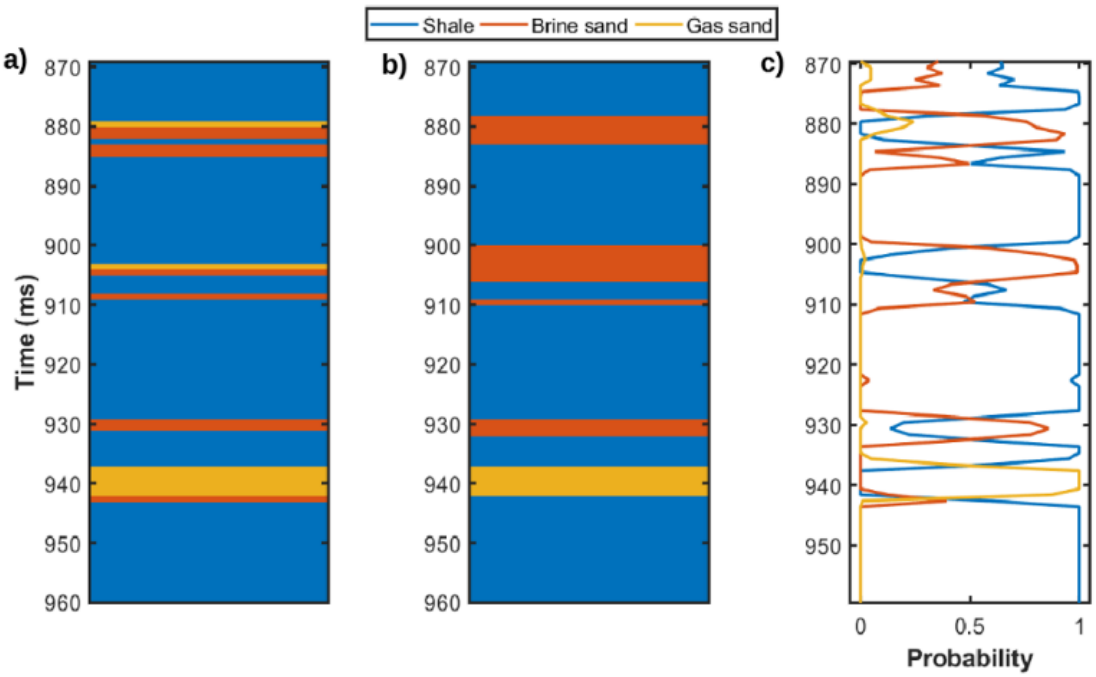


*Figure 26: As in Figure 25 but for the McMC algorithm.*

To quantitatively analyse both results, the coverage probability is computed. This statistical tool allows us to calculate the confidence interval (here fixed at 0.80 probability around the posterior mean model) that contains the true property model. The results are reported in Table (1). From this table, we observe the higher coverage probabilities provided by the McMC approach, meaning that even in data with low contamination by noise, this approach provides more significant and stable results.

| **Coverage probabilities (0.8)** | **P-wave velocity** | **S-wave velocity** | **Density** |
|---|---|---|---|
| **Analytical** | 0.916 | 0.912 | 0.926 |
| **McMC** | 0.944 | 0.902 | 0.934 |

*Table 1: Coverage probabilities for the elastic variables extracted by the analytical approach and the numerical approach.*

The second synthetic CMP gather derived from the remaining well log is contaminated with higher Gaussian-distributed noise, decreasing the S/N ratio to 2. Again, both approaches yield similar elastic distributions (Figures 27 and 28), while the improved performances provided by the

numerical approach can be clearly appreciated by a qualitative and quantitative analysis of the predicted litho-fluid facies (Figures 29 and 30).

Note in particular the differences between the proposed approaches below 934 [ms], where the analytical approach does not correctly predict the background shale sequence with brine sand and introduces a non-existent thin gas sand layer. The coverage probabilities calculated over this test and reported in Table (2) clearly show the current analysis.

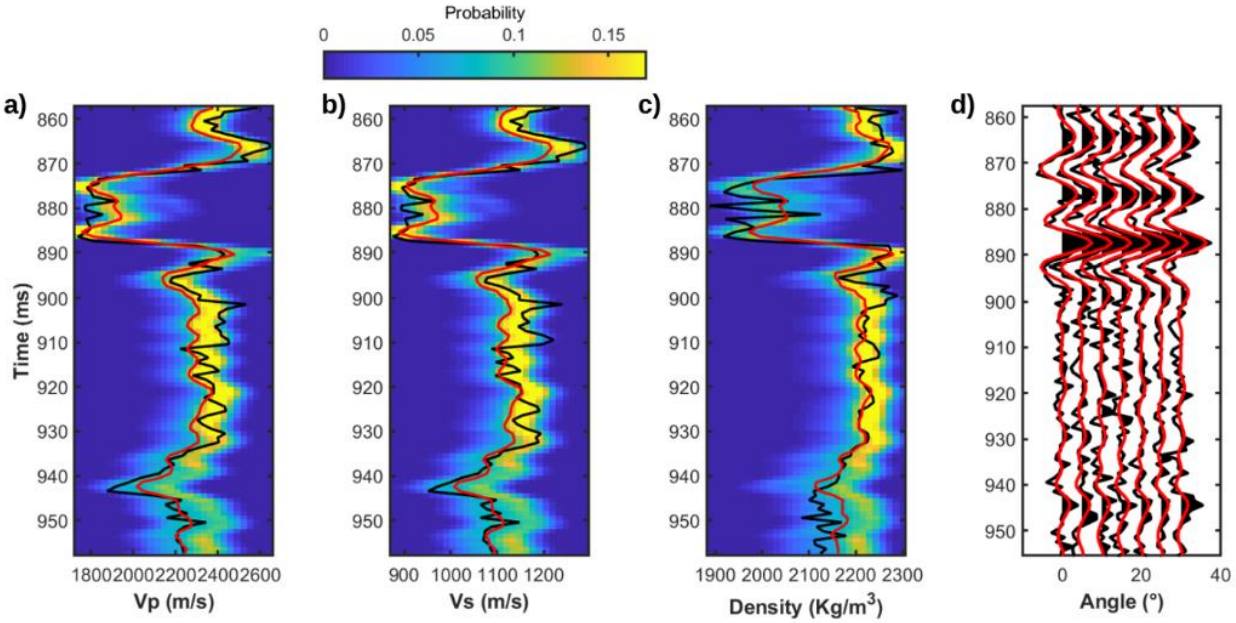


*Figure 27: Results provided by the analytical inversion for the second blind well test. In a), b) and c), the black lines represent the true property values, the red line are the mean models, and the colourmap codes the estimated posterior PDF. d) Comparison of observed (black) and predicted (red) seismic data computed on the posterior mean model.*

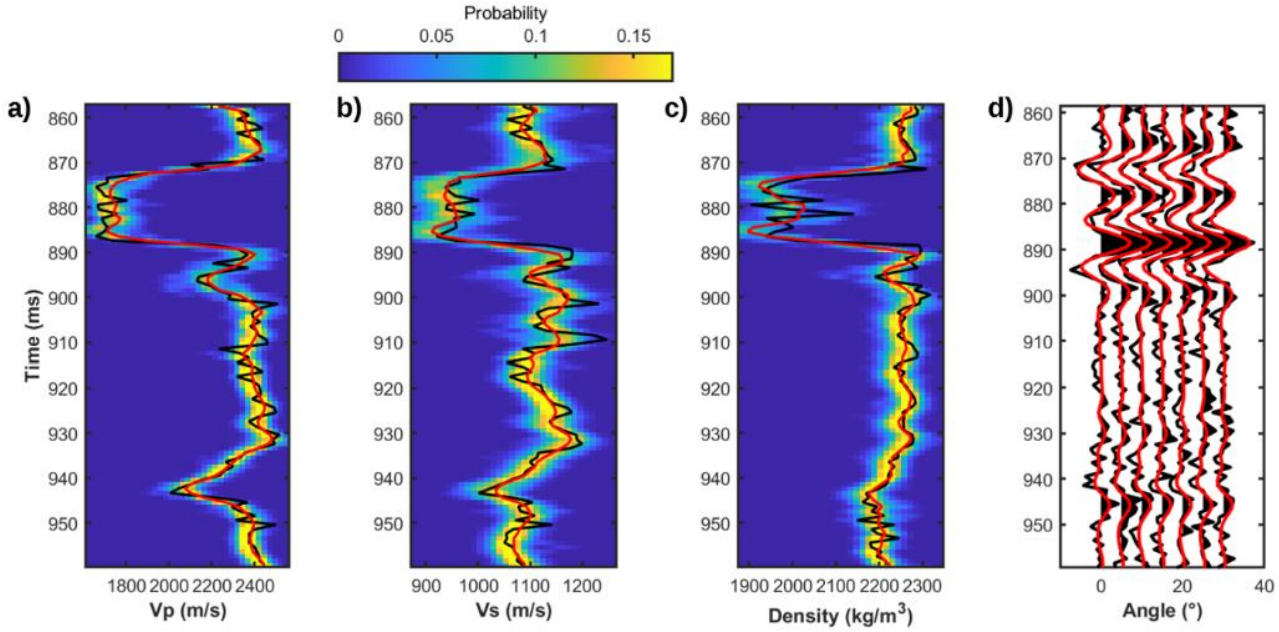


*Figure 28: As in Figure 27 but for the McMC inversion.*

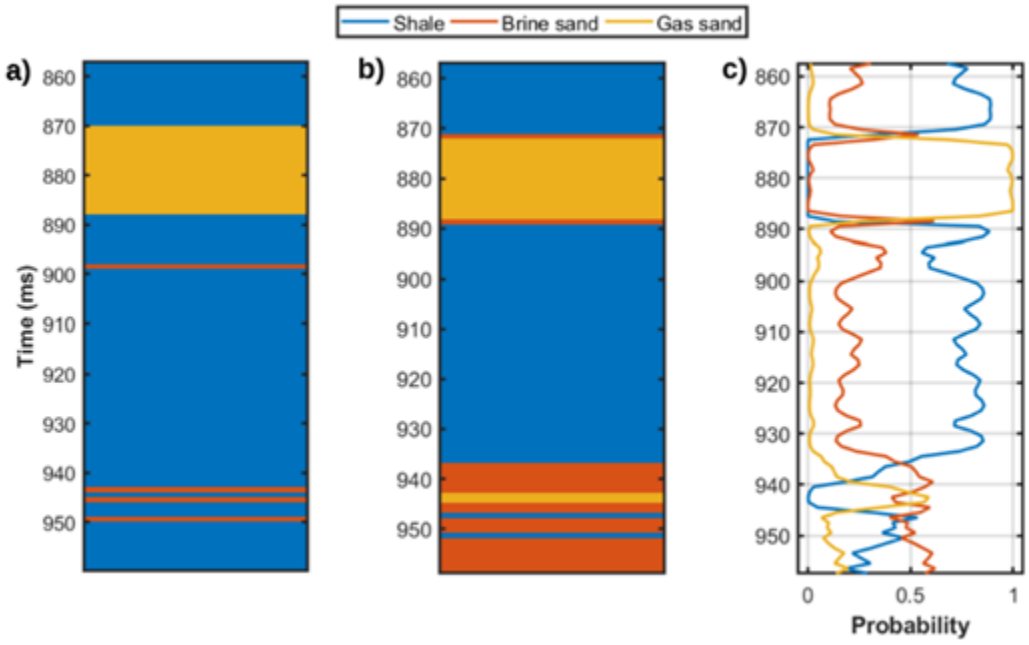


*Figure 29: Facies classification results provided by the analytical inversion for the second blind test. a) True facies profile. b) MAP facies solution. c) Estimated posterior PDF of facies.*

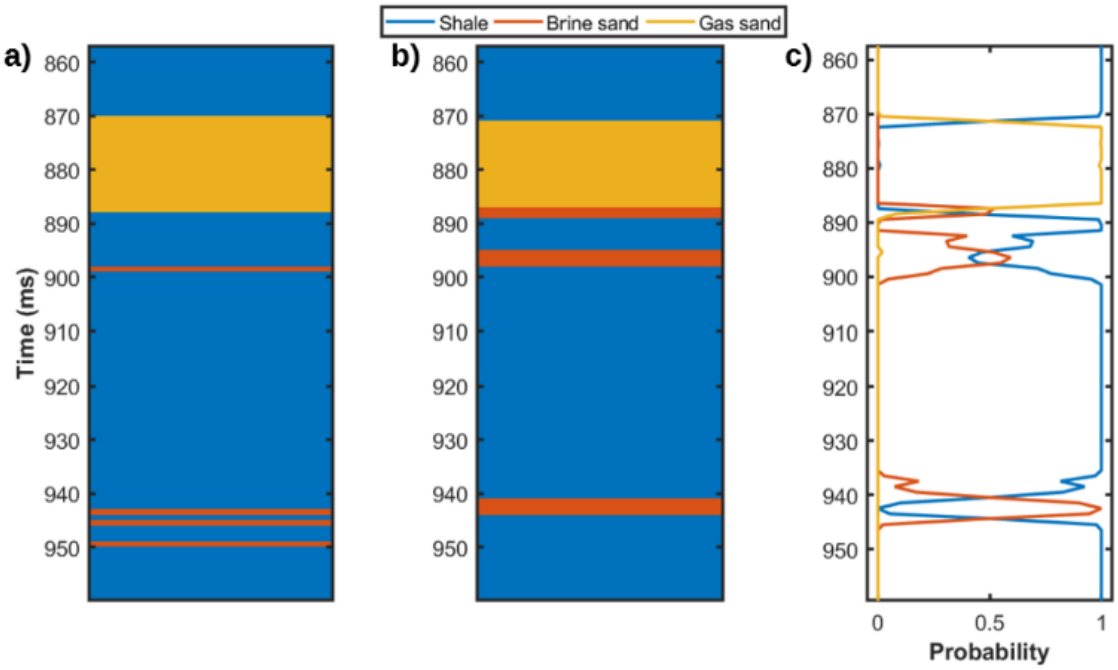


*Figure 30: As Figure 29 but for the McMC algorithm.*

| **Coverage probabilities (0.8)** | **P-wave velocity** | **S-wave velocity** | **Density** |
|---|---|---|---|
| **Analytical** | 0.824 | 0.796 | 0.805 |
| **McMC** | 0.924 | 0.865 | 0.906 |

*Table 2: Coverage probabilities for the elastic variables extracted by the analytical approach and the numerical approach*

We now focus our attention on the second blind well example to analyse the convergence of the interval-oriented algorithm. Figure 31 shows the analysis of the negative log-likelihood associated with the numerical approach of 5 running chains, showing how approximately 5,000 iterations are needed to achieve the stationary regime. Note again the beneficial effect provided by the PT and DRS approach to the convergence rate of the algorithm.

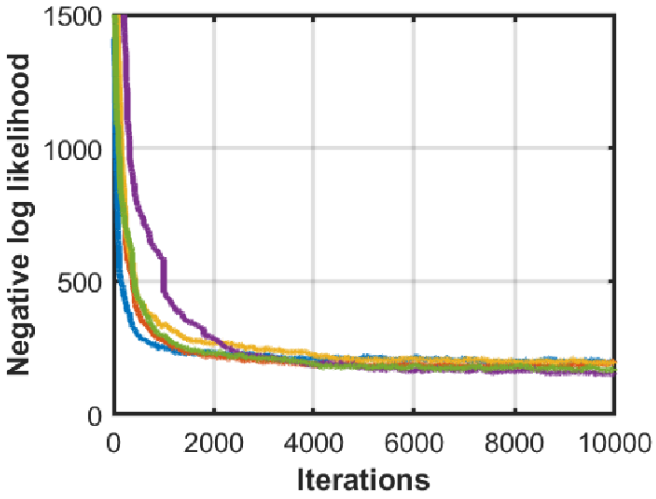


*Figure 31: Examples of the evolution of the negative log-likelihood for 5 of 40 interactive chains.*

Figure 32 shows examples of the evolution of the PSRF value for the Vp, Vs and density values extracted from six different time positions.

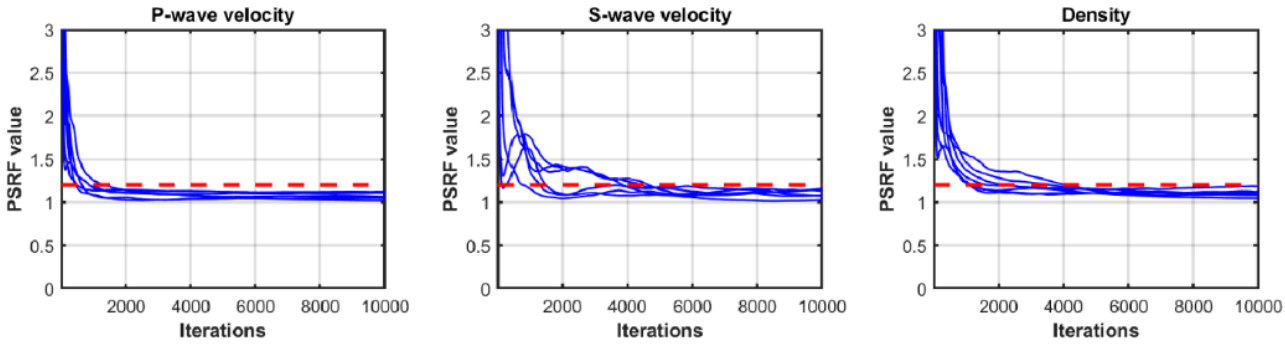


*Figure 32: Examples of the evolution of the PSRF value for the elastic properties estimated at 6 different time positions. In each plot, the dotted red lines show the desired PSRF value of 1.2, whereas the blue lines indicate the elastic properties estimated at different time samples.*

It appears that approximately 1,000 iterations are needed to achieve reliable uncertainty quantifications for the Vp parameter, whereas the convergence for the Vs and density is attained after 5,000 iterations. These different convergence rates are again related to the different influences of Vp, Vs and density on the observed seismic amplitudes. The analysis of the normalized autocorrelation functions is used to set the lag-period value in the McMC sampling (Figure 33).
We observe that approximately 20 consecutive models are needed to obtain a correlation value below the selected threshold of 0.5–0.4.

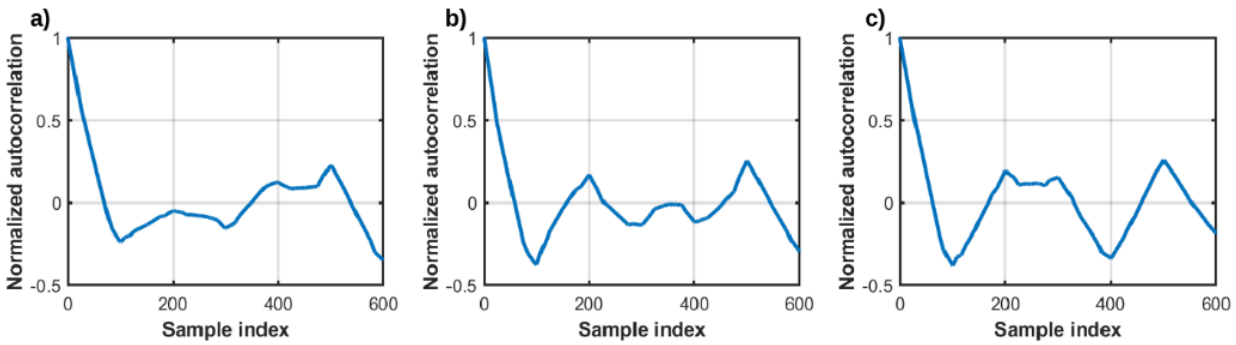


*Figure 33: Close-up of the normalized autocorrelation functions for the* $\boldsymbol{Vp}$ *(a),* $\boldsymbol{Vs}$ *(b), and density (c) models sampled at a given time position.*

By comparing the marginal distributions associated with each elastic property derived by all models sampled by the McMC method, illustrated in Figure (34), we can qualitatively observe the good match with the actual marginal distributions derived by the second well log.

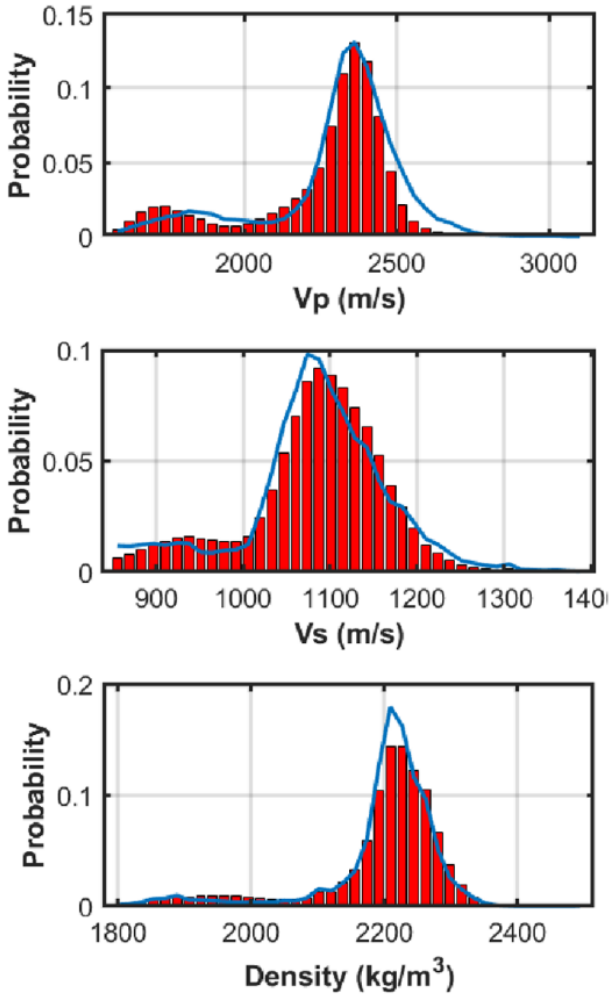


*Figure 34: Comparison between the actual marginal distributions of elastic parameters along the second blind well (blue curves) and the marginal distributions derived from the accepted McMC models (red bars).*

Regarding the computational cost, the serial analytical code runs in a few seconds, while the parallel numerical McMC approach requires approximately 13 minutes with the same computer node characteristics as previously described.

## 2.5 Conclusions

Here, we propose two McMC approaches for target-oriented and interval-oriented AVO inversion. The application of the described algorithms is suitable in the case of complex distributions, non-linear forward models and mixed discrete-continuous properties and requires no assumption about the distribution of the elastic properties investigated. To overcome the strong ill-conditioning of the AVO inversion, we propose the inclusion of geostatistical constraints in the form of lateral and vertical variogram models that guarantee better performances, especially in the case of data strongly contaminated by noise. To hasten the algorithm convergence, we make use of parallel tempering and the delayed rejection scheme and propose to hybridize the adopted approach with a DEMC method. The results obtained are encouraging and produce optimal elastic and litho-fluid facies reconstruction when compared to faster analytical methods that assume Gaussian-mixture distributed elastic variables.

The main downside of this approach is represented by the computational time required. To partially resolve such problems and improve the algorithm performances, more chains could be run, and all hyper-parameters should be optimally set. Furthermore, to reduce the number of iterations needed to reach the stationary regime, a previous fast Bayesian linearized inversion could be run, and the results obtained could be used as starting models for McMC inversion.
Our opinion, however, is that even if all these improvements were made, the extension to 2D and 3D inversions would still require infeasible computational efforts. Indeed, in these high-dimensional spaces, the chain convergence and the algorithm acceptance ratio would be extremely low since the posterior probability density function tends to be highly localized within each model dimension.

Possible further improvements and refinements to reduce the ill-conditioning of the considered AVO inversion could be introduced considering the spatial correlation of noise inside the inversion kernel, considering the noise affecting the data as additional unknowns during the inversion (Bodin *et al.* 2012), or generally adopting a numerical transdimensional approach.

# CHAPTER 3

# A transdimensional approach to reduce AVO inversion ill-conditioning.

In Chapter 2, we propose a numerical approach to address the ill-conditioning of the AVO inversion with the inclusion of variogram models in the inversion framework. If the application of these geostatistical constraints allows stabilization of the inversion, it also strongly increases the computational time needed to derive stable posterior PDFs and easily drives the inversion to unstable results in case of erroneous constraints.

We therefore propose another McMC method that addresses the ill-conditioning of the AVO inversion problem through a transdimensional approach. Unlike Chapter 2, this chapter considers an uninformative (i.e., uniform) prior distribution that brings no information into the inversion; the lack of any constraint increases the ill-conditioning of the inversion that is addressed here considering an approach that assumes that the data themselves can provide information about the model parameterization. Analogous to the method in the preceding Chapter 2, we accelerate the algorithm convergence through parallel tempering and a delayed rejection scheme meta-algorithm.

## 3.1 Introduction

During the inversion process, the main focus is on parameter estimation, where the purpose is to determine the subsurface parameters from the observed data and the model is parametrized using a previously fixed number of unknowns. It is, however, well known that the model parametrization needs to be a compromise between model uncertainty and resolution; underparametrized models

may indeed lead to underestimated seismic data or underestimated model uncertainty and vice versa, resulting in biased posterior PDFs and increased ill-conditioning.
In other words, when we are interested in deriving the elastic parameters underlying a reservoir, we invert the seismic data using models with previously fixed numbers of grid cells or layer thicknesses. However, we do not know a priori what model parametrization optimally fits the observed data and may therefore easily produce under- or overestimated parametrizations. This obviously affects our ability to optimally recover the posterior PDF.

To avoid such problems, standard statistical tools, such as *F-tests*, are commonly applied. In this way, we repeat the inversion with models characterized by different parametrizations and perform statistical analyses to determine the optimal model parametrization (Ando 2010, Sambridge *et al.* 2006). Unfortunately, almost all available statistical tools require linear inverse problems and Gaussian distributions.

To this end, it is usually convenient to adopt an approach that considers the model parametrization as unknown, where the algorithm samples models that are discretized in different iterations. This represents the starting point for all transdimensional approaches (Malinverno and Briggs 2004). The original transdimensional approaches were formulated to solve seismological problems (Bodin *et al.* 2012); recently, they have been successfully applied to hydrocarbon exploration (Sen and Biswas 2017, Zhu and Gibson 2018).

We hence present a 1D transdimensional McMC approach to AVO inversion of pre-stack seismic data to derive the elastic and associated petrophysical properties, together with the locations of reflecting interfaces and the number of layers. Since all transdimensional problems are generally characterized by low acceptance ratios and slow convergence rates, we again adopt a PT strategy and a DRS method to increase the method performance.

Since AVO seismic data do not provide information about the petrophysical properties associated with the exploration area of interest, a calibrated RPM is needed here. In our case, the rock-physics model makes use of theoretical equations based on granular media models to describe the relations between the P-wave velocity, S-wave velocity and density and the porosity ($\varphi$), water saturation (Sw) and shaliness (Sh).

## 3.2 Theoretical Rock-Physics Model

Following Avseth *et al.* (2005), we use the Hertz-Mindlin theory to define the shale and sand dry elastic properties at critical porosity and hydrostatic pressure, whereas the Hashin-Strikmann lower bound is used to simulate the compaction effect on the dry elastic moduli. The density and bulk modulus of gas and water are defined by applying the Batzle and Wang (1992) equations, taking into account temperature and pressure at the target depth. The properties of gas and water are combined according to the Wood equation (Mavko et al., 2009) to derive the preoperties of the saturating fluid. Finally, the Gassmann equation is used to combine the dry elastic properties of the rock matrix and the bulk modulus of the saturating fluid to compute the elastic moduli of the saturated rock. The adopted theoretical model is schematized in Figure (35).

Here, we assume theoretical equations formulated for granular media that independently formulate the anhydrous rock matrix and the saturating fluid properties and subsequently feedforward these into the Gassmann model to simulate the behaviour of a fully saturated rock. It is therefore based on assumptions that lead to a natural oversimplification of the area under investigation: we consider only alternating sandy-clayey sequences, spherical grains are assumed, and the system is assumed to be homogeneous and isotropic.

Based on the analysis of the geological context, we assume a rock volume that has a matrix of smectite and quartz and is saturated with a mixture of gas and brine (salty water).

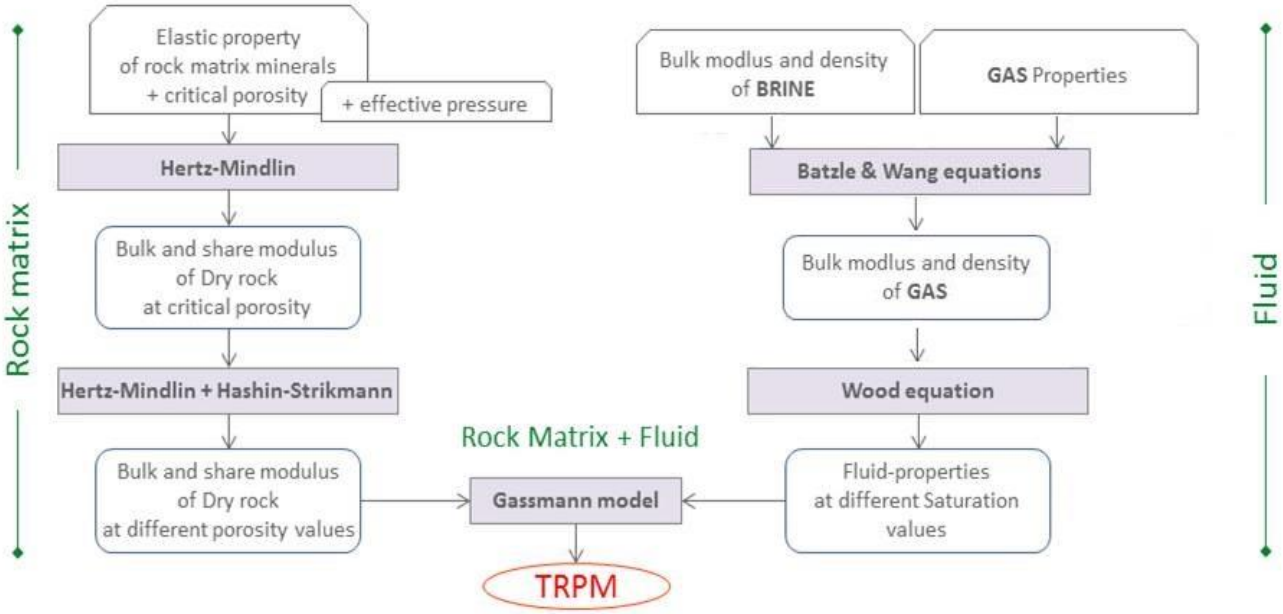

*Figure 35: The theoretical approach adopted based on the theoretical equations of Hertz-Mindlin, Hashin-Strickmann, Batzle and Wang and Gassmann.*

### 3.2.1 Dry Rock Model

From the Hertz-Mindlin granular contact theory (Mavko *et al.* 2009), we derive the elastic properties of the rock matrix. The elastic moduli can be derived as:

$$K_{HM} = \sqrt[3]{\frac{C^2\,(1-\,\varphi_c)^2\,\mu^2\,P}{18\,\pi^2(1-\,\nu)^2}} \quad (37)$$

$$\mu_{HM} = \frac{5-4\,\nu}{5(2-\nu)}\,\sqrt[3]{\frac{3C^2\,(1-\,\varphi_c)^2\,\mu^2\,P}{2\,\pi^2(1-\,\nu)^2}} \quad (38)$$

where $\mu$ and $\nu$ are the shear modulus and Poisson's ratio, respectively; $\varphi_c$ defines the porosity limit over which the grain-to-grain contact is lost, the shear modulus vanishes and the rock is in fluid suspension (critical porosity); and $C$ is the coordination number that describes the number of contacts between adjacent mineral grains.

Once the elastic moduli at the critical porosity are calculated, we interpolate their trend to null porosity values, which occur when only a homogenous rock volume of clayey and sandy minerals is considered. We used the soft-sediment Hashin-Shtrikman (Hashin & Shtrikman 1963, Mavko *et al.* 2009) lower bound trend:

$$K_{dry} = \left[\frac{\varphi/\varphi_c}{K_{HM}+\frac{4}{3}\mu_{HM}}\right]^{-1} - \frac{4}{3}\mu_{HM} \quad (39)$$

$$\mu_{dry} = \left[\frac{\varphi/\varphi_c}{B+\mu_{HM}} + \frac{1-\varphi/\varphi_c}{K_{VR}+\mu_{HM}}\right]^{-1} - \frac{B}{6}\mu_{HM} \quad (40)$$

where

$$B = \left(\frac{9K_{HM} + 8\mu_{HM}}{K_{HM} + 2\mu_{HM}}\right) \tag{41}$$

$$K_{VR} = \frac{1}{2}\left(\frac{Qzt\ \ K_{Qzt} + Clay\, K_{Clay}}{\frac{Qzt}{K_{Qzt}} + \frac{Clay}{K_{Clay}}}\right) \tag{42}$$

where $Qzt$, $Clay$, $K_{Qzt}$ and $K_{Clay}$ identify the quartz and clay percentages and the elastic moduli of quartz and clay, respectively.

The reason we decided to use the Hashin-Shtrikman model is that it provides experimentally deduced equations to derive the elastic moduli while no assumptions are made regarding the grain geometry, which is often rather difficult to approximate due to grain irregularity and heterogeneity.

In this model, the rock volume is represented by spheres, each composed of an external shell and an internal core. The different elastic responses of the system are generated by placing the material with stiffer mechanical behaviour outside and the softer inside (upper *stiff-sediment* configuration) and vice versa (lower *soft-sediment* configuration), as graphically presented in Figure (36).

Given the shallow depth of the considered target (described below), we adopt the lower soft bound trend.

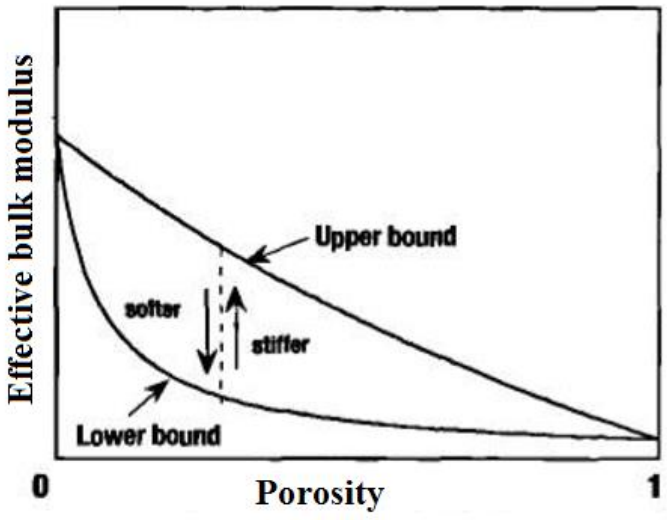

*Figure 36: Graphic representation of bulk elastic bounds* versus *porosity. The Hashin-Shtrikman bounds allow the generation of different mechanical behaviours along the upper or lower limit considered (modified from Avseth et al., 2005).*

### 3.2.2 Fluid Properties Model

The fluid most commonly encountered is brine, an aqueous solution whose composition changes based on the saline solution considered (generally sodium chloride). Through the Batzle-Wang experimental polynomial equations, we can derive its elastic moduli as (Batzle-Wang, 1992):

$$\rho_{brine} = \rho_w + 0.65S^2 + 10^{-6}S[300\,P + 2400\,P\,S + T(80 + 3\,T - 3300\,S - 13\,P + 47\,P\,S)] \quad (43)$$

$$V_{brine} = V_w + S\,(1170 - 9.6\,T + 0.05\,T^2 - 510^{-6}\,T^3 + 2.6\,P - 0.003\,T\,P - 0.05P^2) + S^{1.5}\,(780 - 10\,P + 0.16\,P^2) - 1820\,S^2 \quad (44)$$

where $P$ [MPa] and $T$ [°$C$] are *in situ* pressure and temperature; $S$ stands for brine salinity (weight fraction); and $\rho_w$ is water density, given by:

$$\rho_w = 1 + 10^{-6}(-80\,\mathrm{T} - 3.3\,\mathrm{T}^2 + 0.002\,T^3 + 489\,P - 2\,T\,P + 0.002\,T^2\,P - 10^{-5}\,T^3\,P - 0.33\,P^2 - 0.002\,TP^2) \quad (45)$$

where $V_w$ describes the wave velocity in pure water and is calculated as:

$$V_w = \sum_{i=1}^{5}\sum_{j=1}^{4} W_{ij}\,T^{i-1}P^{j-1} \quad (46)$$

with $W_{ij}$ constant values as reported in Table (3).

| | | | |
|---|---|---|---|
| $\mathbf{W_{11}}$ | 1402.85 | $\mathbf{W_{13}}$ | $3.437*10^{-3}$ |
| $\mathbf{W_{21}}$ | 4.871 | $\mathbf{W_{23}}$ | $1.739*10^{-4}$ |

| | | | |
|---|---|---|---|
| $\mathbf{W_{31}}$ | -0.0478 | $\mathbf{W_{33}}$ | $-2.135*10^{-6}$ |
| $\mathbf{W_{41}}$ | $1.487*10^{-4}$ | $\mathbf{W_{43}}$ | $-1.455*10^{-8}$ |
| $\mathbf{W_{51}}$ | $-2.197*10^{-7}$ | $\mathbf{W_{53}}$ | $5.23*10^{-11}$ |
| $\mathbf{W_{12}}$ | 1.524 | $\mathbf{W_{14}}$ | $-1.197*10^{-5}$ |
| $\mathbf{W_{22}}$ | -0.011 | $\mathbf{W_{24}}$ | $-1.628*10^{-6}$ |
| $\mathbf{W_{32}}$ | $2.747*10^{-4}$ | $\mathbf{W_{34}}$ | $1.237*10^{-8}$ |
| $\mathbf{W_{42}}$ | $-6.503*10^{-7}$ | $\mathbf{W_{44}}$ | $1.237*10^{-10}$ |
| $\mathbf{W_{52}}$ | $7.987*10^{-10}$ | $\mathbf{W_{54}}$ | $-4.614*10^{-13}$ |

*Table 3: Constants for P-wave velocity in pure water.*

The brine bulk modulus can therefore be estimated as:

$$K_{brine} = \rho_{brine}\ V_{brine}^2 \quad 10^{-6} \tag{47}$$

Through the Batzle-Wang equations, we also derive the bulk modulus and the density of the saturating gas:

$$\rho_{gas} \approx \frac{28.8\ G\ P}{Z\ R\ (T + 273.15)} \tag{48}$$

where $G$ is the dimensionless specific gravity of the gas, $R$ [$\mathrm{J \cdot K^{-1} \cdot} mol^{-1}$] is the gas molar constant, $P$ [$MPa$] is the *in-situ* pressure, $T$ [$C$] is the *in-situ* temperature and $Z$ is the compressibility factor, obtained as:

$$Z = P_{pr}\left[0.03 + 0.00527\left(3.5 - T_{pr}\right)^3\right] + E + 0.642\ T_{pr} - 0.007\ T_{pr}{}^4 - 0.52 \tag{49}$$

where

$$E = 0.109\ (3.85 - T_{pr})^2\ exp^{\frac{P_{pr}^{1.2}}{T_{pr}}\left[0.45+8\ (0.56-\frac{1}{T_{pr}})^2\right]} \tag{50}$$

In equations (49) and (50), $T_{pr}$ and $P_{pr}$ represent the pseudo-reduced temperature and pressure, respectively, which are derived as follows:

$$T_{pr} = \frac{T + 273.15}{94.72 + 170.75\,G} \tag{51}$$

$$P_{pr} = \frac{P}{4.892 - 0.4048\,G} \tag{52}$$

The bulk modulus [$GPa$] is instead derived as:

$$K_{gas} \approx \frac{\gamma_0\,P}{10^3\,T\,(\,1 - \frac{P_{pr}}{Z}\,\frac{\partial Z}{\partial P_{pr}})} \tag{53}$$

where:

$$\gamma_0 = 0.85 + \frac{5.6}{2 + P_{pr}} + \frac{27.1}{\left(3.5 + P_{pr}\right)^2} - 8.7\,exp^{-0.65\,(P_{pr}+1)} \tag{54}$$

$$\frac{\partial Z}{\partial P_{pr}} = 0.03 + 0.00527\left(3.5 - T_{pr}^3\right) + 0.109\left(3.85 - T_{pr}^2\right) + F \tag{55}$$

$$F = -1.2\,\frac{{P_{pr}}^{0.2}}{T_{pr}}\left[0.45 + 8\,(0.56 - \frac{1}{T_{pr}})^2\right]\,exp^{\frac{{P_{pr}}^{1.2}}{T_{pr}}\left[0.45+8\,(0.56-\frac{1}{T_{pr}})^2\right]} \tag{56}$$

### 3.2.3 Rock-Fluid Model

Gassmann (1951) derived simple fundamental equations to predict the elastic properties of a saturated porous medium using only the bulk properties of rock and fluid and with no assumption about the specific pore geometric distribution. These theoretical equations are based on the following assumptions:

I. The pore pressure is identically homogeneous. This implies that the saturated rock shear modulus is equal to the anhydrous rock shear modulus, $\mu_{sat} = \mu_{dry}$ (Berryman J., 1999).
II. There is no chemical interaction between the saturating fluid and the rock matrix. This implies that phenomena such as chemical precipitation, which significantly reduce the total porosity of rocks, are not considered here.
III. The high-frequency effects of the relative motion between the rock matrix and fluid are not described. The Gassmann equations are basically the extension to lower frequency ranges (seismic length scale) of the more general Biot equations, where the high frequencies trigger a dissipative relative movement between rock and saturating fluids.

The bulk modulus for the saturated rock is finally deduced as:

$$K_{sat} = K_{frame} + \frac{\left(1 - \frac{K_{frame}}{K_{matrix}}\right)^2}{\frac{\varphi}{K_{flu}} + \frac{1-\varphi}{K_{matrix}} - \frac{K_{frame}}{{K_{matrix}}^2}} \quad (57)$$

where $K_{sat}$, $K_{frame}$, $K_{matrix,}$ and $K_{flu}$ are the bulk modulus of the saturated rock, porous rock frame (drained of any pore-filling fluid), mineral matrix, and pore fluid, respectively, and $\varphi$ is the porosity (as fraction). The saturated rock density is derived using a balancing equation, where each component is weighted with the relative porosity values:

$$\rho_{sat} = \varphi\, \rho_{fluid} + (1 - \varphi)\ \rho_{matrix} \quad (58)$$

The elastic P-wave and S-wave velocities are finally estimated through:

$$V_p = \sqrt{\frac{K_{sat} + \frac{4}{3}\,\mu_{sat}}{\rho_{sat}}} \quad (59)$$

$$V_s = \sqrt{\frac{K_{sat}}{\rho_{sat}}} \quad (60)$$

## 3.3 The Adopted Method

We recall the Bayes theorem (equation 3) that, given a parametrization $\mathcal{H}$ of the model space, becomes:

$$p(\mathbf{m}|\mathbf{d},\mathcal{H}) = \frac{p(\mathbf{d}|\mathbf{m},\mathcal{H})p(\mathbf{m}|\mathcal{H})}{p(\mathbf{d}|\mathcal{H})} \tag{61}$$

If $\mathcal{H}$ describes a model parametrization with *n* unknowns, it can be written as:

$$p(n,\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m},n)p(\mathbf{m}|n)p(n)}{p(\mathbf{d})} \tag{62}$$

where again the left term describes the posterior distribution, which is the target of the inversion process and is computed through all models collected by the McMC random walk in the model space. For the transdimensional approaches, the Metropolis-Hastings rule that determines with a probability α whether the perturbed models are accepted or rejected becomes:

$$\alpha = p(\mathbf{m}',n'|\mathbf{m},n) = \min\ [1, \text{prior ratio } \times \text{likelihood ratio } \times \text{proposal ratio } \times |\mathbf{J}|] = \tag{63}$$
$$\min\left[1, \frac{p(\mathbf{m}',n')}{p(\mathbf{m},n)} \times \frac{p(\mathbf{d}|\mathbf{m}',n')}{p(\mathbf{d}|\mathbf{m},n)} \times \frac{q(\mathbf{m},n|\mathbf{m}',n')}{q(\mathbf{m}',n'|\mathbf{m},n)} \times |\mathbf{J}|\right]$$

where the current model **m** has dimension *n* and the perturbed model $\mathbf{m}'$ has dimension $n'$. The Jacobian matrix **J** must be involved when the current and the perturbed models have different dimensions; the prior term again describes information derived from the available well logs, and the proposal $q()$ term describes the $\mathbf{m}'$ model as a random perturbation of **m**.

### 3.3.1 Prior And Likelihood Models

When performing the elastic inversion, the model **m** includes the elastic variables and **z**, the location of the interfaces; $\mathbf{m} = [\mathbf{e},\mathbf{z}] = [\boldsymbol{Vp},\boldsymbol{Vs},\boldsymbol{\rho},\mathbf{z}]$**.** When we perform the petrophysical inversion, the model **m** includes the petrophysical properties **r** and the locations of the interfaces;

$\mathbf{m} = [\mathbf{r}, \mathbf{z}] = [\boldsymbol{\varphi}, \mathbf{Sw}, \mathbf{Sh}, \mathbf{z}]$**.** For the currently adopted approach, the prior distribution is composed of three different terms: the prior over the model dimensions $p(n)$, over the elastic or petrophysical layer properties $p(\mathbf{e}, \mathbf{r}|n)$ and over the layer partitioning $p(\mathbf{z}|n)$, here assumed to be independently distributed. We can therefore express the total prior model as:

$$p(\mathbf{m}, n) = p(\mathbf{e}|n)p(\mathbf{z}|n)p(n) \quad (64)$$

$$p(\mathbf{m}, n) = p(\mathbf{r}|n)p(\mathbf{z}|n)p(n) \quad (65)$$

Here, we make use of the theoretical assumption of uncorrelated petrophysical and elastic variables, through which equations (64) and (65) can be reformulated as:

$$p(\mathbf{m}, n) = p(\mathbf{Vp}|n)p(\mathbf{Vs}|n)p(\boldsymbol{\rho}|n)p(\mathbf{z}|n)p(n) \quad (66)$$

$$p(\mathbf{m}, n) = p(\boldsymbol{\varphi}|n)p(\mathbf{Sw}|n)p(\mathbf{Sh}|n)p(\mathbf{z}|n)p(n) \quad (67)$$

where $p(n)$ is defined as a uniform bounded distribution over the interval $(n_{min}, n_{\max})$:

$$p(n) = \begin{cases} \frac{1}{\Delta n}, & \text{if } n_{min} \leq n \leq n_{max} \\ 0, & \text{otherwise} \end{cases} \quad (68)$$

where $\Delta n = n_{max} - n_{min} + 1$.

The prior over the layer partitioning is:

$$p(\mathbf{z}|n) = \frac{n!\,(N-n)!}{N!} \quad (69)$$

The considered N+1 time samples can be split into N cells. The prior over the petrophysical and elastic properties can be defined as a uniform bounded distribution over the interval $(r_{min}, r_{\max})$ and $(e_{min}, e_{\max})$, respectively:

$$p(e|n) = \begin{cases} \prod_{i=1}^{n} p(e_i|n) = \prod_{i=1}^{n} \frac{1}{\Delta e} = \frac{1}{(\Delta e)^n}, & \text{if} \quad e_{min} \leq e \leq e_{max} \\ 0, & \text{otherwise} \end{cases} \tag{70}$$

$$p(r|n) = \begin{cases} \prod_{i=1}^{n} p(r_i|n) = \prod_{i=1}^{n} \frac{1}{\Delta r} = \frac{1}{(\Delta r)^n}, & \text{if} \quad r_{min} \leq r \leq r_{max} \\ 0, & \text{otherwise} \end{cases} \tag{71}$$

where $\Delta e = e_{max} - e_{min}$ and $\Delta r = r_{max} - r_{min}$. The prior model for elastic and petrophysical variables can be compactly rewritten as:

$$p(\mathbf{m}, n) = \frac{p(\mathbf{z}|n)}{\Delta n \prod_{i=1}^{3} (\Delta e_i^n)} \tag{72}$$

$$p(\mathbf{m}, n) = \frac{p(\mathbf{z}|n)}{\Delta n \prod_{i=1}^{3} (\Delta r_i^n)} \tag{73}$$

Since inter-parameter correlations can be strong, the assumption we previously made over the prior distributions may result in an overall poorer algorithm performance and slower convergence rate. To address this issue, we follow the procedure presented by Dosso *et al.* (2014), to which the interested reader is referred for mathematical details. We apply singular value decomposition (SVD) to the posterior covariance matrix and move to a rotated model space where the elastic/petrophysical properties are uncorrelated. In this context, rotated parameters are perturbed individually according to a Gaussian distribution with zero mean and a variance given by the singular values of the currently estimated posterior covariance matrix. Subsequently, we rotate back to the original space before evaluating the likelihood function for the elastic inversion or the RPM application for the petrophysical inversion.

Following Xiang *et al.* (2018) for the starting model in the original (non-orthogonal) space, we adopt a posterior model covariance matrix calibrated around the starting model. As the sampling continues, this first approximation is refined, making use of all new sampled models.

Finally, under the assumption of Gaussian-distributed white noise, the likelihood function is again, as previously described, derived as:

$$p(\mathbf{d}|\mathbf{m},n)=\frac{1}{\sqrt{(2\pi)^{L}|\mathbf{C_d}|}}\exp\left(-\frac{\varphi(\mathbf{m},n)}{2}\right)\propto\exp\left(-\frac{\varphi(\mathbf{m},n)}{2}\right) \quad (74)$$

### 3.3.2 The Rj-McMC Inversion Scheme

The adopted inversion scheme is described as follows:

Odd iterations:

1. Property perturbation: a random perturbation is applied to all parameters of a randomly chosen layer. The perturbation is applied to the principal component space through a Gaussian-distributed proposal with a mean equal to the current parameter value.

Even iterations (all proposed perturbations are chosen with uniform probability):

2. Birth move: a new layer is created in a randomly chosen temporal position. To define its elastic/petrophysical values, a principal component perturbation is applied to its upper or lower layer values (chosen with equal probability).
3. Death move: an interface is randomly chosen and deleted. To define the property value pertaining to the delated layer, a random choice is made between the upper and lower layer.
4. Interface perturbation: the vertical position of a randomly chosen layer is perturbed. The perturbation is applied in accordance with a Gaussian-distributed proposal. Note that to avoid the creation of layers that are too thin, this perturbation can be rejected.

Since the newly created models can fall outside the prior distribution, we impose the restriction that the prior of such a proposed model is null, therefore rejecting the new model. In addition, to avoid the creation of unphysical models, we insert constraints over the Vp/Vs and ρ/Vp ratios for both the elastic and petrophysical inversions. If the newly created model does not match these criteria, it is discarded until a new acceptable model is generated. Note that for petrophysical inversion, these criteria are checked by applying the RPM to the sampled petrophysical properties.

Since shaliness and water saturation span the [0, 1] interval, while the porosity is defined over [0, 0.5], it is convenient to apply the logit transformation to define new variables defined over $\mathbb{R}$

(Bosch *et al.* 2007). We therefore define the standard $\mathbf{L^s}$ and modified $\mathbf{L^m}$ logit forward transformation for the petrophysical variables as:

$$\mathbf{L^s}(x) = \ln\left(\frac{x}{1-x}\right) \qquad (75)$$

$$\mathbf{L^m}(x) = \ln\left(\frac{x}{1-2x}\right) \qquad (76)$$

The backward inverse transformation, applied before the RPM application, is instead defined as:

$$(\mathbf{L^s})^{-1}(x) = \left(\frac{e^x}{1+e^x}\right) \qquad (77)$$

$$(\mathbf{L^m})^{-1}(x) = \left(\frac{e^x}{1+2e^x}\right) \qquad (78)$$

For the petrophysical inversion we also propagate the uncertainty related to the adopted RPM for the final PDF estimation. To this end we follow Aleardi *et al.* (2018a), perturbing the elastic properties with a Gaussian-distributed error model derived and comparing the true logged elastic properties with those derived by the RPM application. This perturbation is applied before the forward model evaluation.

To assess the convergence towards a stable PDF, standard measures of convergence, such as the potential scale reduction factor (Gelman *et al*. 2013), cannot be applied. We therefore carefully analyse the data misfit of each running chain until it reaches the stationary regime. Another method involves the estimation of the posterior PDFs after the burn-in stage by analysing the posterior PDFs in the first and second stages of sampling. If these are comparable, no further sampling is needed.

Other hyper-parameters, such as the number of chains and temperature levels, are set after a trial-and-error approach, trying to reach a compromise between the computational cost-width and stable convergence to the PDF.

### 3.3.3 Acceptance Probabilities

Here, we briefly describe the acceptance probabilities for each move for the adopted McMC approach. Since the same method is used for both elastic and petrophysical inversions, we limit this analysis to elastic inversion (**e** instead of **r**). The method we follow is derived from Bodin and Sambridge (2009) and Agostinetti and Malinverno (2010), to which the interested reader is referred.

#### 3.3.3.1 Fixed-dimensional moves

When the current and perturbed methods share the same dimensions, the algorithm can apply elastic or interface perturbations. In the former case, the principal component perturbation is applied as previously described. In the latter case, the newly created interface $\mathbf{z}'$ is:

$$\mathbf{z}' = \mathbf{z} + g\sigma_z \mathbf{u_i} \qquad (79)$$

where a Gaussian-distributed proposal is applied with a mean equal to the current interface location and standard deviation $\boldsymbol{\sigma_z}$. The parameter $g$ is a random number extracted from a Gaussian distribution $N(0,1)$ and $\mathbf{u_i}$ is the unit vector lag that identifies the perturbed interfaces.
For this elastic/interface perturbation, the prior ratio is equal to 1, as the proposal perturbation; the Metropolis-Hastings acceptance rule therefore becomes:

$$p(\mathbf{m}', n'|\mathbf{m}, n) = \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}', n')}{p(\mathbf{d}|\mathbf{m}, n)}\right] = \min\left[1, \exp\left(-\frac{\varphi(\mathbf{m}', n') - \varphi(\mathbf{m}, n)}{2T}\right)\right] \qquad (80)$$

#### 3.3.3.2 Transdimensional Moves

##### 3.3.3.2.1 Prior Ratio

The prior ratio can be expressed as:

$$\frac{p(\mathbf{m}',n')}{p(\mathbf{m},n)}=\frac{p(\mathbf{e}'|n')}{p(\mathbf{e}|n)}\frac{p(\mathbf{z}'|n')}{p(\mathbf{z}|n)}\frac{p(n')}{p(n)} \tag{81}$$

Extending equation (72) we obtain the following for a birth move:

$$\frac{p(\mathbf{m}',n')}{p(\mathbf{m},n)}=\frac{(n+1)!\,(N-(n+1))!}{N!\,\Delta n\,\prod_{i=1}^{3}(\Delta\mathbf{e}_i^{n+1})}\times\frac{N!\,\Delta n\,\prod_{i=1}^{3}(\Delta\mathbf{e}_i^{n})}{n!\,(N-n)!}=\frac{n+1}{(N-n)\prod_{i=1}^{3}(\Delta\mathbf{e}_i)} \tag{82}$$

where N defines the time samples of the considered vertical window. Analogously for the death move we obtain:

$$\frac{p(\mathbf{m}',n')}{p(\mathbf{m},n)}=\frac{(n-1)!\,(N-(n-1))!}{N!\,\Delta n\,\prod_{i=1}^{3}(\Delta\mathbf{e}_i^{n-1})}\times\frac{N!\,\Delta n\,\prod_{i=1}^{3}(\Delta\mathbf{e}_i^{n})}{n!\,(N-n)!}=\frac{\prod_{i=1}^{3}(\Delta\mathbf{e}_i)\;(N-n+1)}{n} \tag{83}$$

#### 3.3.3.2.2 Proposal Ratio

The proposal ratio can be expressed as follows:

$$\frac{q(\mathbf{m},n|\mathbf{m}',n')}{q(\mathbf{m}',n'|\mathbf{m},n)}=\frac{q(\mathbf{z}|\mathbf{z}',n')}{q(\mathbf{z}'|\mathbf{z},n)}\frac{q(\mathbf{e}|\mathbf{e}',n')}{q(\mathbf{e}'|\mathbf{e},n)} \tag{84}$$

The probability of generating new properties on the (n+1)th layer is (birth move):

$$q(\mathbf{e}'|\mathbf{e},n)=\frac{1}{\prod_{i=1}^{3}\sqrt{2\pi}\sigma_{\mathbf{e}_i}}\exp\left(-\sum_{i=1}^{3}\frac{(\mathbf{e}_i'-\mathbf{e}_i)^2}{2\sigma_{\mathbf{e}_i}^2}\right) \tag{85}$$

where $\sigma_{\mathbf{e}_i}$ is the proposal distribution standard deviation, and $\mathbf{e}_i$ is the *i*th elastic property for the current model. For the reverse step (death move), the probability of removing the elastic properties when a layer is deleted is:

$$q(\mathbf{e}|\mathbf{e}', n') = 1 \qquad (86)$$

The probability of creating the new layer becomes:

$$q(\mathbf{z}'|\mathbf{z}, n) = \frac{1}{N - n} \qquad (87)$$

The probability of the reverse step in a birth move (deleting the layer positions) is:

$$q(\mathbf{z}|\mathbf{z}', n') = \frac{1}{1 + n} \qquad (88)$$

Combining these equations, we therefore obtain the following expression for a birth perturbation:

$$\frac{q(\mathbf{m}, n|\mathbf{m}', n')}{q(\mathbf{m}', n'|\mathbf{m}, n)} = \frac{(N - n)\prod_{i=1}^{3}\sqrt{2\pi}\sigma_{\mathbf{e}_i}}{n + 1}\exp\left(\sum_{i=1}^{3}\frac{(\mathbf{e}_i' - \mathbf{e}_i)^2}{2\sigma_{\mathbf{e}_i}^2}\right) \qquad (89)$$

Analogously, for a death perturbation, the proposal becomes:

$$\frac{q(\mathbf{m}, n|\mathbf{m}', n')}{q(\mathbf{m}', n'|\mathbf{m}, n)} = \frac{n}{(N - n + 1)\prod_{i=1}^{3}\sqrt{2\pi}\sigma_{\mathbf{e}_i}}\exp\left(-\sum_{i=1}^{3}\frac{(\mathbf{e}_i' - \mathbf{e}_i)^2}{2\sigma_{\mathbf{e}_i}^2}\right) \qquad (90)$$

### 3.3.3.2.3 Acceptance Probabilities

Combining the likelihood ratio with the prior ratio and proposal ratio, we obtain for a birth perturbation:

$$[p(\mathbf{m}', n'|\mathbf{m}, n)]_{birth} = \min\left[1, \frac{\prod_{i=1}^{3}\sqrt{2\pi}\sigma_{\mathbf{e}_i}}{\prod_{i=1}^{3}\Delta\mathbf{e}_i}\exp\left(\left(\sum_{i=1}^{3}\frac{(\mathbf{e}_i' - \mathbf{e}_i)^2}{2\sigma_{\mathbf{e}_i}^2}\right) - \frac{\varphi(\mathbf{m}', n) - \varphi(\mathbf{m}, n)}{2\sigma_d^2 T}\right)\right] \quad (91)$$

Analogously, for a death move, we obtain:

$$[p(\mathbf{m}', n'|\mathbf{m}, n)]_{death} = \min\left[1, \frac{\prod_{i=1}^{3}\Delta\mathbf{e}_i}{\prod_{i=1}^{3}\sqrt{2\pi}\sigma_{e_i}}\exp\left(\left(\sum_{i=1}^{3} - \frac{(\mathbf{e}_i' - \mathbf{e}_i)^2}{2\sigma_{\mathbf{e}_i}^2}\right) - \frac{\varphi(\mathbf{m}', n) - \varphi(\mathbf{m}, n)}{2\sigma_d^2 T}\right)\right] \quad (92)$$

The entire implemented transdimensional approach is summarized in the flowchart presented in Figure (37).

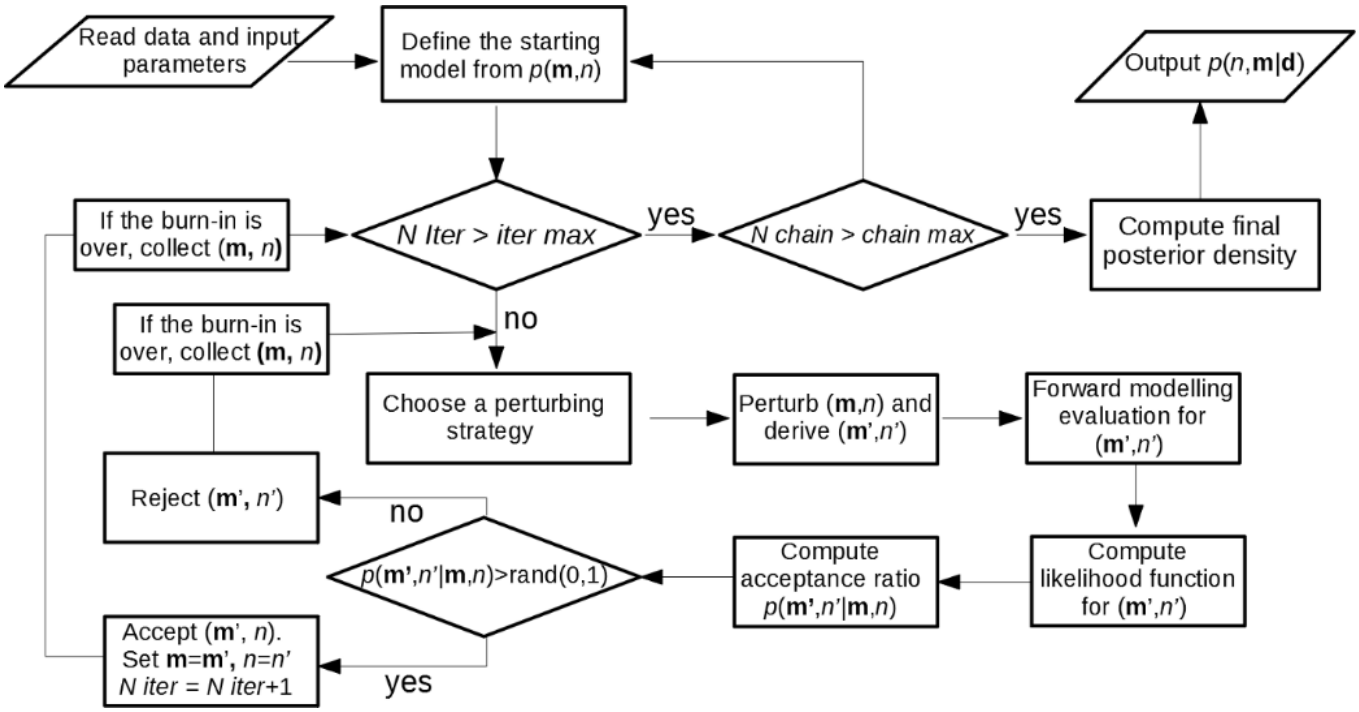


*Figure 37: Flowchart of the implemented transdimensional McMC inversion. Iter max and chain max represent the maximum numbers of iteration and chains, respectively. N iter and N chain are the actual iteration and chain numbers, respectively. Note that for graphical convenience, this chart does not include the parallel tempering strategy and the delayed rejection scheme.*

## 3.4 Inversion Tests

To initially test the stability and robustness of the proposed approach, we start by applying it to synthetic seismic data.

### 3.4.1 Synthetic Inversions

For the present approach, we adopt 20 running chains whose starting models are spread along the model space in accordance with the prior distribution. The first 10 chains have temperature values associated with $T$ = 1, and the remaining 20 chains have temperature values ranging in a logarithmically spaced interval [1,100]. The upper limit value of temperature is set to obtain a satisfactory acceptance ratio. For the elastic inversion, the algorithm runs for 30,000 and the burn-in is set to 3,000, while for the petrophysical inversion, it runs for 20,000 with a burn-in of 2,000 samples.

For the previously described approach, the synthetic CMP gathers are generated through a convolutional model, where a 40 [Hz] Ricker wavelet is assumed as the source signature and the forward Zoeppritz equations constitute the non-linear forward model. The angle range adopted is 0-40 [°]. To better simulate field data, all CMP gathers are again contaminated with Gaussian-distributed white noise with a standard deviation of 0.003.

The algorithms run on parallel codes on a single computer node equipped with 2 deca-core Intel E5-2630 processor at 2.2 GHz and need approximately 3 minutes to invert a single CMP gather for elastic and petrophysical inversions.

To verify that the proposed approach works properly, we start by assessing the algorithm performance in deriving the prior distribution where the low-frequency model is obtained by prior geological analyses over the investigated area. From the original Bayes theorem (equation 3), we observe that if the likelihood value is set equal to 1, the posterior PDF converges to the prior distribution; i.e., the algorithm is forced to accept all models proposed in accordance only with the prior distribution. We illustrate in Figure (38) only the elastic inversion obtained, but analogous results can be derived for the petrophysical inversion.

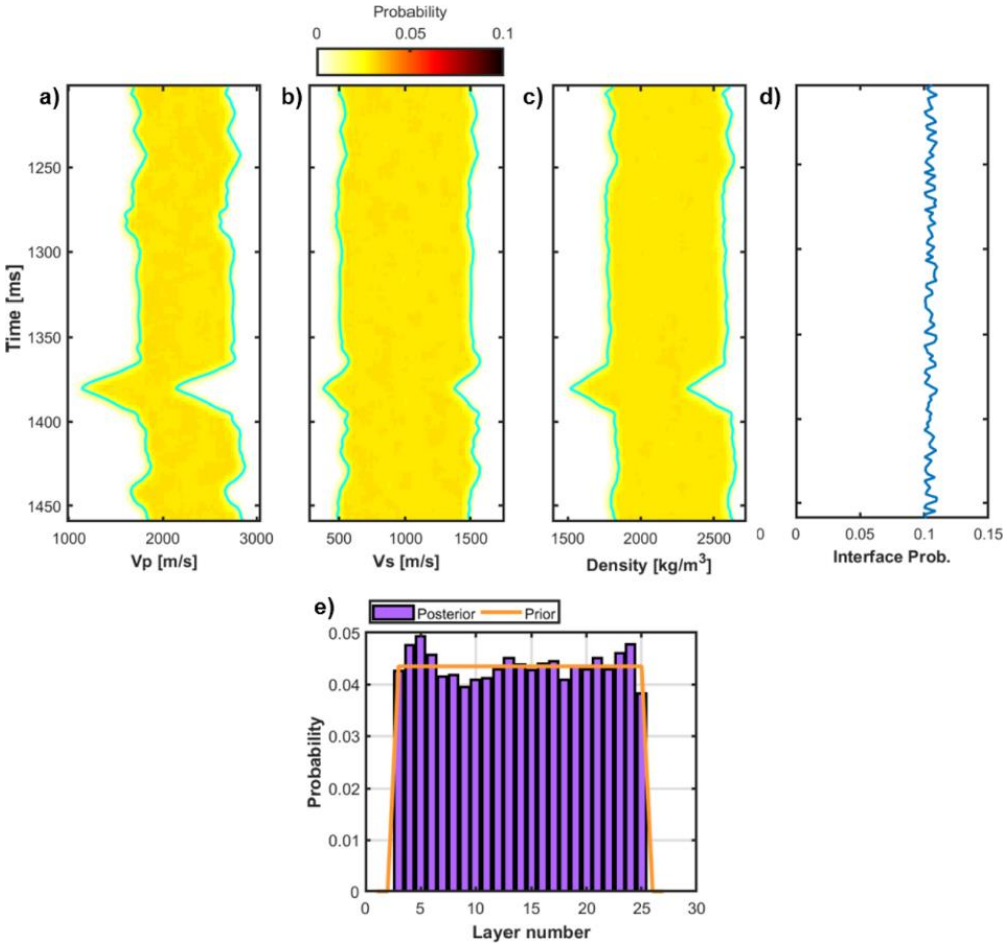


*Figure 38: Marginal PDFs derived by setting the likelihood value to 1 for all the models. a) Marginal PDF for Vp. b) Marginal PDF for Vs. c) Marginal PDF for density. d) Marginal PDF for layer boundaries. e) Marginal PDF for the number of layers. In a), b) and c), the cyan lines represent the admissible parameter range.*

As expected, the results for interface layer locations show that these are uniformly distributed in the allowed range of layer numbers (here fixed at [3-26]), while the elastic variables are almost uniformly distributed in the investigated range. This effectively reveals that in the absence of any information brought by the seismic data, the algorithm is not biased towards any model.

We now apply elastic inversion to a synthetic CMP gather whose elastic properties are extracted from a column of the Marmousi model (Martin *et al.* 2006). The layer interfaces in the reference model are 25, but since a few of them are very thin, below the reference resolution λ/4, we expect the algorithm to reconstruct 22 layers. The prior on the layer number is therefore set over the range [10-40]. The low-frequency model, over which the prior probabilities $p(\mathbf{Vp}|,n)$, $p(\mathbf{Vs}|,n)$ and $p(\boldsymbol{\rho}|,n)$ are defined, is derived as a very smoothed version of the true properties. We recall from Ray and Chopra (2016) that a low-frequency model is always a crucial operator for convolutional-based inversion approaches.

Figure (39) illustrates the obtained elastic inversion results:

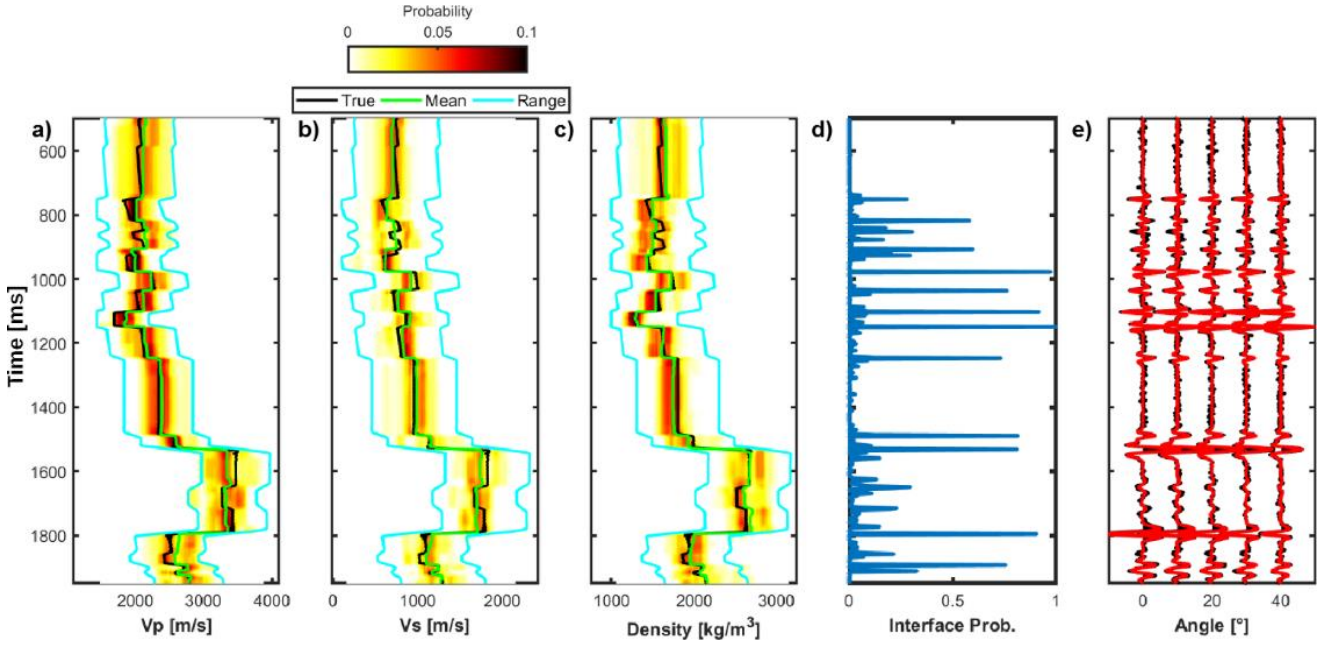


*Figure 39: Marginal PDFs derived in the elastic AVO inversion of synthetic data. a) Marginal PDF for Vp. b) Marginal PDF for Vs. c) Marginal PDF for density. In a), b) and c), the cyan lines represent the admissible parameter range, the black lines show the true properties, and the green lines depict the estimated mean model. d) Posterior probability density functions for the interface locations. e) Comparison between observed (black) and predicted (red) data computed on the mean estimated model.*

These results highlight the great algorithm performances in deriving the actual vertical profiles of elastic variables, where almost every interface is reconstructed with optimal agreement marking the major strong contrasts. Furthermore, the predicted data, show how the actual subsurface complexity has been correctly derived with a high level of confidence by the proposed algorithm.

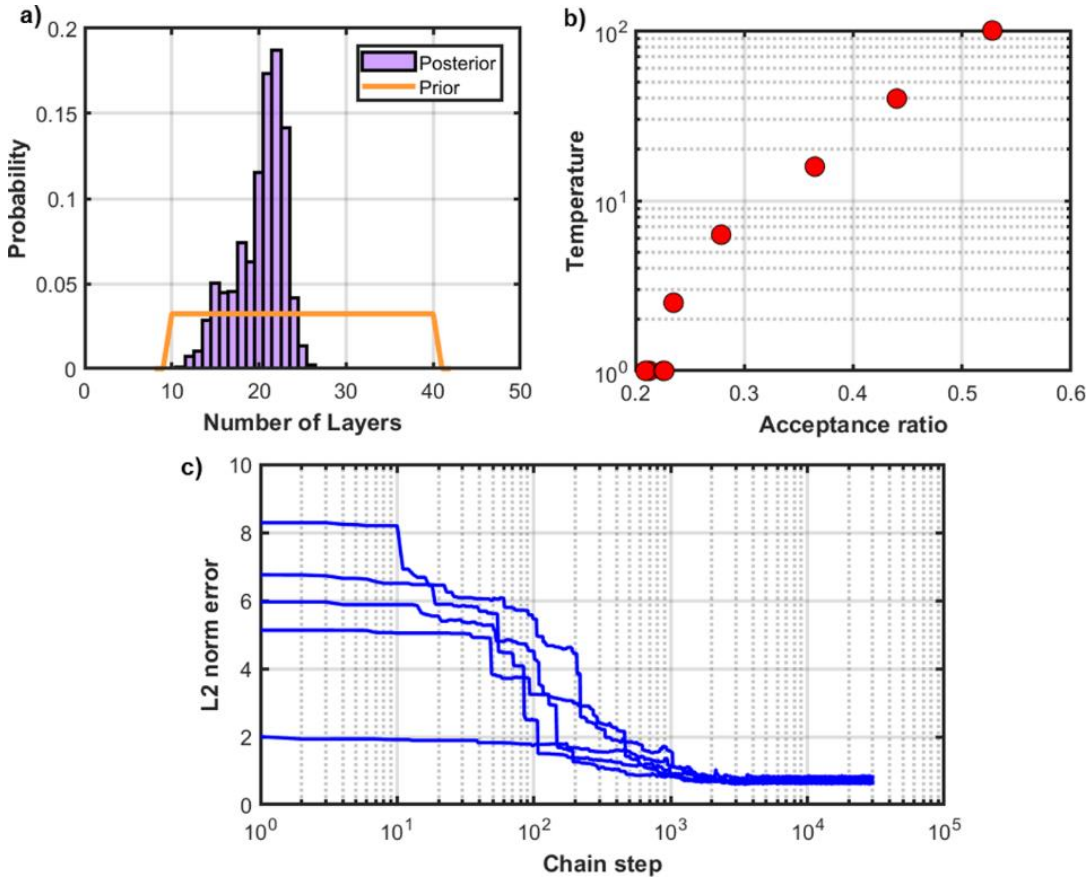


*Figure 40: a) Marginal PDF for the number of layers in the elastic AVO inversion. b) Variation of the acceptance ratio with changing temperature values. c) Example of evolution of the L2 norm data misfit for 5 chains.*

The analysis of the PDF (Figure 40.a) over the number of layers shows that a higher probability is associated with models characterized by 22 layers, as expected. Figure (40.b) represents the ratio between the acceptance ratio and temperature values associated with the PT adopted approach. From this figure, we can appreciate how higher temperature values, associated with flatter likelihood distributions, effectively generate an increase in accepted models. From Figure (40.c), we finally observe the DRS and PT effects on the model exploration for 5 running chains, where approximately 3,000 iterations are needed to reach a stationary regime.

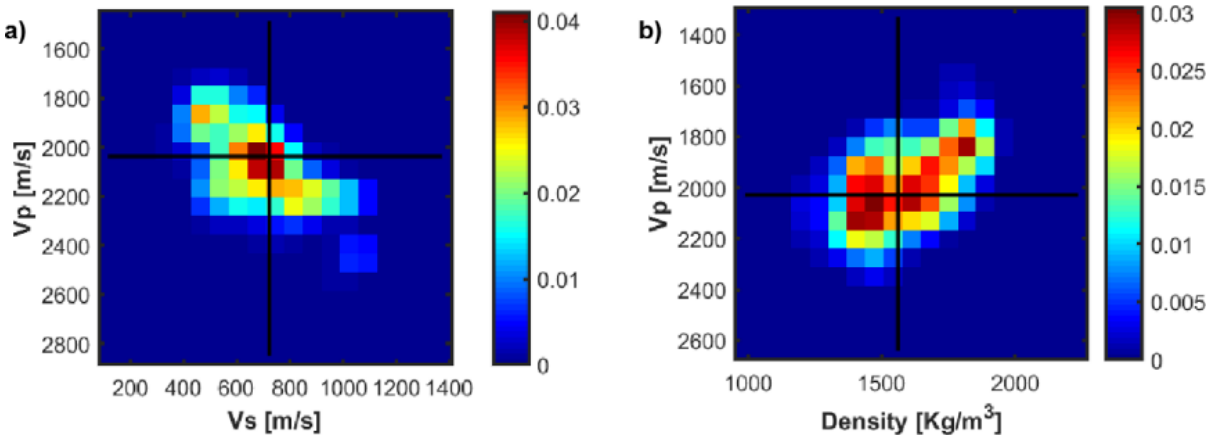


*Figure 41: Example of 2D marginal pdfs for Vs-Vp (a) and Vp-density (b) computed from the results shown in Figure 3 at a given vertical position. The solid black lines show the correct elastic property values.*

Another interesting analysis of the algorithm performances is the comparison of the marginal PDFs associated with a fixed time sample projected in the Vp/Vs and Vp/ρ space as shown in Figure (41). We can effectively observe how the true model, identified by the black crosses, is perfectly identified by the ensemble of sampled models. From these figures, we can further analyse the effects of single elastic parameters on the AVO response. According to the theoretical Aki-Richards equations, we know that Vp and density exert proportional impacts (both with positive signs), while Vp and Vs generate opposite effects. This means that to hold the reflection coefficient constant over the considered range of incidence angles, an increase in the P-wave velocity must be counterbalanced by a decrease in density and by an increase in S-wave velocity.

We now describe the results obtained for the synthetic petrophysical inversion. The geological background is represented by a clayey sequence, encasing gas-saturated and brine-saturated sands. A reference model with 11 layers is built, with a smaller range of depth than that of the elastic inversion. The prior distribution over the number of layers is [2,20], the Vp/Vs ratio is set within [1.6,2.6] and the ρ/Vp ratio is [0.8,1.2]. The forward model operator, the source signature, and the range of angles are the same as those for the previous elastic inversion.

The results again show the great algorithm performance. In particular, we note that all interface locations are correctly identified and that the PDFs associated with each petrophysical variable derive the true vertical profile with high accuracy. From the analysis of the posterior distributions, we particularly note that the uncertainties related to porosity are significantly smaller than those related to the other parameters (note that such analysis cannot be easily derived graphically from Figure (42) due to the different ranges of definitions of the petrophysical variables). This effect is

expected since we know that porosity plays a major role in defining the elastic parameters and therefore the seismic data. The analysis of the posterior distribution over the number of layers, reported in Figure (43.a), again confirms that the most likely distribution is associated with the 11-layer configuration that effectively represents the true model. Figure (43.b) shows the increment of accepted models as the temperature increases, and Figure (43.c) shows the L2 norm error associated with 5 running chains, from which we can derive the burn-in around the 3,000$^{th}$ sampled model.

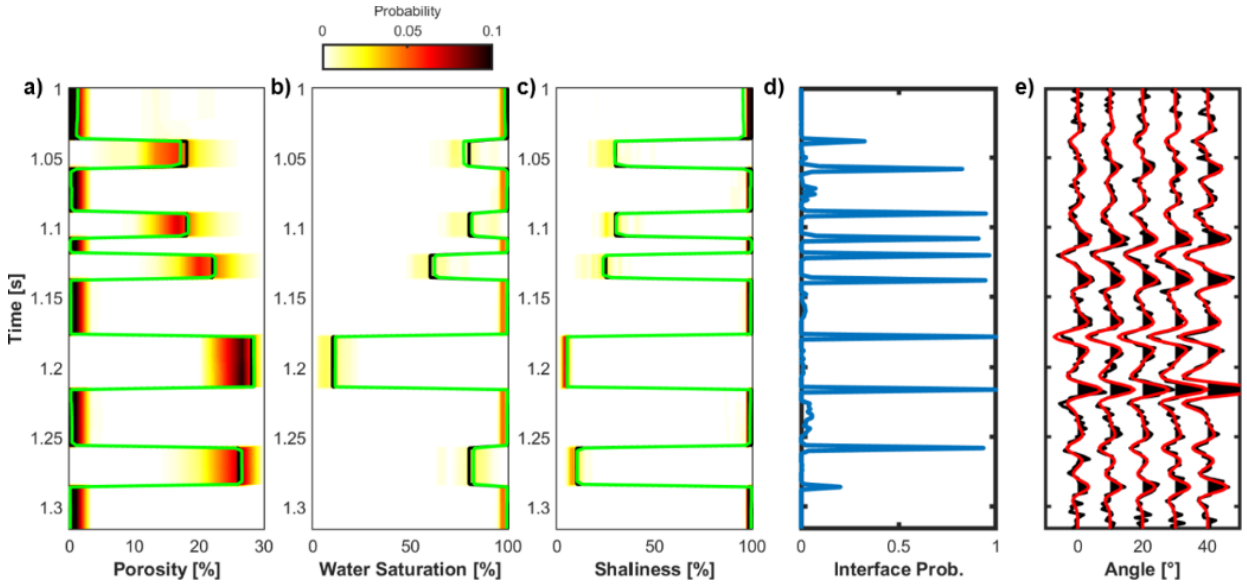


*Figure 42: Marginal PDFs derived in the petrophysical AVO inversion test of synthetic data. a) Marginal PDF for porosity. b) Marginal PDF for water saturation. c) Marginal PDF for shaliness. In a), b) and c), the yellow lines represent the admissible parameter range, the black lines show the true properties, and the green lines depict the estimated mean model. d) Posterior probability density functions for the interface locations. e) Comparison between observed (black) and predicted (red) data computed on the mean estimated model.*

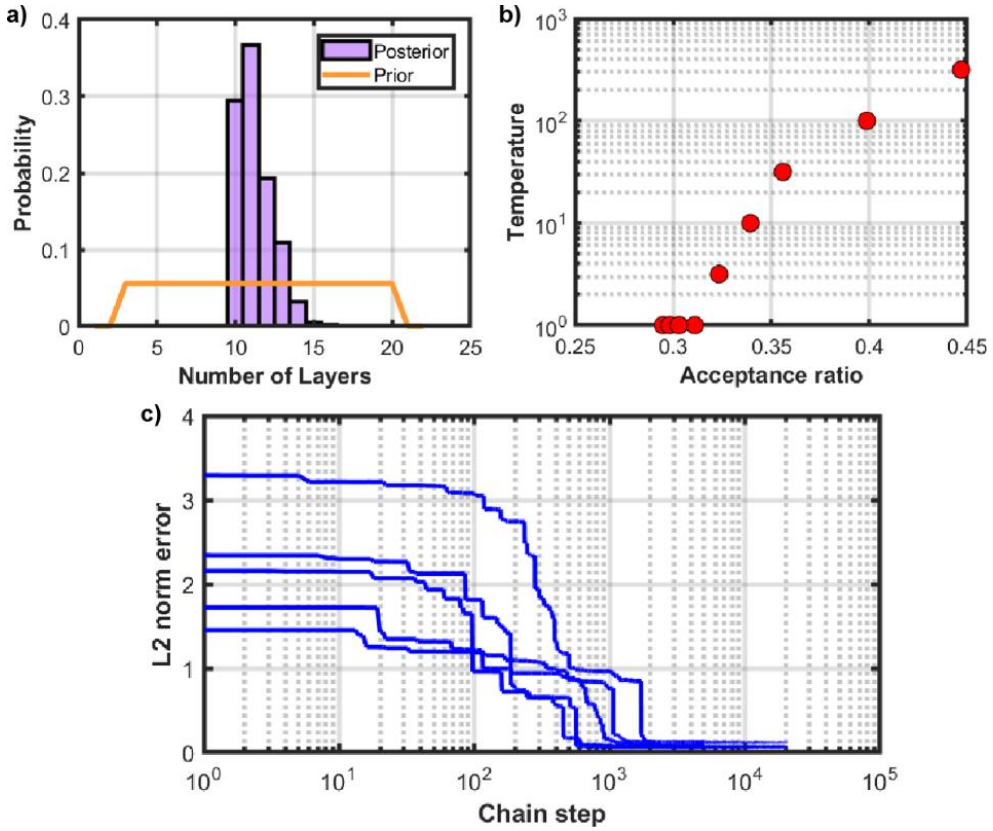


*Figure 43: a) Marginal PDF for the number of layers in the petrophysical AVO inversion. b) Variation in the acceptance ratio with changing temperature values. c) Examples of evolutions of the L2 norm data misfit for 5 chains.*

### 3.4.2 Field Data Inversions

Given the promising results from synthetic seismic data, we now apply the proposed approach to field data.

The adopted algorithm is applied to data from an onshore gas field, where the geological setting is dominated by a turbiditic sequence hosting a gas-saturated sandy body. The elastic contrast associated with the encased reservoir can be observed in the close-up view presented in Figure (44), where it generates high-amplitude reflections.

Note that since the reservoir is located at a depth of approximately 1100 [m], we assume that stratigraphic compaction played a minor role in this reservoir geometry configuration; therefore, a soft-sediment approach for the RPM definition (see previous RPM section) is more appropriate.

The available information on the acquired seismic data is as follows: dominant frequency of 40-45 [Hz], maximum available offset of 4.5 [km] and time sampling of 4 [ms].

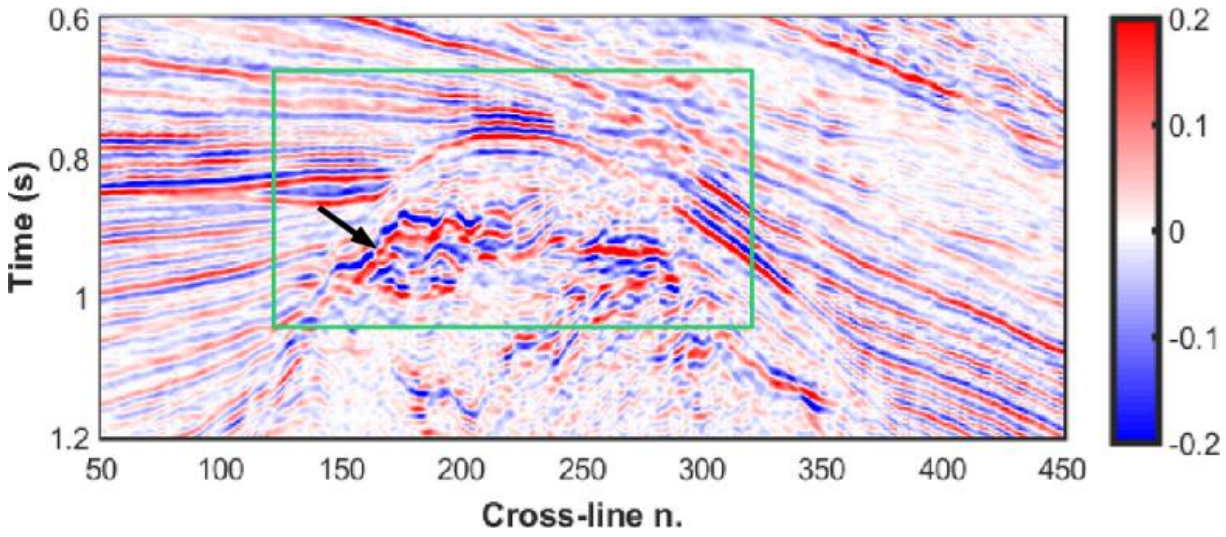


*Figure 44: Example of a stack section along an in-line direction extracted from the 3D seismic volume. The green rectangle delimits the target zone where the inversion algorithms are applied, while the black arrow points towards the top reflections of the reservoir.*

The proposed algorithms have been benchmarked again the commonly used elastic linear Bayesian inversion approach proposed by Buland and Omre (2003) and the petrophysical linear Bayesian derivation proposed by Aleardi *et al.* (2018a). Since these approaches invert for P- and S-wave seismic impedance, we convert the elastic results obtained by the transdimensional McMC algorithm to this other domain.

While both of the adopted linear methods make use of a vertical variogram to include strong lateral constraints, the McMC method does not. To slightly attenuate the noise among adjacent CMP gathers, a simple three-point moving average window is therefore applied to the proposed numerical approach.

The low-frequency model that our numerical approach uses is derived by integrating the available borehole information and by the results of the proposed linear inversion approaches.

The data covariance matrix is directly calculated over the available data under the assumption of uncorrelated noise along incidence angles. We assume that adjacent CMP gathers share the same elastic variables; therefore, differences are only due to noise contamination (Aleardi *et al.* 2018a).

For both the elastic and petrophysical inversions, the set hyper-parameters are as follows: the prior over the layer number is (6-30), the Vp/Vs ratio is set from (1.6-2.6), while the ρ/Vp ratio is (0.7-1.2). We again make use of 20 running chains at temperature $T$ = 1 and 20 other chains at logarithmically spaced temperature values in the range (1-150). Chains run for 30,000 steps, and a burn-in of 3,000 is assumed. Due to the reservoir geometry and depth, available incidence angles ranged from 20-40 [°].

Both inversions run in approximately 6 hours on parallel codes on a computer with 2 deca-core Intel E5-2630 processor at 2.2 GHz.

#### 3.4.2.1 Elastic Inversion

Figure (45) reports the observed impedance values obtained with the linear Bayesian inversion and the transdimensional approach. Note the high similarity, especially in retrieving the anticlinal trap, and the strong amplitude contrast that marks the reservoir layer. Figure (46) shows the posterior standard deviation for the predicted elastic Ip and Is variables, where we observe that the strongest variations are again associated with the reservoir layer.

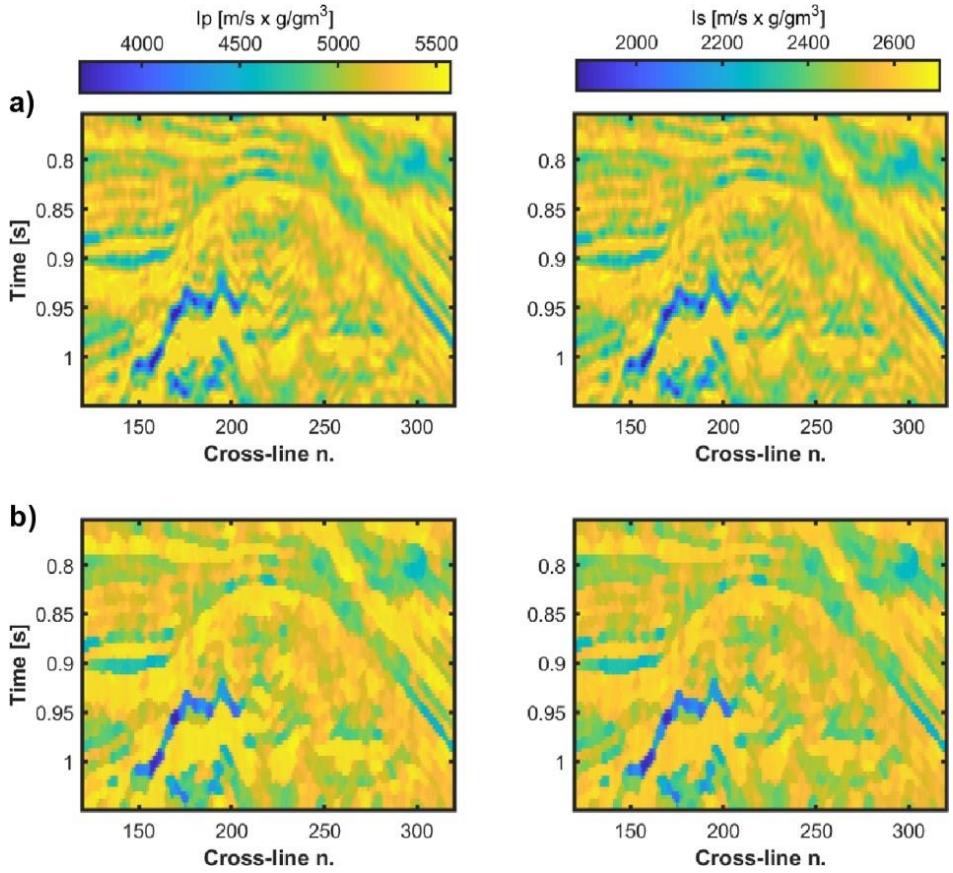


*Figure 45: Mean elastic properties of Ip and Is predicted along the considered 2D section by a linear Bayesian inversion (a) and by the implemented Rj-McMC algorithm (b). Ip and Is are presented on the left and right, respectively.*

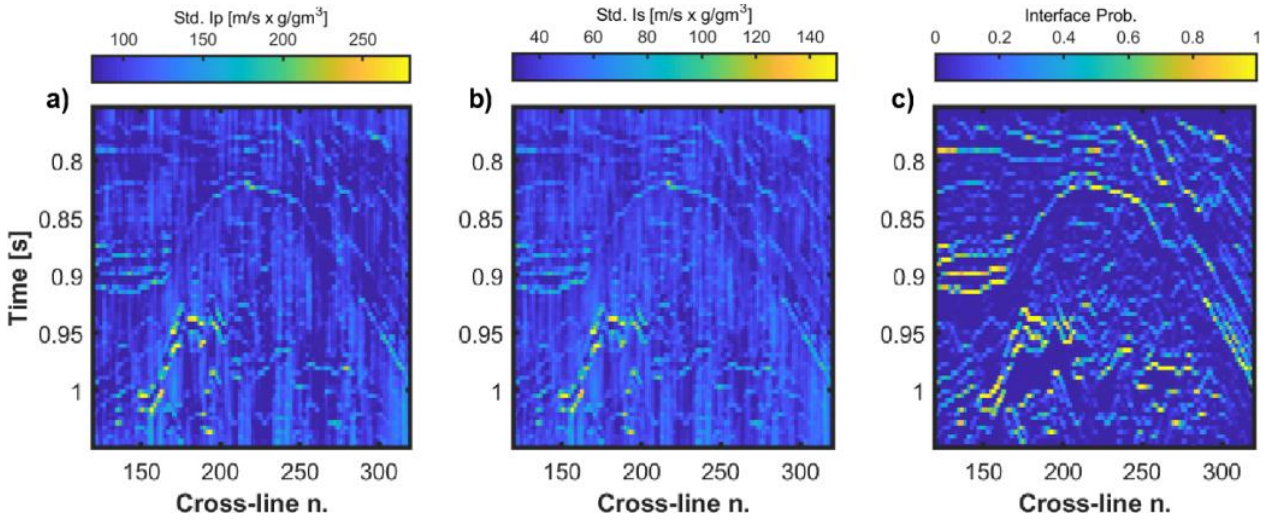


*Figure 46: a) Standard deviation affecting the Ip model predicted by the Rj-McMC inversion. b) Standard deviation for Is. c) Marginal PDF for interface locations along the considered in-line section.*

This effect is typical of the existing trade-off between uncertainty in the model and uncertainty in the spatial location, where "a stronger elastic contrast causes higher uncertainty in the model property values at the interface location, but lower uncertainty in the location of the discontinuity itself" (Zhu and Gibson 2018). The strong differences among the CMP gathers can confidently be related to the different S/N values affecting adjacent positions. Note how the highest probability interface is again related to the stronger elastic reservoir contrast, meaning that almost all collected models correctly identify the target reservoir.

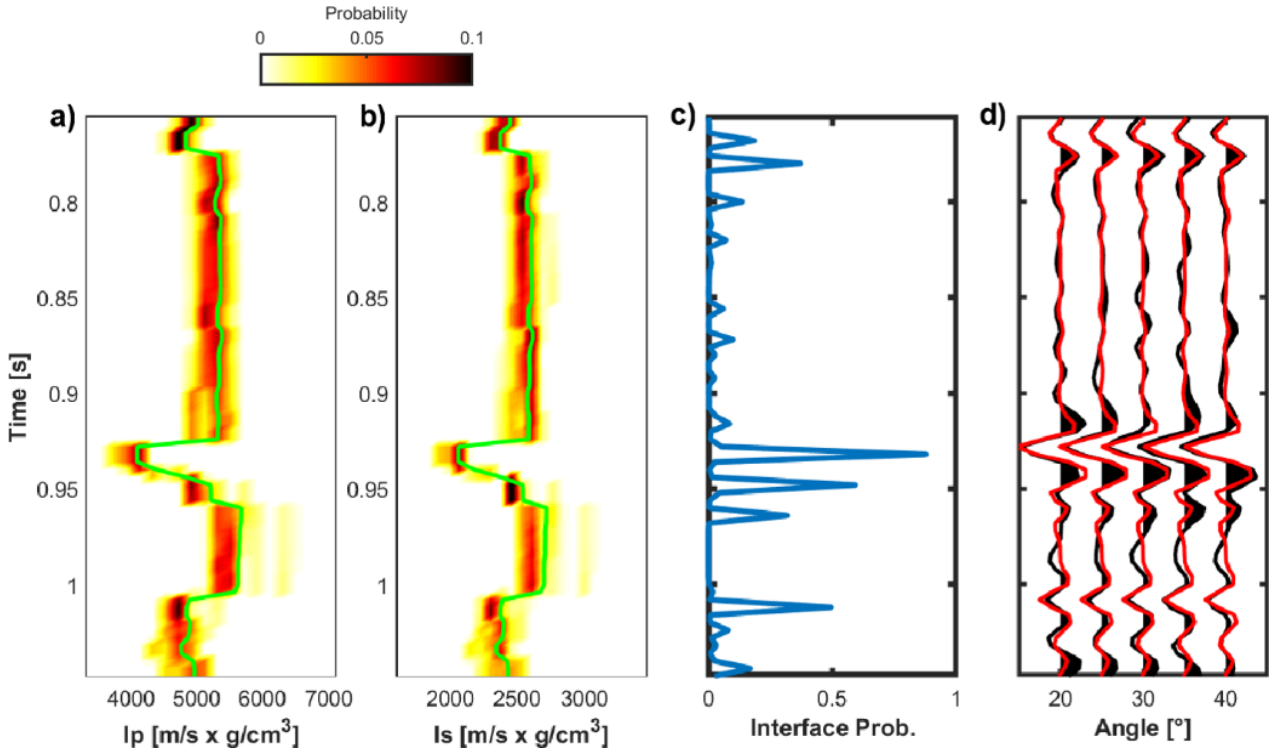


*Figure 47: First example of inversion results provided by the elastic Rj-McMC inversion for a given CMP location extracted from the 2D section of Figure 44. a) Mean Ip model (green line) superimposed on the corresponding marginal PDF. b) Mean Is model (green line) superimposed on the corresponding marginal PDF. c) Posterior probability density function for the interface locations. d) Comparison between observed (black) and predicted (red) data computed for the mean estimated model.*

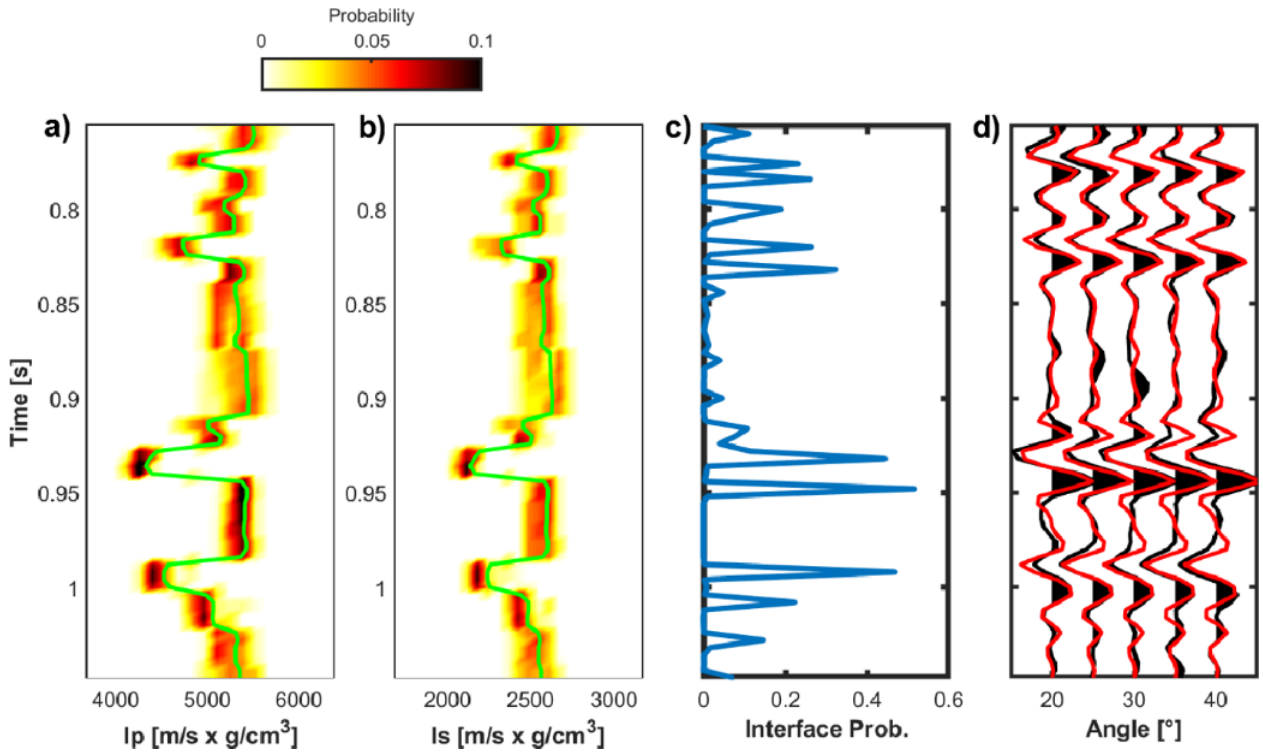


*Figure 48: As in Figure 47 but for a different CMP gather.*

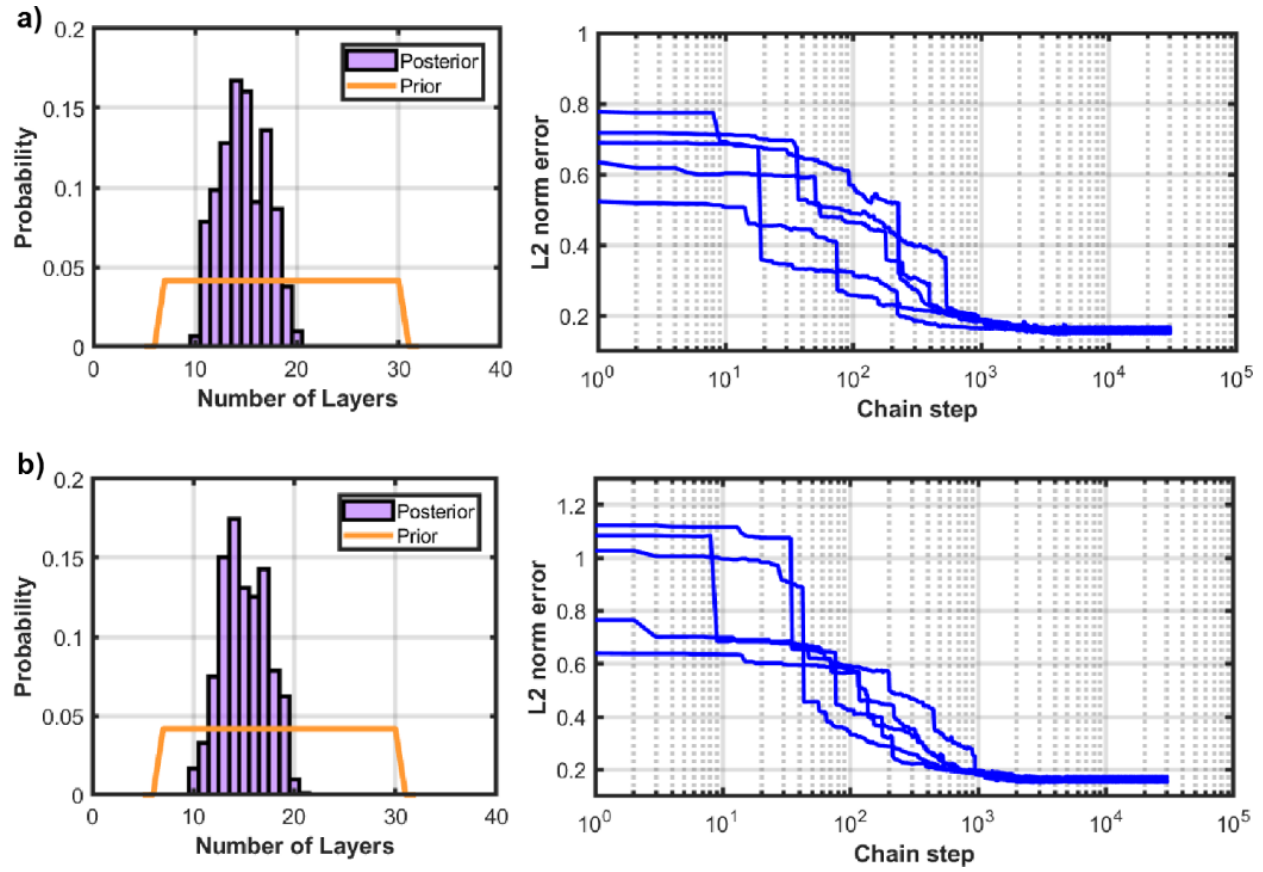


*Figure 49: Left side: Marginal PDFs for the number of layers for the two CMPs shown in Figures 47 and 48 (panels a and b, respectively). Right side: Examples of evolution of the L2 norm data misfit for the two CMP gathers shown in Figures 47 and 48 (panels a and b, respectively).*

Analogous results can be obtained from the analysis of Figures (47) and (48), where the derived elastic variables over two different CMP gathers extracted by the same line of the Marmousi model are illustrated. We can observe how the vertical elastic profile is correctly matched with a high level of confidence and the boundary locations are clearly derived. From the analysis of the posterior distribution (Figure 49), we note that in both cases, the optimal model parametrization produces approximately 15 layers. From the misfit of 5 running chains, we finally observe that, similar to the synthetic inversions, approximately 3,000 sampled models are necessary to achieve the stationary regime.

#### 3.4.2.2 Petrophysical Inversion

Figure (50) shows the petrophysical inversion derived with the linear Bayesian approach and the transdimensional McMC. As in the previous elastic case, we here note the good agreement between the proposed approaches, especially in identifying the strong variation associated with the reservoir layer, characterized by high porosity values and sharp decreases in the water saturation and shaliness content. As expected, we quantify an increased uncertainty in the posterior standard deviation associated with the latter petrophysical parameters (Figure 51), which confirms what is deduced by the analytical inversions: the stronger influence of the porosity than that of the remaining petrophysical parameters on the elastic properties and the AVO response.

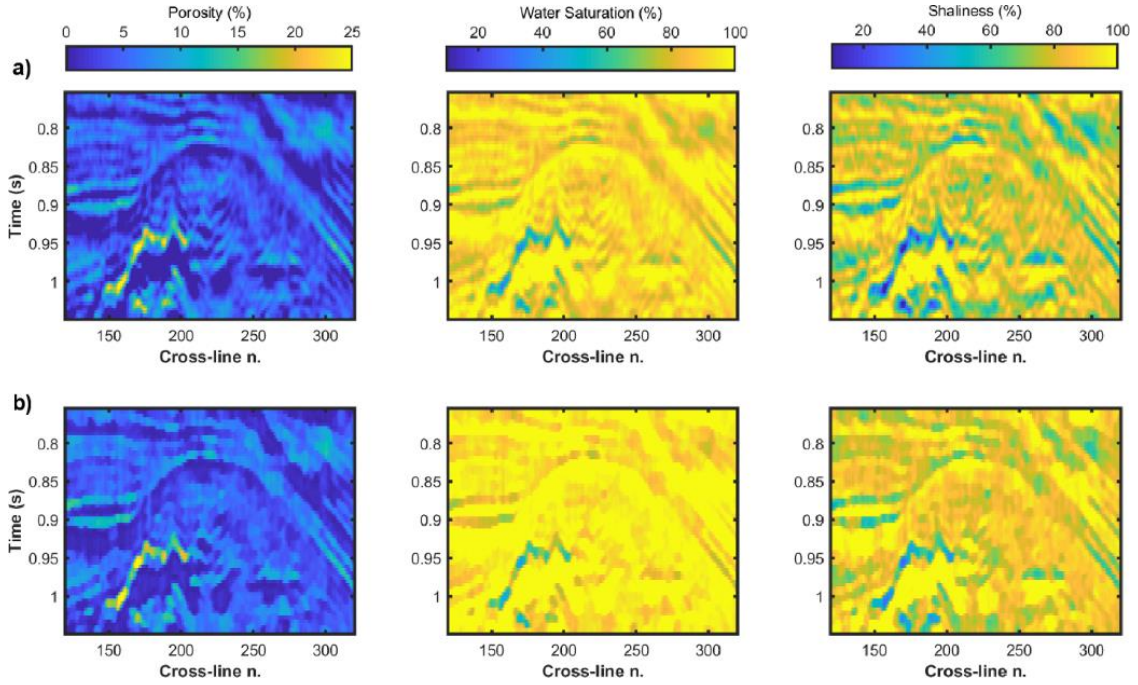


*Figure 50: Mean petrophysical properties of porosity, water saturation and shaliness predicted along the considered 2D section by the linear Bayesian petrophysical inversion (a) and by the implemented Rj-McMC algorithm (b). In a) and b) porosity, water saturation and shaliness are represented from left to right.*

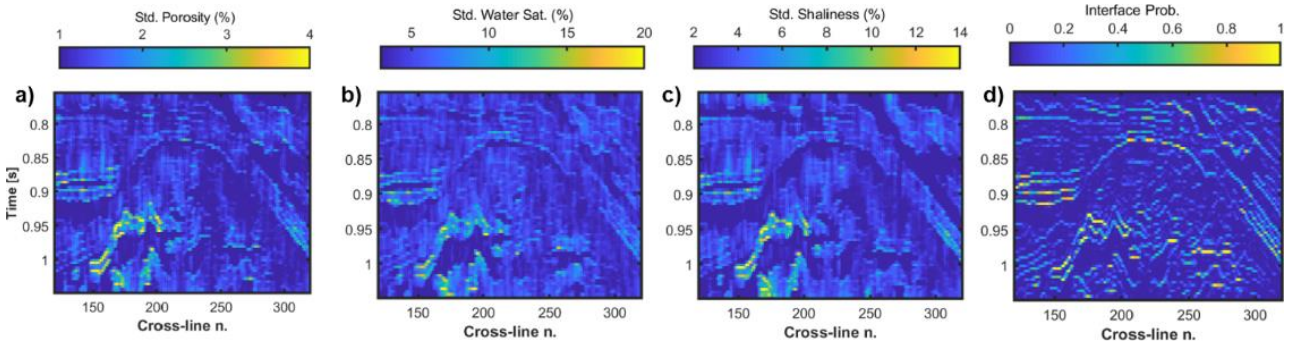


*Figure 51: a), b) and c) Standard deviation for porosity, water saturations and shaliness (panels a, b and c, respectively) predicted by the Rj-McMC inversion along the considered 2D section. d) Marginal PDF for interface location along the inverted in-line section.*

Similar results can be obtained from the analysis of the CMP gathers extracted for the elastic inversion, as shown in Figures (52) and (53). In particular, we observe that the reconstructed vertical profile allows matching the observed seismic data with a very high level of confidence, that low uncertainty affects the estimated porosity model and that the uncertainty in the water saturation estimates is significant.

Analogous to the elastic inversions, the petrophysical inversions in Figure (54) show that the posterior distribution for the number of layers is again approximately 15 layers, showing the power

and stability of the proposed approach, while approximately 3,000 iterations are needed to attain the stationary regime.

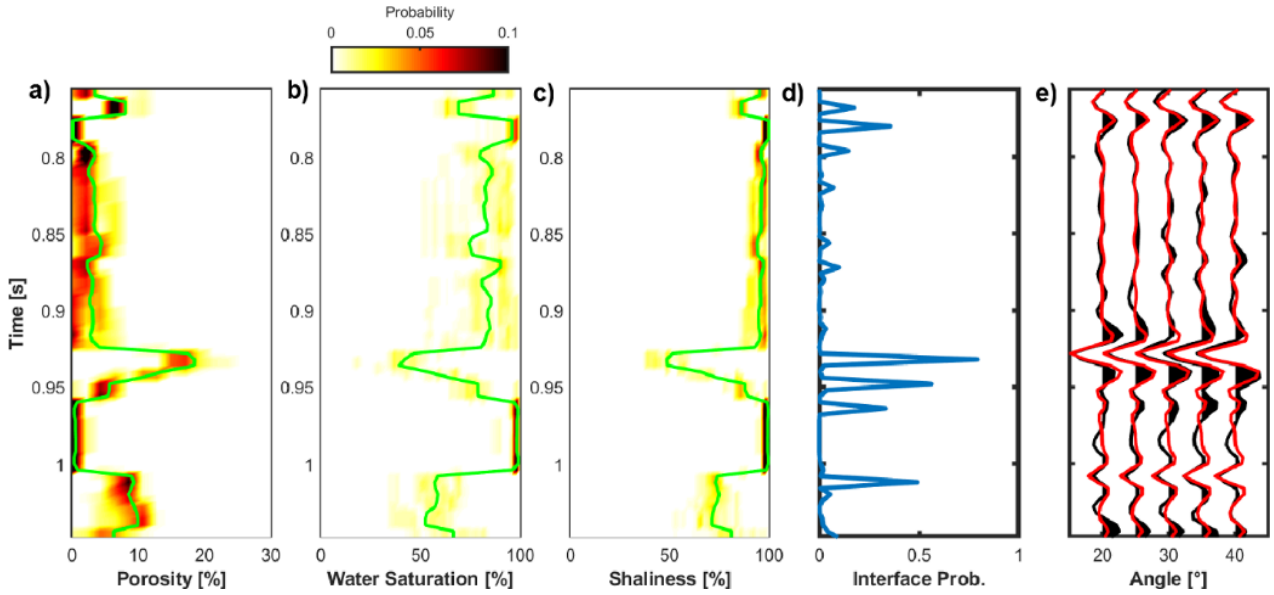


*Figure 52: First example of inversion results provided by the petrophysical Rj-McMC inversion for the CMP gather already shown in Figure 47. a), b) and c) represent the mean porosity, water saturation and shaliness models (green lines), respectively, superimposed on the corresponding marginal PDFs. d) Posterior probability density function for the interface location. e) Comparison between observed (black) and predicted (red) data computed for the mean estimated model.*

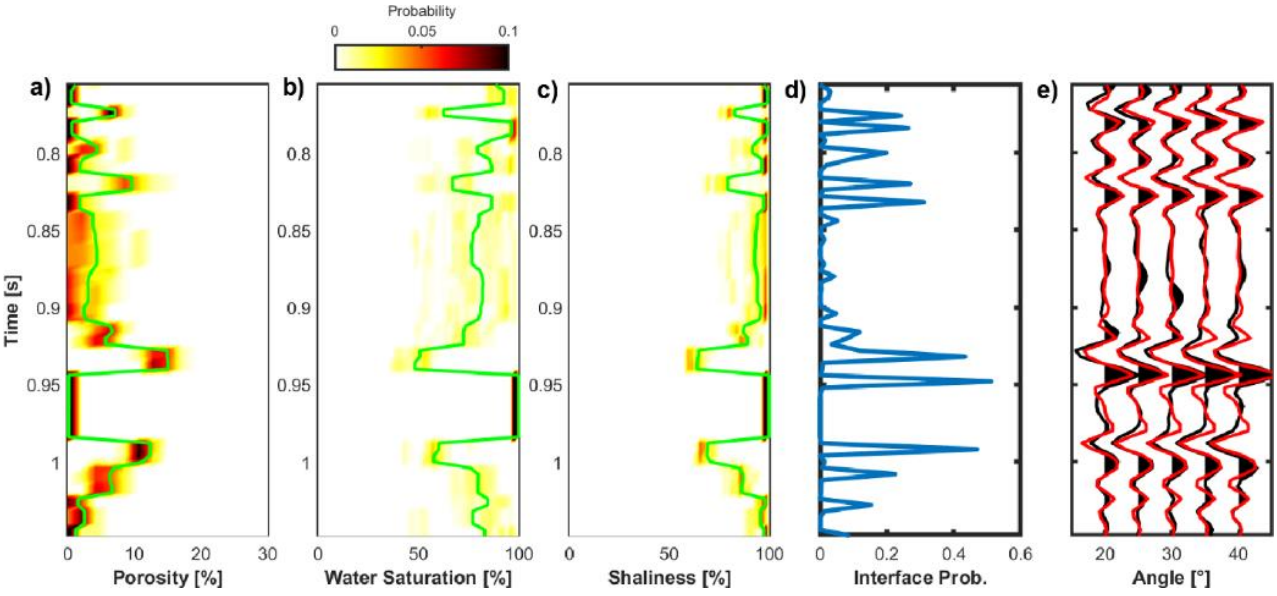


*Figure 53: As in Figure 52 but for the CMP gather already shown in Figure 48.*

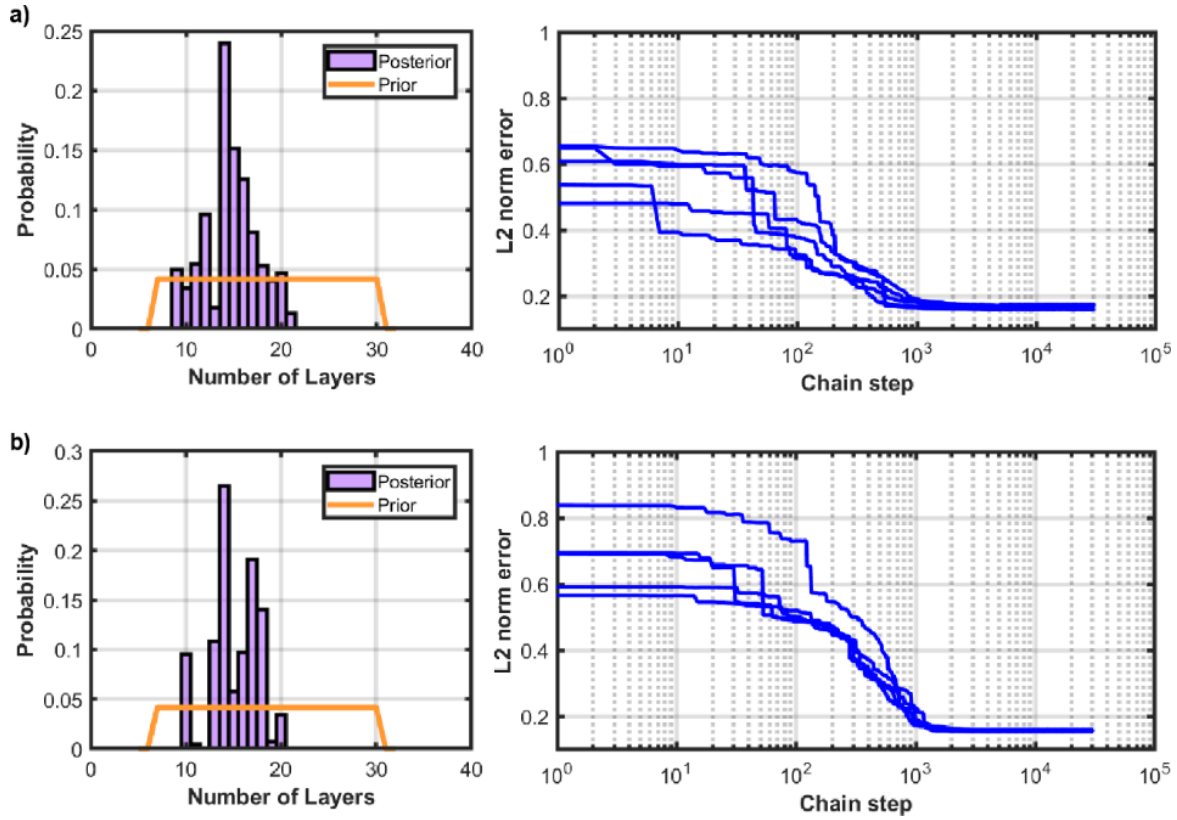


*Figure 54: Left side: Marginal PDFs for the number of layers for the two CMPs shown in Figures 52 and 53 (panels a and b, respectively). Right side: Examples of evolution of the L2 norm data misfit for 5 chains for the two CMP gathers shown in Figures 52 and 53 (panels a and b, respectively).*

## 3.5 Conclusions

Here, we have presented a transdimensional McMC algorithm for elastic and petrophysical AVO inversion of pre-stack seismic data.

The proposed approach demonstrates its ability to reconstruct real complex subsurface geometries compared to classic analytical methods. The reliability of this transdimensional algorithm allows it to consider the model parametrization as an additional unknown generating different discretized models, which reduces the ill-conditioning of the problem and optimally derives the best model parameterization based on the statistical match with the observed data.

Its ability lies in the possibility of not only accurately assessing the posterior model in the case of non-linear forward models but also and more importantly providing stable and reliable predictions with a minimum of external prior constraints included in the inversion kernel. The inversion

framework indeed automatically adjusts the underlying model parametrization to produce solutions with appropriate levels of complexity to fit the data to statistically meaningful levels.

The application of meta-algorithms such as parallel tempering and the delayed rejection scheme greatly reduces the computational effort that is usually required for all numerical approaches. Even so, the extension of the described approach to 3D seismic data implies unfeasible computational efforts, for which a more optimized procedure needs to be applied. Furthermore, we remark the difficulty of optimally setting all the required hyper-parameters, which otherwise easily drive the inversion to unstable results.

Actual and more advanced McMC methods involve a Hamiltonian-based inversion approach to further increase the model space exploration and more accurately derive the posterior distribution, simultaneously reducing the correlations among adjacent sampled models and increasing the acceptance ratio.

# CHAPTER 4

# A transdimensional approach to generate data-driven constraints.

In Chapter 2 we saw that one of the most popular strategies devoted at attenuating the ill-conditioning of inverse problems is the inclusion of lateral constraints in the inversion framework. In Chapter 3 we proposed an alternative approach to the geostatistical constraints using a transdimensional method that inherently reduces the ill-conditioning of the problem by considering the model parametrization as unknown the algorithm solve for. In Chapter 4 we propose another approach that exploits its transdimensional nature to automatically infuse constraints in the inversion framework without any prior assumption about the parameters distribution, but simply through a data-driven regularization strategy.

## 4.1 Introduction

The most popular strategies devoted at attenuating the ill-conditioned inverse problems are the inclusion of lateral constraints into the inversion framework. Several advanced regularization strategies exist, such as the inclusion of geostatistical constraints in the form of isotropic model correlation functions (Buland *et al.* 2003), multipoint statistics (Cordua *et al.* 2012, Hansen *et al.* 2012), stratigraphic constraints (Tetyukhina 2011) or constraints derived by the seismic bandwidth (Haas and Dubrule 1994). The main limit of all these approaches is that they rely on prior structural knowledge of the investigated area and force the recovered model to honor such prior constraint. These are all essentially model-driven regularization strategies that could provide biased model parameter estimations in case of erroneous prior assumptions.

We here propose another transdimensional approach that uses data-driven regularization strategies: the purpose is to locally adapt the automatically-inferred constraints to the structural properties of the subsurface model from the local characteristics (i.e. variability) of the observed data.

On the line of these data-driven approaches, we here present a numerical transdimensional Markov chain Monte Carlo (Rj-McMC) algorithm approach for target-oriented AVO inversion to derive the number of Voronoi cells, the position of Voronoi nuclei, and the elastic properties associated to each polygon of a synthetic 2D horizon. Since generally all transdimensional problems are characterized by lower acceptance ratios compared to other McMC methods, we again adopt a PT strategy method to further increase the algorithm potentialities.

## 4.2 The adopted Method

As previously deeply discussed, in a transdimensional inversion the number of model parameters (that codes the optimal subsurface model parameterization) is considered unknown and is estimated using a probabilistic sampling.

In our case the inverted 2D synthetic horizon is composed of Voronoi cells, whose number and shape are automatically determined by the Rj-McMC sampling. The algorithm hence autonomously partitions the considered 2D synthetic horizon on the basis of the spatial variability of data, producing subsurface 2D models discretized in Voronoi polygons each one enclosing CDP positions with similar AVO responses. This means that all the CDPs falling within the same Voronoi polygon share similar elastic properties and for this reason an average of these elastic property values is assigned to these CDPs. Similarly, the observed data for each polygon is computed by averaging AVOs responses of all the CDPs falling within each Voronoi cell, as graphically reported in figure (55).

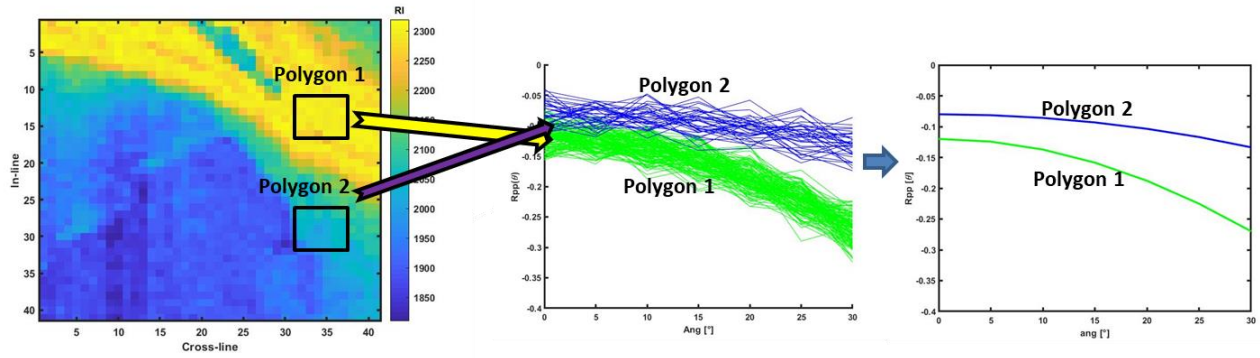


*Figure 55: The inversion adopted method. The 2D synthetic area to invert is automatically divided into polygons that share analogous AVOs responses. These polygons are subsequently averaged and the mean value is associated to each Voronoi polygon.*

From the one hand, this strategy constitutes a data-driven approach to reduce the ill-conditioning of the problem by including lateral constraints into AVO inversion: such constraints are automatically inferred from the lateral variability of the data and not arbitrarily infused into the inversion framework. From the other hand, averaging AVO responses pertaining to CDPs falling within the same cell inherently increases the S/N ratio of the observed data. These two aspects turn out to be crucial for stabilizing the inversion even in case of severely noise-contaminated data.

In order to fully explore the method potentialities, the implemented inversion is here applied to synthetic seismic data characterized by different S/N ratios and is benchmarked against a more standard Bayesian AVO inversion without lateral constraints.

### 4.2.1 The Forward Model

Zoeppritz equations describe the partitioning of wave seismic energy at an interface in term of P-wave velocity, S-wave velocity and density of the upper and lower layers. In order to reduce the ambiguity within the AVO method (Drufuca and Mazzotti 1995) and further constrain the inversion, the number of unknowns can be reduced. For present work we adopt the Ursenbach and Stewart equations (Ursenbach and Stewart, 2008) as forward modelling that describes the system in terms of relative contrasts of P-impedance and S-impedance ($R_I$ and $R_J$, respectively).

$$Rpp(\Theta) = \left(1 + \sin^2\Theta \frac{4\gamma^2 + cos^2\Theta - 1}{5}\right)\frac{R_I}{cos^2\Theta} - 8\gamma \sin^2\Theta R_J \ . \quad (93)$$

where *Rpp* is the P-wave reflectivity, $\Theta$ is the angle of incidence across the interface, $\gamma = \frac{Vs1+Vs2}{Vp1+Vp2}$ with the subscripts 1 and 2 respectively refers to the upper and lower layer,

$$R_I = \frac{\Delta Ip}{2\,\widetilde{Ip}};\ R_J = \frac{\Delta Is}{2\,\widetilde{Is}} \ . \quad (94)$$

where Ip and Is are respectively the P-wave and S-wave impedance; the operator Δ indicates the contrast and the symbol ˜ indicates the mean across the interface.

### 4.2.2 Space Discretization

Here, the model space is discretized using the Voronoi polygons over equally spaced grid points representing the CDP positions. A number of discrete set of points, the center of Voronoi cells, partitions the plane and the cells are composed by those points whose distance from the center is smaller. So polygons don't overlap each other (all points of the grid fall in just one polygon) and the parametrization is totally defined by the position of the polygons nuclei. Our models are therefore represented by all the CDPs over the grid, to each of whom we associate a P-impedance and S-impedance value. Figure (56) shows an example of Voronoi model space discretization which forms an irregular set of cells that partitions the plane.

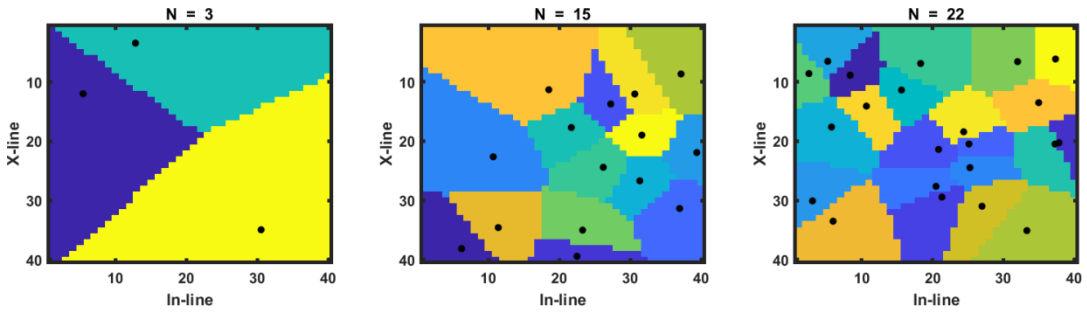


*Figure 56: Example of model space partition in 3, 15 and 22 randomly distributed polygons. Black dots indicate the nuclei centers.*

### 4.2.3 Proposed Inversion Procedure

Given the analogy with the previously proposed inversion approach, we do not repeat here the analogous calculations to derive the likelihood, the prior and the proposal models with the relative acceptance probabilities and the Metropolis-Hastings acceptance rules.

For each chain the inversion starts from a single Voronoi cell representing a laterally invariant $R_I$ and $R_J$ value drawn from the prior distribution. The values of $n_{min}$ and $n_{max}$ defining the prior range of the Voronoi cells number are respectively equal to 1 and three-quarters of maximum number of CDPs within the inverted area (chosen for computational reason).

To all the $N$ CDPs positions falling within the same cell are assigned the same elastic property values that are computed as the average of their $R_I$ and $R_J$ values. Similarly, the averaged AVO responses of the $N$ CDPs constitute the observed data for the considered cell. Then the algorithm evolves by sampling the model space, that is by sampling the $R_I$ and $R_J$ values, the number of cells, and the positions of their nuclei.

During the iterations the algorithm tends to gather within the same Voronoi cell adjacent CDPs with similar AVO responses to which are assigned the elastic properties that produces a good fit with the observed data. For each iteration, the algorithm applies a perturbation to the current model chosen with equal probability from the following list:

1- *Death move*: One polygon from the Voronoi tessellation is deleted and the shape of the remaining polygons based on the positions of their nuclei is rearranged (Figure 57 B).

2- *Cell move*: One Voronoi cell is randomly chosen and the position of the corresponding center is changed without modifying the associated $R_I$ and $R_J$ values. This perturbation produces a slightly rearrangement of the Voronoi tessellation over the considered 2D horizon (Figure 57 C).

3- *Elastic move*: One Voronoi cell is randomly chosen and the $R_I$ and $R_J$ values for all the CDP positions enclosed in the selected cell are perturbed. This perturbation follows a Gaussian proposal centered on the current $R_I$ and $R_J$ values (Figure 57 D).

4- *Birth move*: A new polygon within the Voronoi tessellation is created and the elastic properties of the newly created polygon are assigned drawing a random realization from the prior $R_I$ and $R_J$ distribution. Note that only the neighboring cells of the new-born cell have their geometry changed during this step (Figure 57 E).

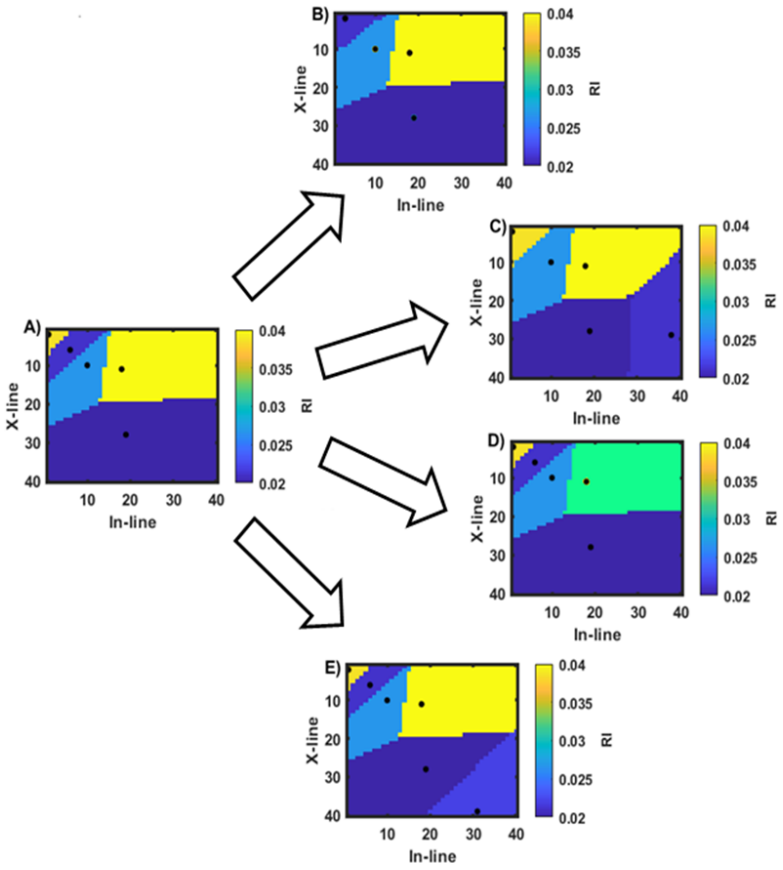


*Figure 57: Proposed perturbation strategies adopted: A) starting model with 5 nuclei, B) perturbed model after death step, C) perturbed model after cell move, D) perturbed model after elastic property change, E) perturbed model after birth step.*

Figure (58) illustrates an example of evolution of the $R_I$ models sampled by the algorithm. These models represent the $R_I$ values estimated over the 2D stratigraphic horizon. The algorithm starts from $R_I$ values very different from the reference model and from a Voronoi tessellation with a number of cells that is not sufficient to successfully predict the observed data. As the iterations proceed, the sampled $R_I$ values get closer to the reference model, and the Voronoi tessellation successfully partitions the considered 2D horizon by gathering within the same polygon CDPs with similar AVOs responses. In other terms, the algorithm successfully recognizes adjacent CDPs positions with similar AVOs responses and similar relative contrasts in the P- and S-impedance values.

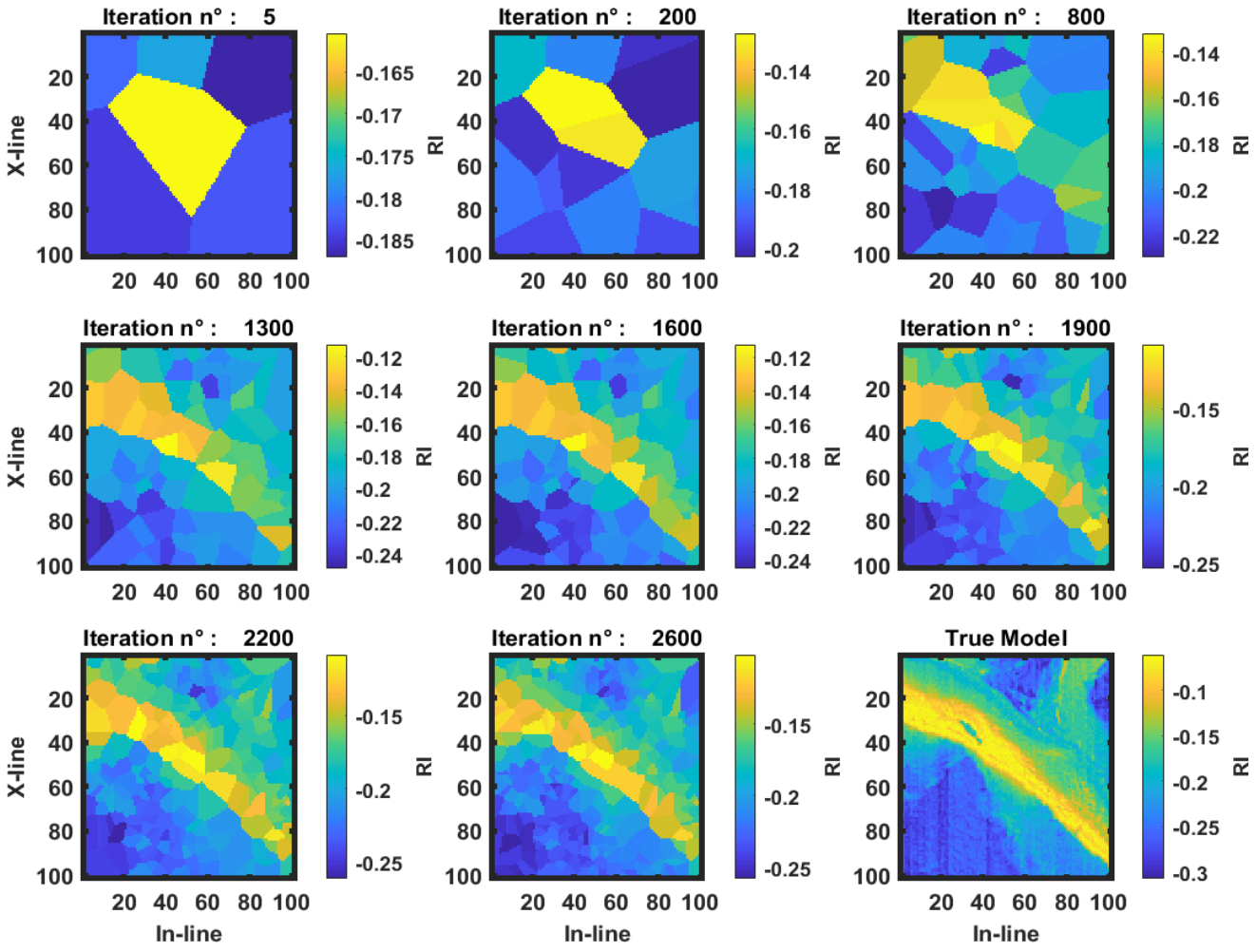


*Figure 58: Example of evolution of the sampled RI models from iteration 5 (top left) to iteration 2600 (bottom center) over a regular spaced grid of 100x100 CDPs. The reference model is shown on the bottom right panel.*

## 4.3 Synthetic Inversion Tests And Results

The previously described Rj-McMC is compared with a classical Bayesian linear approach. This is a standard, fast marching method with good reliability where the model $\mathbf{m}$ is obtained as:

$$\mathbf{m} = \left(\mathbf{C_d}^{-1}\,(G(\mathbf{m})'\,G(\mathbf{m})) + \mathbf{C_m}^{-1}\right)^{-1} G'\,\left(\mathbf{d} - \left(G(\mathbf{m_{prior}})\right)\right)\mathbf{C_d}^{-1}. \qquad (95)$$

with $\mathbf{C_d}$ covariance matrix of data noise, $\mathbf{C_m}$ model covariance matrix, $G$ represents the forward model operator (Equation 2), $\mathbf{d}$ is the observed seismic AVO and $\mathbf{m_{prior}}$ indicates the prior model.

The synthetic observed AVO data are computed from a reference elastic model constituted by a regularly spaced grid of 320 by 480 CDPs (leftmost part of Figure 60, 61 and 62). This reference model represents a stratigraphic section of a deltaic depositional area characterized by twisted river channel systems where yellow regions are representative of predominantly sandy channels, nested in blue-colored background shale portions with significant lateral contrasts in the $R_I$ and $R_J$ values.

The Ursenbach and Stewart equation is applied to the reference model to compute the AVO response for each CDP gather position within an angle range between 0 and 45 [°] and a 40 [Hz] Ricker source wavelet with a sampling interval of 2 [ms]. We consider three scenarios with different S/N ratios and with Gaussian-distributed noise affecting the data; in the first, second and last scenario the noise standard deviation is respectively equal to 0.001, 0.05 and 0.07. On each scenario we use 20 chains each one running for $10^5$ iterations. Figure (59) indicates that after the first 3,000 iterations, approximately, the MCMC algorithm reaches the stationary regime after which the algorithm begins to sample the posterior PDF. The Vp/Vs ratio oscillates from 1.9 to 2.2.

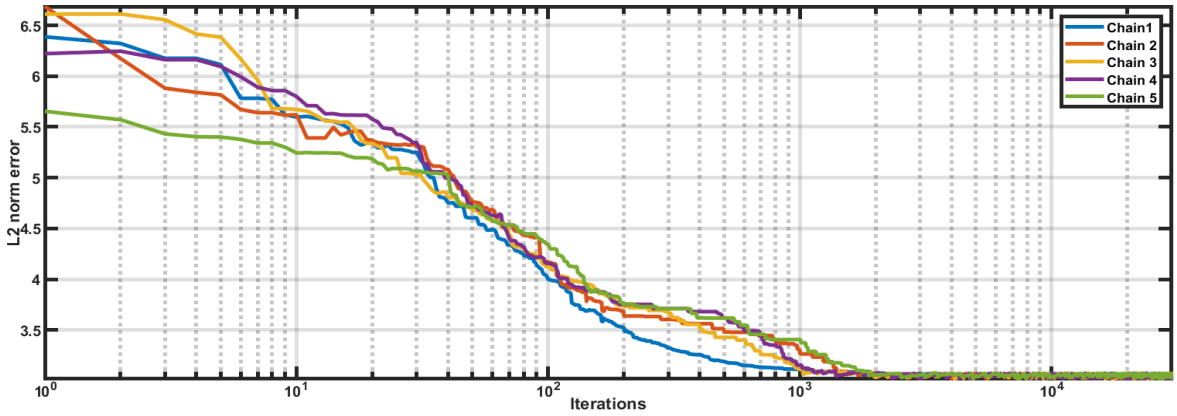


*Figure 59: Example of evolution of the L2 norm misfit for 5 chains of 20.*

In order to significantly reduce the computational cost of the inversion procedure on the simulated 2D model for the Rj-McMC, we split the inversion area in smaller areas of 40 x 40 CDPs. The standard Bayesian approach has been instead applied separately on each CDP.

Figure (60) shows that for a S/N ratio of 70 dB both approaches used provide final MAP solutions in good agreement with the reference model and where the formation boundaries can be mapped with high accuracy. Differently, with increased noise, the standard Bayesian algorithm without lateral constraints provides an estimated model characterized by significant scattering that is produced by the noise propagation from the data to the model space (Figure 60 and 61). On the contrary, the implemented Rj-McMC algorithm efficiently attenuates the ill-conditioning of the inversion procedure: from the one hand, averaging the AVO response within the same Voronoi cell significantly increases the S/N ratio of the observed data. From the other hand, averaging the elastic properties estimated for the CDPs position falling within the same cell, inherently introduces lateral constraints into the inversion framework, that can be particularly highlighted where the lateral contrasts are stronger (as in *xline* 150-250, *inline* 150-300 or *xline* 50-150, *inline* 350-450 of Figures 60, 61 and 62). Both these characteristics of the Rj-McMC algorithm ensure a more stable inversion procedure and more reliable results.

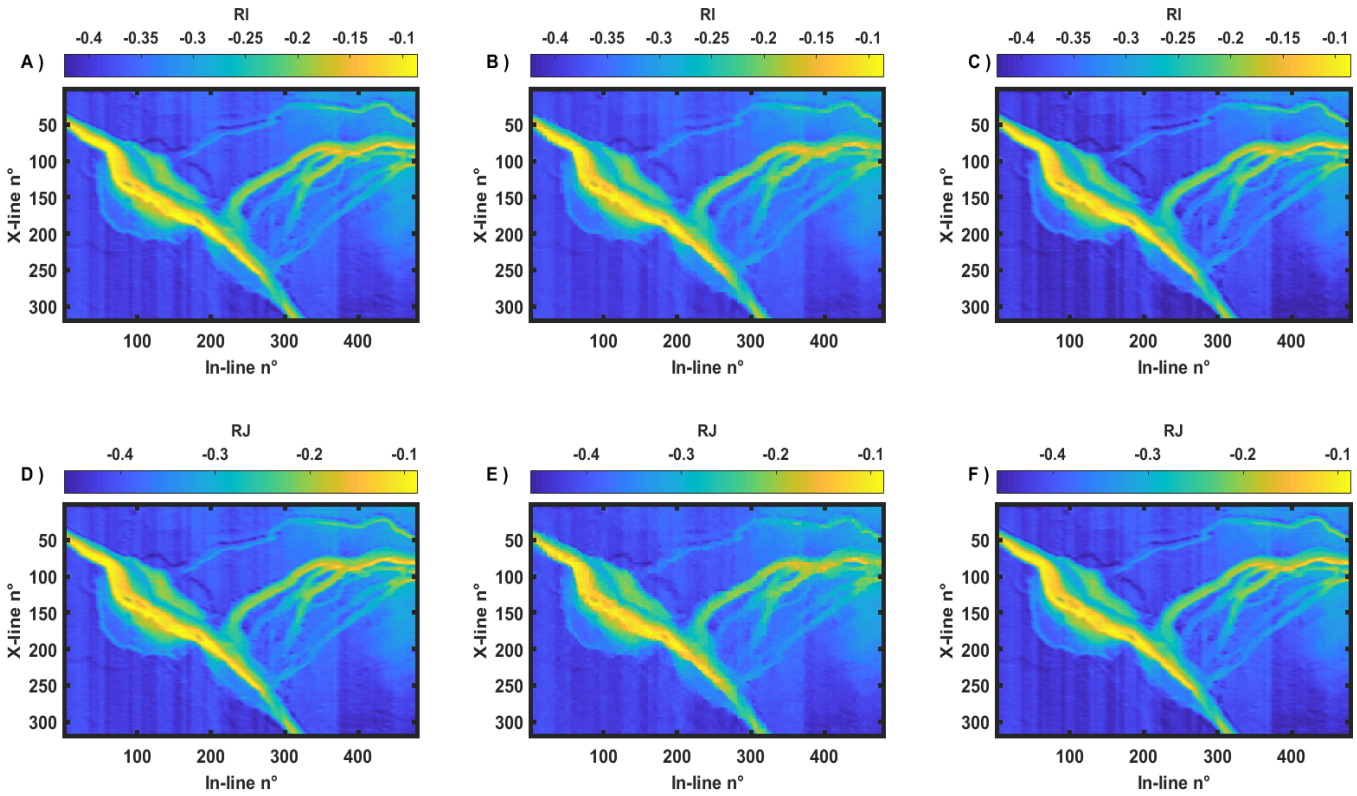


*Figure 60: Comparison between the reference model (A and D), the solution provided by the Rj-McMC (B and E)and by the standard Bayesian approach (C and F)for noise standard deviation is equal to 0.001. RI and RJ represent the relative contrasts in the Ip and Is values at the reflecting interface, respectively.*

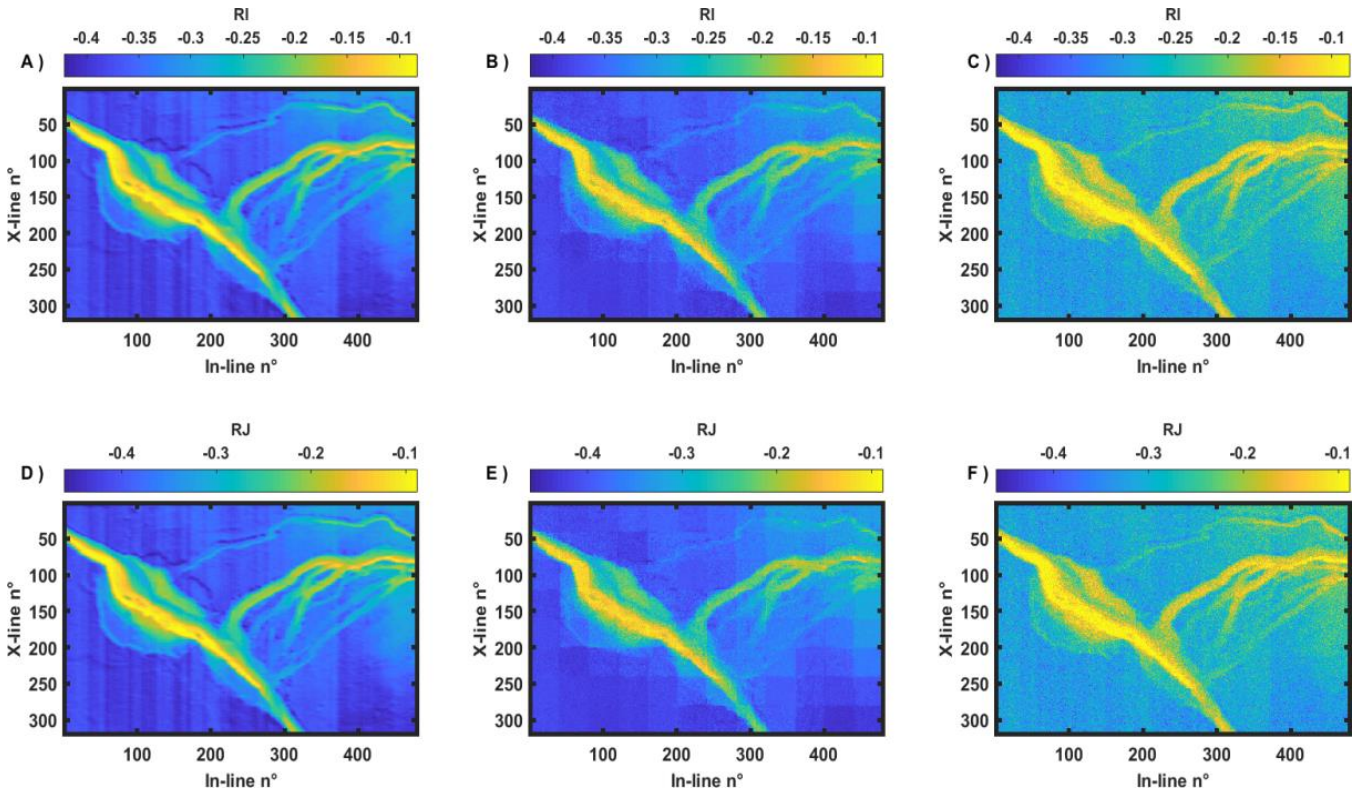


*Figure 61: Comparison between the reference model (A and D), the solution provided by the Rj-McMC (B and E)and by the standard Bayesian approach (C and F)for noise standard deviation is equal to 0.05. RI and RJ represent the relative contrasts in the Ip and Is values at the reflecting interface, respectively.*

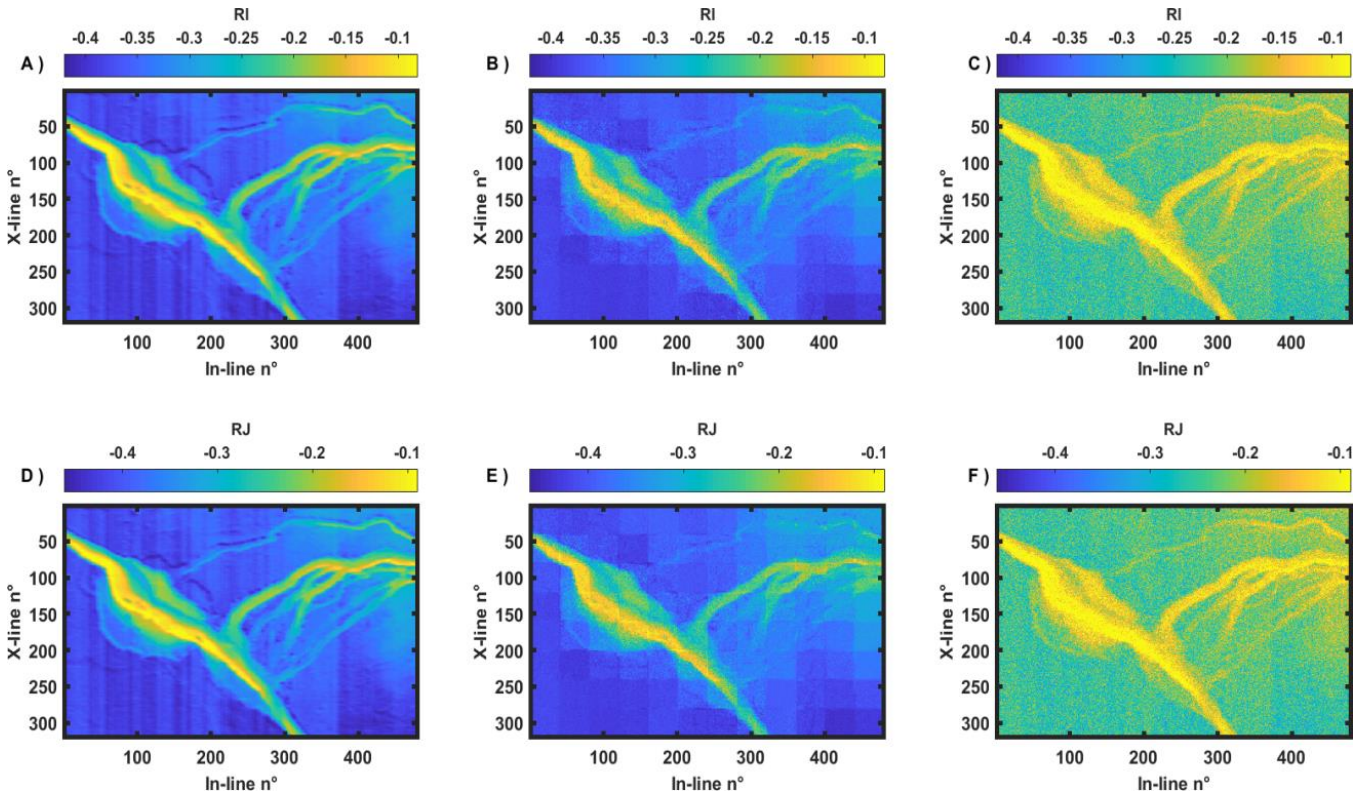


*Figure 62: Comparison between the reference model (A and D), the solution provided by the Rj-McMC (B and E)and by the standard Bayesian approach (C and F)for noise standard deviation is equal to 0.07. RI and RJ represent the relative contrasts in the Ip and Is values at the reflecting interface, respectively.*

In order to give a quantitative indicator to compare the results obtained, we calculate the correlation coefficient (*RR*) between the Rj-McMC and the standard Bayesian approach for the poorly- and highly-contaminated noise scenarios, as reported in Table 4.

| ***RR* for noise standard deviation of 0.001** | $\boldsymbol{R_I}$ | $\boldsymbol{R_J}$ |
|---|---|---|
| *Rj-McMC* | 0.984 | 0.977 |
| *standard Bayesian* | 0.999 | 0.999 |
| ***RR* for noise standard deviation of 0.07** | $\boldsymbol{R_I}$ | $\boldsymbol{R_J}$ |
| *Rj-McMC* | 0.949 | 0.949 |
| *standard Bayesian* | 0.777 | 0.776 |

*Table 4: Correlation coefficients derived from the posterior PDFs estimated by the Rj-McMC and the standard Bayesian approach.*

Results obtained confirm what previously qualitatively highlighted, that for high S/N ratio both approaches show a very good match with the synthetic data. The slightly lower match observed with the Rj-McMC for both the $R_I$ and $R_J$ parameters can be explained considering that in these synthetic cases, the $n_{max}$ value has been fixed to three-quarters of the maximum number of CDPs within the inverted area. By doing so the algorithm generates models where naturally one or more CDPs shared the same elastic properties. This yields to a natural lower correlation with the synthetic data respect to the standard Bayesian approach, where single CDPs were inverted. In the lower S/N ratio case, where the ill-conditioning is severe, the differences between the outcomes of the two approaches can be clearly appreciated, in which the better match obtained with the Rj-McMC can be quantitatively calculated about 17 % more respect the standard Bayesian method.

The implemented inversion algorithm runs in approximately 12 hours on two compute nodes equipped with two deca-core Intel(R) Core(TM) i7-8550U @2.00 GHz (8 Gb RAM). Clearly the computational cost is primarily related to the number of chains used, the maximum number of Voronoi cells the algorithm is allowed to create, the maximum number of iterations, the seismic

data size (i.e the considered range of angles of incidence over which the AVO is calculated) and the starting model adopted for the inversion. We observe that doubling the numbers of unknowns (i.e bigger inversion areas or increased incidence angles) lead to an exponential increase of the computational cost; for which a further optimized code should be applied.

## 4.4 Conclusions

We implemented a Reversible jump-Markov chain Monte Carlo algorithm for target-oriented AVO inversion on 2D synthetic seismic data. Our aim is to estimate the elastic variables and their lateral variability over a reference model characterized by strong lateral contrasts. Results are benchmarked against a standard Bayesian linear approach for different S/N ratios.

Since the convergence of the McMC algorithms is generally very slow, a Parallel Tempering approach is used to ensure optimal exploitation and exploration capabilities and to preserve the mutual correlation between the elastic properties.

Our synthetic tests show that for very high S/N ratio the standard Bayesian approach fails to reconstruct the actual subsurface structures and produce a final prediction totally covered by noise. Differently, the proposed Rj-McMC ensures much more stable results, in which the lateral discontinuities are recovered accurately.

The reliable solution provided by the proposed Rj-McMC method is guaranteed by its transdimensional framework that inherently adapts the parametrization of the proposed models to the lateral variability of the data. Indeed, the implemented algorithms do not require any statistical test to choose the adequate model parameterization and do not include any regularization operator to force the model to honour some external constraints. Instead, the Rj-McMC automatically adjusts the underlying model parametrization to produce solutions with appropriate level of complexity to fit the data to statistically meaningful levels.

# CHAPTER 5

# A Hamiltonian Monte Carlo approach for AVO inversion.

In Chapter 5, we propose an advanced numerical algorithm that incorporates the principles of Hamiltonian dynamics into the standard McMC method. This approach speeds up the convergence of McMC sampling towards the stationary regime exploiting the derivative information of the likelihood function. An additional, outstanding benefit of this method is that it produces largely independent models that result in more accurate posterior estimations and in a significant reduction of the computational effort. To reduce the ill-conditioning of the inversion approach, we again make use of geostatistical constraints in the inversion framework.

## 5.1 Introduction

The Hamiltonian Monte Carlo approach was originally formulated to solve lattice quantum chromodynamics problems (Duane *et al.* 1987) but the theoretical background on which it is formulated and the promising results obtained immediately raised interest in the scientific community. Despite this interest, applications of the HMC to solve inverse geophysical problems can only be recently found (Fichtner *et al.* 2019 or Sen and Biswas, 2017).

The HMC approach is built on a solid and strong mathematical foundation that makes it suitable for solving high-dimensional ill-conditioned geophysical inverse problems. This approach makes use of approximated Hamiltonian dynamic simulations to describe the model as a particle that follows a given trajectory. This trajectory is defined by all accepted model perturbations and is uniquely determined by three parameters: the potential energy, the mass matrix and the kinetic energy. The

first is associated with the well-known misfit function, while the other two are physical parameters created to introduce the derivative operators. The gradient information is used here to make longer jumps in the model space, producing perturbed models dramatically independent with respect to the current.

The introduction for the gradient information of the posterior PDF into the inversion kernel, allows us to focus on the exploration of the most promising model space areas. This strategy ultimately results in a smaller number of iterations needed to attain the stationary regime and a decreased sampled model correlation.

Analogous to the approach proposed in Chapter 2, to reduce the null-space and derive physically plausible solutions, geostatistical constraints in the form of a variogram model are introduced into the AVO inversion framework.

We therefore present a Hamiltonian Monte Carlo approach to solve target-oriented and interval-oriented Bayesian AVO inversion to derive the elastic variables of P-wave velocity, S-wave velocity and density from pre-stack seismic data. To assess the algorithm performances, it is first been applied to synthetic data and then to field data.

## 5.2 The adopted Method

A Hamiltonian description of the model trajectory can uniquely be obtained using the momentum and position vectors, on the "phase space" definition domain.
In particular, the particle trajectory is derived through the potential energy (*U*), described as the negative logarithm operator of the misfit function:

$$U(\mathbf{m}) = -\ln\big(p(\mathbf{m}|\mathbf{d})\big) \tag{96}$$

The kinetic energy (*K*) is instead defined on the basis of the auxiliary unphysical introduced momentum variable **p**, defined as a multivariate normal distribution independent of the potential energy:

$$K(\mathbf{p}) = N(\mathbf{p}; 0, \mathbf{M}) \tag{97}$$

where **M** is the so-called mass matrix, an extremely important parameter that ensures stable and fast algorithm convergence (Fichtner *et al.* 2019).

The joint Hamiltonian density $H(\mathbf{p}, \mathbf{m})$ that describes the total energy of the moving particle is then defined as:

$$H(\mathbf{m}, \mathbf{p}) = U(\mathbf{m}) + K(\mathbf{p}) \tag{98}$$

Next, the joint system constructed for the current model **m** and momentum **p** evolves via Hamilton's equations with respect to an artificially introduced time variable $\tau$.

$$\frac{dm_i}{d\tau} = \frac{\partial K}{dp_i}, where\ i = 1,2,3, \dots, Q, \tag{99}$$

$$\frac{dp_i}{d\tau} = -\frac{\partial U}{dm_i}, where\ i = 1,2,3, \dots, Q, \tag{100}$$

Note how the right term of equation (100) is the partial derivative of the misfit function with respect to the model **m**, which allows introducing the gradient information into the inversion McMC sampling procedure.

For the current model **m**, the adopted HMC procedure is as follows:

1. Random realizations are generated from the Gaussian distribution $N(\mathbf{p};0,\mathbf{M})$ to derive the momenta $p_i$.
2. The Hamilton two-state differential equations (99) and (100) are solved to derive the proposed perturbed model $\mathbf{m}(\tau)$ and momentum $\mathbf{p}(\tau)$ through a numerical integrator algorithm (leapfrog method).
3. The proposed model and momentum are accepted or rejected through the Metropolis-Hastings algorithm with a probability $\alpha$:

$$\alpha = min\left[1, \frac{\exp\left(-H\left(\mathbf{p}(\tau), \mathbf{m}(\tau)\right)\right)}{\exp\left(-H(\mathbf{p}, \mathbf{m})\right)}\right] \tag{101}$$

If the perturbations are not accepted, we set $\mathbf{p} = \mathbf{p}$ and $\mathbf{m} = \mathbf{m}$; otherwise, we set $\mathbf{p} = \mathbf{p}(\tau)$ and $\mathbf{m} = \mathbf{m}(\tau)$ and the proposed model and momentum are used as starting variables in the next iteration.

4. The process returns to step 1 and reiterates until a previously fixed number of iterations is reached.

Note that the HMC updating strategy involves all model parameters being perturbed simultaneously, producing more efficient long-distance jumps in the model space and dramatically decreasing the proposed model correlation and increasing the acceptance level.

For both the target-oriented and interval-oriented inversions, the prior model can be compactly written as:

$$p(\mathbf{m}) = N\left(\mathbf{m}_{\boldsymbol{prior}}, \mathbf{C}_{\mathrm{m}}\right) \tag{102}$$

with $\mathbf{m}_{\boldsymbol{prior}}$ as the prior model mean vector and $\mathbf{C}_{\boldsymbol{m}}$ as the prior model covariance matrix, both are extracted from available well log data. To introduce geostatistical constraints to further increase the algorithm performances, we again make use of the variogram model in the inversion kernel for both inversion approaches under the assumption of Gaussian-distributed variables. Following this strategy, the covariance matrix $\mathbf{C}_{\boldsymbol{m}}$ is again obtained by a Kroeneker product between the stationary covariance matrix, expressing the covariance of the elastic properties, and a first-order exponential function (Buland and Omre 2003).

The potential energy ($U$) is:

$$\begin{aligned} U(\mathbf{m}) &= -ln\left(p(\mathbf{m}|\mathbf{d})\right) \\ &= \frac{1}{2}\left[\left(\mathbf{d} - G(\mathbf{m})\right)^T \mathbf{C}_{\mathrm{d}}^{-1}\left(\mathbf{d} - G(\mathbf{m})\right) + \left(\mathbf{m} - \mathbf{m}_{prior}\right)^T \mathbf{C}_{\mathrm{m}}^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right)\right] \end{aligned} \tag{103}$$

the relative derivative with respect to model **m** (equation 100) is computed as:

$$\frac{\partial U(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{J}\, \mathbf{C}_{\mathrm{d}}^{-1}\left(\mathbf{d} - G(\mathbf{m})\right) + \mathbf{C}_{\mathrm{m}}^{-1}\left(\mathbf{m} - \mathbf{m}_{prior}\right) \tag{104}$$

where $G$ represents the non-linear Zoeppritz equations forward operator, **d** is the observed seismic data, $\mathbf{C}_d$ is the data covariance matrix and **J** is the Jacobian matrix computed with a forward finite difference approach (Aster *et al.* 2018). Through this, the partial derivative of the *i*-th data point with respect to the *k*-th model parameter is:

$$\frac{\partial d_i}{\partial m_k} = \frac{d_i(m_k + h_k) - d_i(m_k)}{h_k} \tag{105}$$

Note that this forward finite difference method has a second-order accuracy; to increase it, a central difference method may be employed at the expense of greater computational effort.

Since we use of Zoeppritz equations as analytical forward model, the computational cost of a finite-difference approach is affordable; it may become unfeasible in cases of non-analytical operators. To this end, and to promote a more general approach that indiscriminately could be used, we proposed to replace the Jacobian operator with a new matrix derived by the Aki-Richards's approximation of Zoeppritz equations.

The leapfrog numerical integrator algorithm we adopt to solve equations (99) and (100) iteratively updates the momentum and position variables. Among all proposed algorithms, it indeed yields the best performance and implementation feasibility. This algorithm initially updates the momentum variable of the $\delta/2$ time sample, subsequently updates the position variable of the $\delta$ time samples, and finally updates the momentum variable of another $\delta/2$ time sample. This updating strategy is repeated for *L* time steps, after which the algorithm defines the new perturbed position.

$$p_i(\tau + \delta/2) = p_i(\tau) - \frac{\delta \partial U}{2 \partial m_i} | \tau' \tag{106.1}$$

$$m_i(\tau + \delta) = m_i(\tau) + \delta \frac{\delta K}{\partial p_i} \Big| \tau + \frac{\delta}{2} \tag{106.2}$$

$$p_i(\tau + \delta) = p_i\left(\tau + \frac{\delta}{2}\right) - \frac{\delta \partial U}{2 \partial m_i} | \tau + \delta' \tag{106.3}$$

For theoretical insights into the numerical integrators the interested reader is referred to Neal (2011) and Betancourt (2017).

As previously stated the mass matrix is a crucial component of this approach. Indeed, it controls the algorithm convergence rate and the "phase space" traverse. According to Betancourt (2017), "it decorrelates the target distribution. A proper setting of this matrix maximizes the exploration of independent models and prevents the exploration of similar model space regions". From the analyses performed, we deduce that no general formulation exists for the mass matrix, but typically, it is a scalar-multiplied identity matrix. After some trial-and-error configurations, we decided to adopt the Fichtner *et al.* (2019) model.
Following this description, the mass matrix is derived as the inverse of the posterior covariance matrix, thus introducing information about the curvature of the misfit function around the current model:

$$\mathbf{M} = \mathbf{J}^{\mathrm{T}}\mathbf{C}_{\mathrm{d}}^{-1}\mathbf{J} + \mathbf{C}_{\mathrm{m}}^{-1} \tag{107}$$

After the initial application of this proposed mass matrix, we reformulate it to strongly reduce the computational efforts needed to derive it, jointly introducing a negligible error in the inversion process (see next paragraph).

A stable HMC convergence process also depends on the choice of optimal $L$ and $\delta$. Indeed, these parameters control the sampling trajectory in the model space. Too short trajectory jumps will produce a perturbed state too close to the current model, thus slowing the HMC sampling; on the other hand, too long jumps may wrongly lead the algorithm to previously-visited states, thus wasting computational time. To partially overcome this problem, we adopt the Mackenzie (1989) proposed strategy, where the $L$ parameter value is not previously fixed but for each algorithm iteration is randomly generated from a uniform distribution.

The major differences between the HMC and standard McMC procedures can therefore briefly be summarized as:

- Distances between successive generated points with HMC are typically large, so we need fewer iterations to obtain representative sampling (no random walk), and we achieve higher acceptance ratios.
- The "price" of a single HMC iteration is higher since we need to compute derivatives, but the HMC still requires fewer iterations to reach a stable PDF estimation.

In addition to previous cases, to further test the algorithm applicability, we introduce another source of uncertainty in the target-oriented inversion process. We now assume that the upper layer properties, so far always noted, are here inaccurately derived. The forward operator is therefore considered inaccurate, and this source of error needs to be correctly addressed and propagated to the PDF (see Menke 2018). We adopt a Monte Carlo simulation, which proceeds as follows:

1. On the basis of available prior information, the mean and covariance matrix associated with shale facies are derived (we recall that we always assume a shale-sand configuration).
2. The mean value is perturbed by adding a random vector drawn from the shale-associated covariance matrix.
3. The current elastic perturbed properties are converted into the AVO (i.e., the forward model is applied), holding the lower layer properties fixed.
4. Steps 2 and 3 are repeated for *n* times, and all seismic responses are collected, each generated by slightly different elastic properties.
5. $\mathbf{C}_{\mathrm{u}}$, the covariance of all the collected AVO responses, is computed.

Hence, the differences between the *n* simulated seismic data quantify the fluctuations in the seismic response related to uncertainty in the upper layer properties. The data covariance matrix $\mathbf{C}_{\mathrm{d}}$ is therefore the sum of $\mathbf{C}_{\mathrm{u}}$ and the matrix containing the random noise affecting the observed AVO data.

We here again make use of Gaussian variogram models to include geostatistical constraints to decrease the ill-conditioning of the proposed approaches. In both the interval oriented and the target oriented approach the lateral constrains are again deduced by available well log data and prior geological knowledge of the investigated area.

The low frequency elastic model, needed to attain stable results in the interval oriented approach, is again derived by interpolation of all available well log data.

## 5.3 Avoiding The Numerical Computation Of The Jacobian

As previously stated, the Jacobian matrix is a powerful and essential tool for the HMC approach; nevertheless, it must be calculated for each iteration, requiring high computational effort. The

approach we propose to avoid the numerical computation of the Jacobian matrix can be widely applied for the interval-oriented and target-oriented inversions, but for brevity, we report only the former.

We recall the Buland and Omre (2003) temporal extension of the Aki-Richards equations:

$$\begin{aligned} Rpp(t,\vartheta) &= a_{v_p}(t,\theta)\frac{\partial}{\partial t}InVp(t) + a_{v_s}(t,\theta)\frac{\partial}{\partial t}InVs(t) + \alpha_p(t,\theta)\frac{\partial}{\partial t}In\,p(t) \\ &= \frac{1}{2}(1+tan^2\theta)\frac{\partial}{\partial t}In\,Vp(t) \\ &+ 4\frac{Vs^2(t)}{Vp^2(t)}sin^2\theta\frac{\partial}{\partial t}In\,Vs(t)\ \frac{1}{2}\left(\frac{Vs^2(t)}{Vp^2(t)}sin^2\theta\right)\frac{\partial}{\partial t}In\,p(t) \end{aligned} \tag{108}$$

The convolutional model that generates the observed data **d** can be reformulated as:

$$\mathbf{G = SADe = Ge} \tag{109}$$

where the matrix **S** contains the wavelet for each incidence angle, **A** contains the coefficients $\alpha Vp(t)$, $\alpha Vs(t)$ and $\alpha\rho\ (t)$ of equation (108), **D** is the first-order derivative operator and **e** represents the logarithm of the elastic properties. The linear forward operator is therefore:

$$\mathbf{G = SAD} \tag{110}$$

If we replace the Jacobian matrix **J** with the **G** operator, we can rewrite the mass matrix as:

$$\mathbf{M = G^T C_d^{-1} G + C_m^{-1}} \tag{111}$$

According to equation (108), knowledge of the Vp/Vs ratio along the time interval to be inverted is needed to derive the numerical values forming the matrix **G** and for this reason, we propose recomputing this matrix operator for each sampled model and the associated Vp/Vs ratio (since the computational cost for computing **G** is negligible).

Note that if we replace the Jacobian with the matrix **G**, we are inherently assuming that the curvature of the misfit function is constant over the entire model space. In the synthetic interval-

oriented inversion (see next section), this approach, for example, makes it possible to reduce the computational cost of a single HMC inversion from approximately 10 to 1.5 minutes and still provided satisfactory model predictions and uncertainty quantifications. This reduction in the computational effort occurs at the expense of decreased convergence towards the stationary regime and an overall decrease in the independence of successively sampled models. In any case, using this strategy an AVO-HMC inversion becomes much less computationally expensive than any McMC run. In addition, the validity of the linear operator depends on the considered angle range and on the elastic contrasts at the interface, and for this reason, the suitability of this strategy should be evaluated case by case.

## 5.4 Synthetic Inversion

### 5.4.1 Target-Oriented Inversion

As in the previously described approaches, we start by assessing the algorithm performances on synthetic seismic data and subsequently on field data.

The true model represents a slice of a stratigraphic grid where a meandering sand delta-channel system is hosted in a shale sequence. Due to the large dimension of the simulated 2D model (400 in-line and 400 cross-line sections defined over a regular grid of 25×25 m) and for computational feasibility reasons, we split the entire model into blocks with dimensions of 50 cross-lines and 50 in-lines. This results in a total number of unknowns for each block equal to $50 \times 50 \times 3 = 7500$. The inversion is independently run for each block, and the final 2D maps of elastic properties are obtained by merging the inversion results retrieved in each block**.** For each block, we run the HMC algorithm for 50,000 iterations, and the models sampled after 100 iterations are used to numerically compute the posterior model. The angle range over which the Zoeppritz equations are applied to simulate the observed P-wave reflection coefficients for each CMP gather is (0 – 40) [°]. Gaussian random noise is added to the simulated data to obtain a S/N ratio of 5.

In Figure (63), we show the true model that simulates a complex meandering system with isolated and interconnected sand bodies surrounded by shales. The variogram model needed to define the lateral constraints is directly determined from the spatial autocorrelation of the true model, and it is

the same for the three elastic properties we invert for. Figure (63.b) shows the starting point for the HMC inversion, which is a heavily smoothed version of the true model obtained by applying a simple moving average filter to the true parameter values. Figure (63.c) shows the mean posterior solution estimated by the HMC algorithm; we can appreciate how the inversion correctly estimates the lateral variability of the three elastic properties, while the lateral constraints efficiently preserve the lateral transition boundaries between the shale and sand formations.

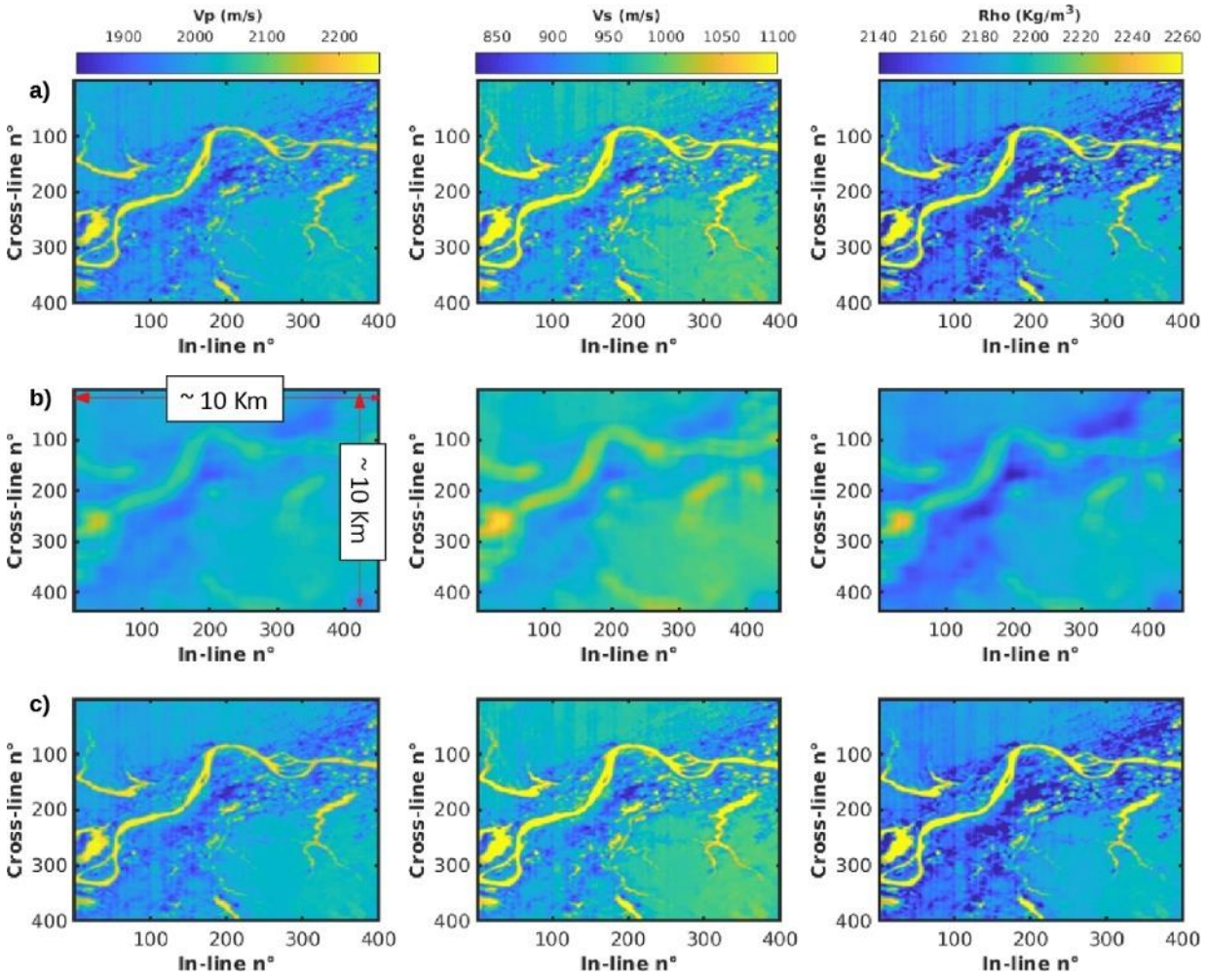


*Figure 63 : a) The true model. b) A heavily smoothed version of the true model used as the starting point for the HMC inversion. c) The estimated posterior mean. In a), b), and c), Vp, Vs, and density are represented from left to right.*

Figure (64) shows some examples of 1D marginal posterior distribution for Vp, Vs and density pertaining to different CMP gather positions. Note that the peak of the posterior is always very close to that of the true model. In this case, the posterior is very close to a Gaussian distribution since the inversion is weakly non-linear in the angle range we consider.

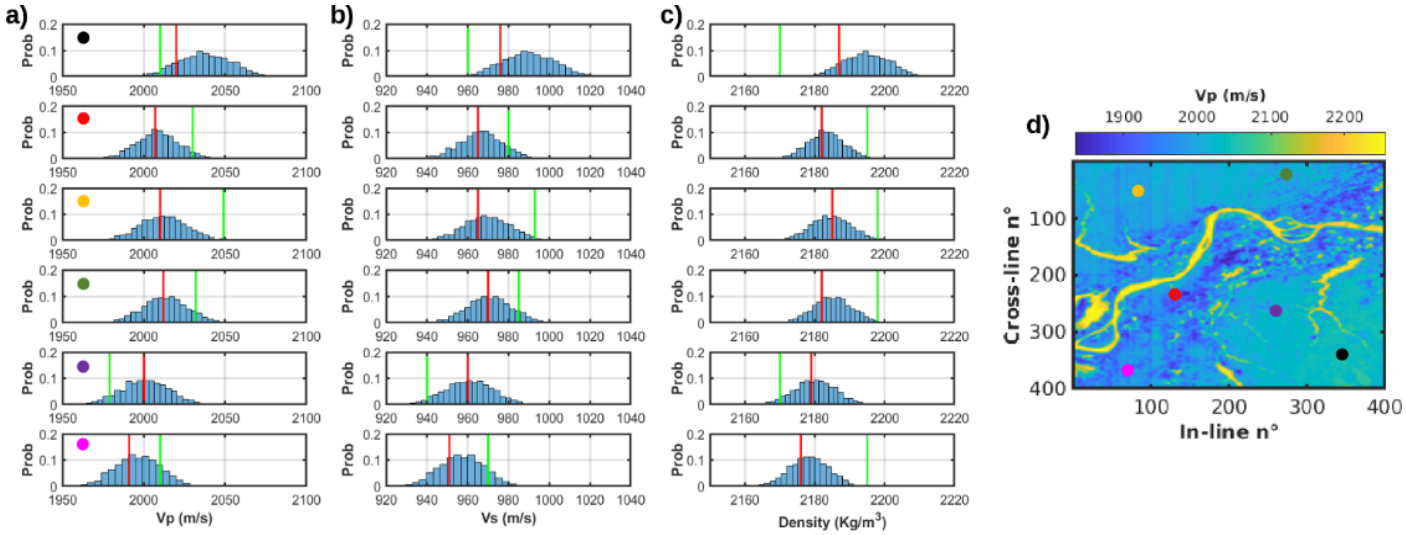


*Figure 64: Examples of 1D marginal posterior distributions for Vp (a), Vs (b), and density (c) pertaining to different CMP gather positions. The red and green lines represent the true elastic properties and the starting model, respectively. d) The CMP locations are identified by the coloured circles overlapping the estimated mean Vp model.*

Finally, Figure (65) shows that fewer than 100 iterations are required to converge to the stationary regime after which the misfit value fluctuates around a stable value.

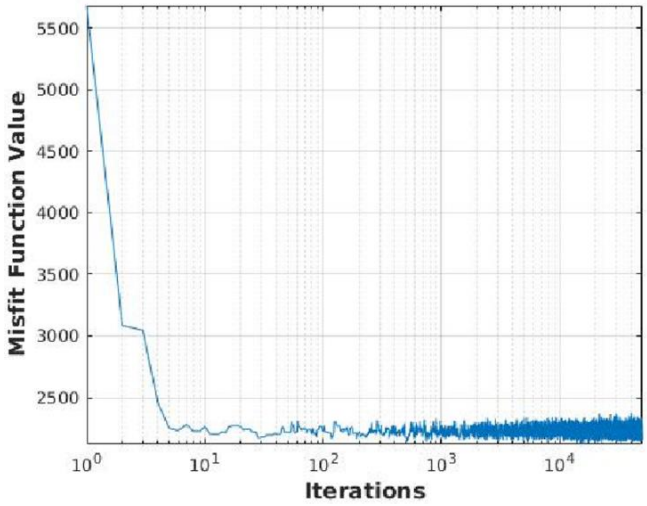


*Figure 65: Evolution of the misfit function value for the synthetic target-oriented inversion.*

We implement a parallel MATLAB code running on two deca-core Intel processor E5-2630 at 2.2 GHz (128 Gb RAM). The acceptance rate for this inversion is higher than 0.9 in the pre-burn in phase, in which the chain moves rapidly from the starting model towards a high-probability region in the model space. The acceptance rate stabilizes at approximately 0.6 in the sampling, post-burn-in stage.

The computational time to invert a single cell of 50 in-line and 50 cross-line is approximately 15 minutes.

### 5.4.2 Interval-Oriented Inversion

The synthetic seismic data are derived from an available well log investigating a productive field, while the prior information is derived from five other available neighboring well logs. The source signature is again a Ricker wavelet with a peak frequency of 50 [Hz], a sampling time of 0.001 [s] and an incidence range is (5-40) [°]. White Gaussian random noise is added to simulate the field data with S/N ratio of 5.

Figure (66) shows the results obtained with the proposed HMC approach. These results highlight the strong algorithm performances in deriving the actual vertical profiles of elastic variables. As expected, the uncertainty associated with the P-wave velocity is significantly smaller than those of the S-wave velocity and density. Furthermore, the predicted data show how the actual subsurface complexity is correctly derived with a high level of confidence by the proposed algorithm.

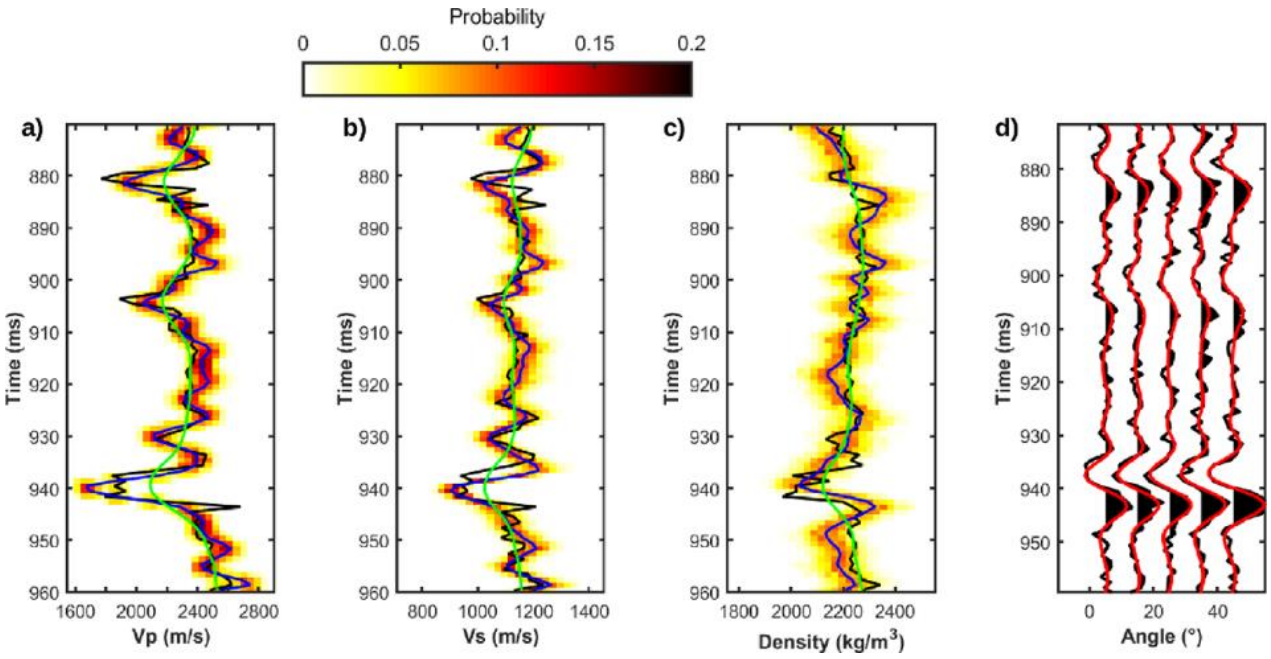


*Figure 66: Results for the interval-oriented inversion of synthetic data. In a), b) and c), the black lines represent the true property values, the blue lines are the estimated mean models, the green lines show the*

*starting model, and the colourmap codes the estimated PDF. d) Comparison of observed (black) and predicted (red) seismic data computed on the posterior mean model.*

From the evolution of the misfit illustrated in Figure (67), we note that approximately 100 iterations are needed to attain a stable value and reach the stationary regime.

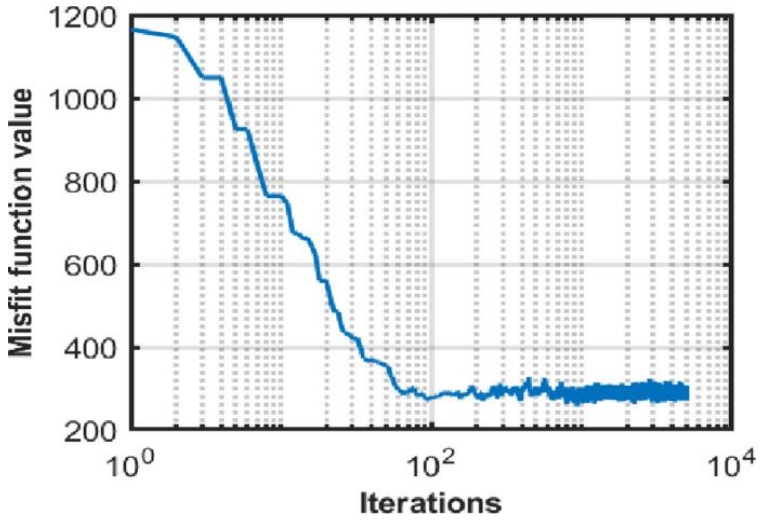


*Figure 67: Evolution of the misfit function value for the synthetic interval-oriented inversion.*

The acceptance ratio is initially very high (approximately 0.9) and subsequently decreases to 0.6 after the burn-in. The total computational time is 10 minutes running on two deca-core Intel E5-2630 processors at 2.2 GHz (128 Gb RAM).

## 5.5 Field Data Inversion

The interval oriented and target oriented approaches are applied to the same productive field as previously investigated with the transdimensional approach.

Figure (68) represents the root-mean-square (RMS) stack amplitude map extracted along a time slice representing the top reflection of the investigated reservoir interval. Low amplitudes correspond to shale, while high amplitudes identify the sand bodies. The green rectangle encloses the area considered in the target-oriented inversion, whereas the black arrow indicates the main gas sand body object of this study, where the high RMS values indicate significant elastic contrasts with

respect to the overlying layer. The lateral constraints included in the target-oriented approach are again intended to attenuate the effect of this residual noise contamination on the final solution.

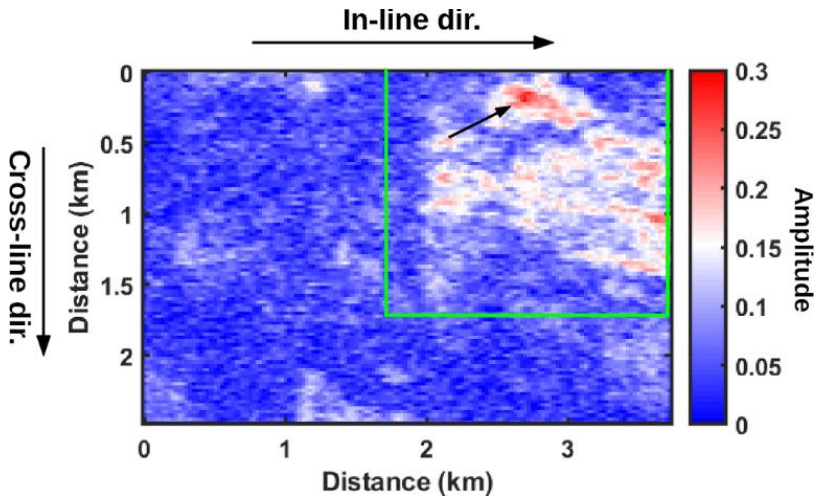


*Figure 68: RMS amplitude map at the top of the investigated reservoir. The green rectangle encloses the area considered in the target-oriented inversion. The black arrow points towards the main gas sand body.*

Analogous to the transdimensional approach, this approach derives the prior information used to define the lateral and vertical variability, the mutual correlation of elastic properties, the elastic properties of the overlying layer (for the target-oriented inversion) and the low-frequency elastic background model (for the interval-oriented inversion) from the available well log data from 7 exploration wells investigating the area.

The noise standard deviation is derived by comparing adjacent AVO responses (in the target-oriented inversion) and by comparing adjacent CMP gathers (in the interval-oriented inversion), thus again assuming that these variations are due to noise contamination (see Aleardi *et al.* 2018). The angle-dependent wavelet input for the inversion stage has been estimated through a least square truncated SVD inversion in which the reflectivity matrix is derived from available borehole information (Bianco, 2016).

### 5.5.1 Target-Oriented Inversion

For the target-oriented inversion, the AVO responses pertaining to the reflection of interest are extracted from the pre-stack Kirchhoff time-migrated CMP gathers. The proposed approach is here benchmarked with the standard Bayesian non-linear AVO inversion without lateral constraints; the interested reader can find more details in Aleardi *et al.* (2017).

Figures (69.a) and (69.b) present two of the available AVO CMP gathers where it is clearly possible to visualize the strong negative amplitude anomaly marking the shale-sand transition. Figures (69.c) and (69.d) show the relative AVO extracted responses.

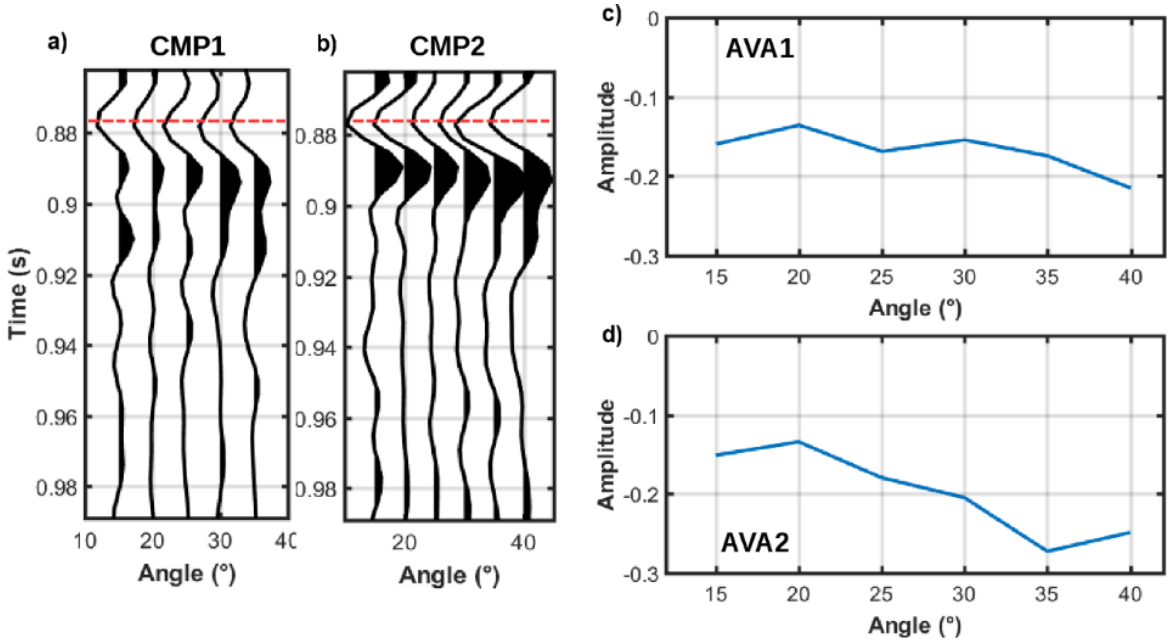


*Figure 69: a) and b) show close-ups of two pre-stack seismic gathers. The strong amplitude anomaly at approximatively 0.88 s, (red dotted lines), identifies the top reservoir reflection. c) and d) represent the AVO responses of the target reflection extracted from the CMPs shown in a) and b), respectively.*

For this approach, we involve a single HMC chain running for 70,000 iterations and a burn-in period of 100 iterations. All the models collected by this single chain after the burn-in are considered for the computation of the MAP, for the marginal distribution of the parameters and for the related standard deviation maps.

Figures (70.a-c) illustrate the starting model derived by interpolation of all available well logs and the relative estimated elastic inversion results. The lower elastic values are associated with the gas-saturated sand bodies, the higher with the encasing shales. Note the black arrows that indicate the major gas-saturated body, which is the main objective of the inversion.

Figures (70.d-f) show the HMC inversion elastic results. We can particularly appreciate how the lateral constraints included in the HMC inversion efficiently attenuate the noise propagation from the data to the model space and promote the lateral continuity of the results.

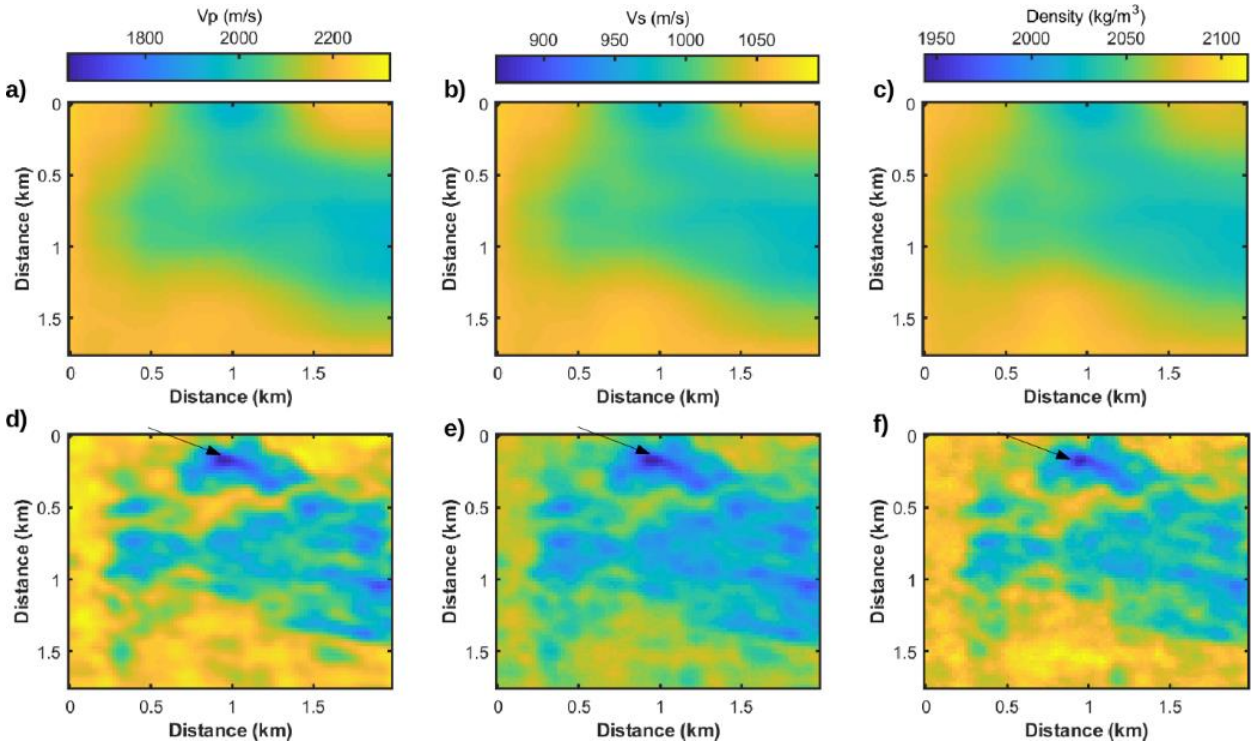


*Figure 70: a), b) and c) show the starting Vp, Vs and density models, respectively. d), e), and f) show the posterior mean Vp, Vs and density models, respectively. The black arrows highlight the very low elastic property values associated with the main gas-saturated sand body.*

In Figure (71), we represent the estimated posterior standard deviation maps for Vp, Vs and density numerically computed from the ensemble of sampled models. Note that the uncertainty for a given model parameter is inversely proportional to the curvature of the misfit function, so it depends on the likelihood function and on the prior information infused into the inversion (see equation 103 and Menke 2018). In our case, the forward modelling is non-linear, and for this reason, such curvature and thus the model uncertainty, are expected to vary over the model space.

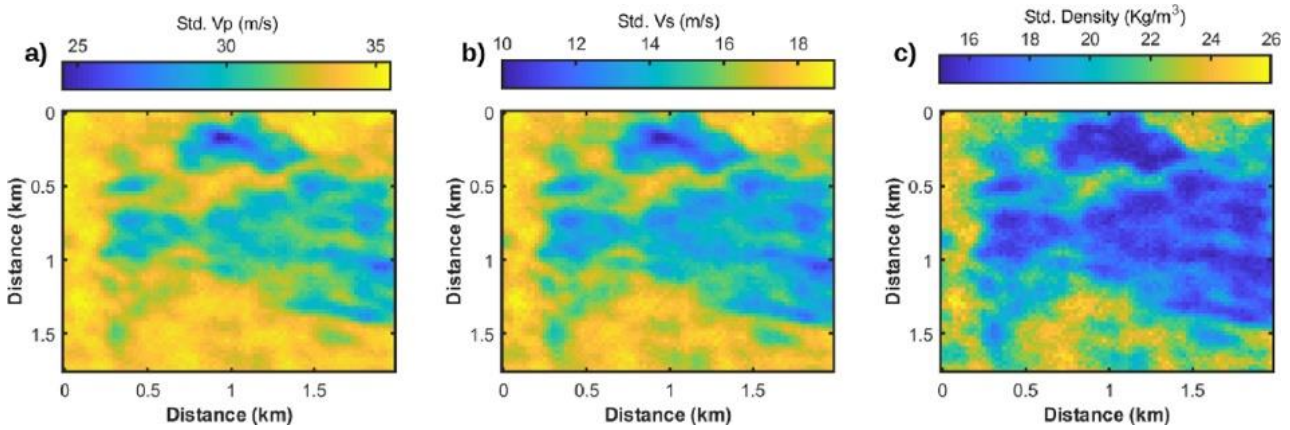


*Figure 71: Maps of estimated standard deviations for Vp (a), Vs (b), and density (c) represented for each inverted CMP position.*

Following the synthetic inversion analyses performed, Figure (72) shows 6 different CMP gathers marginal posterior PDFs. Recalling the relative effects of the elastic parameters on the AVO response, we observe an increasing uncertainty moving from Vp, Vs and density. Note also the almost Gaussian distribution associated with all the inversion results, the effect of the mildly non-linear forward model.

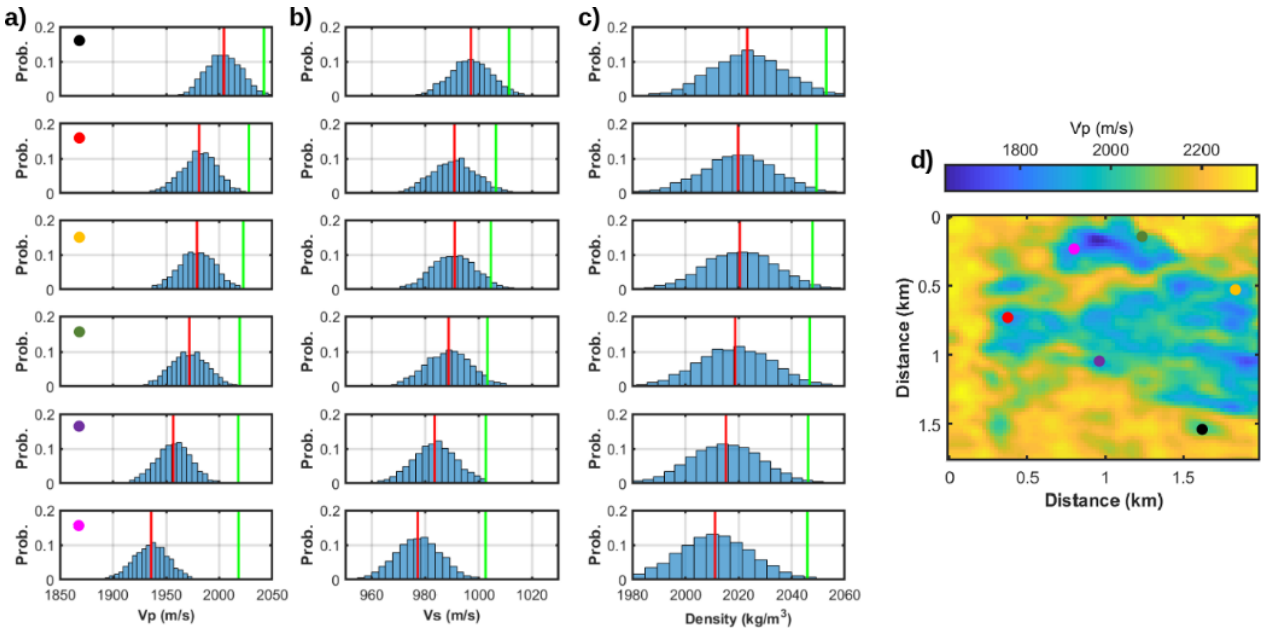


*Figure 72: a), b) and c), show some examples of 1D marginal PDFs for Vp, Vs, and density, respectively, extracted for different CMP gather positions. Red and green lines represent the posterior mean and the starting model values, respectively. d) The CMP locations are identified by the coloured circles overlapping the estimated mean Vp model.*

Figure (73) shows the normalized autocorrelation function related to 2000 sampled Vp, Vs and density models. We can here appreciate how rapidly it drops to 0 for all the elastic variables, compared with other McMC methods (compared for example to Figure 20 or Figure 33), showing how the HMC approach effectively generates highly uncorrelated models.

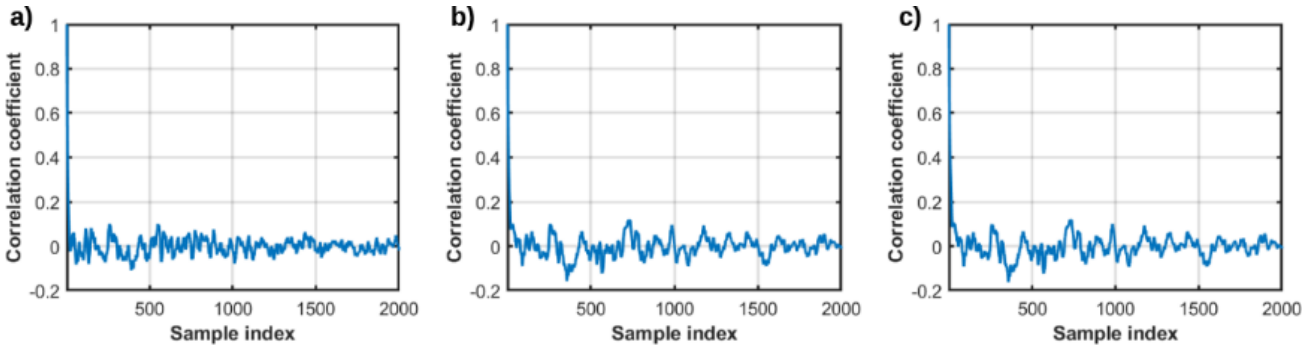


*Figure 73: Examples of normalized autocorrelation coefficients for 2.000 successively sampled models of Vp (a), Vs (b), and density (c) at a given CMP location.*

Finally, Figure (74) shows the misfit function for the proposed running chain, with a close-up view. The extremely fast convergence in this first application of field data is also favoured by the weak non-linearity of the proposed inversion approach.

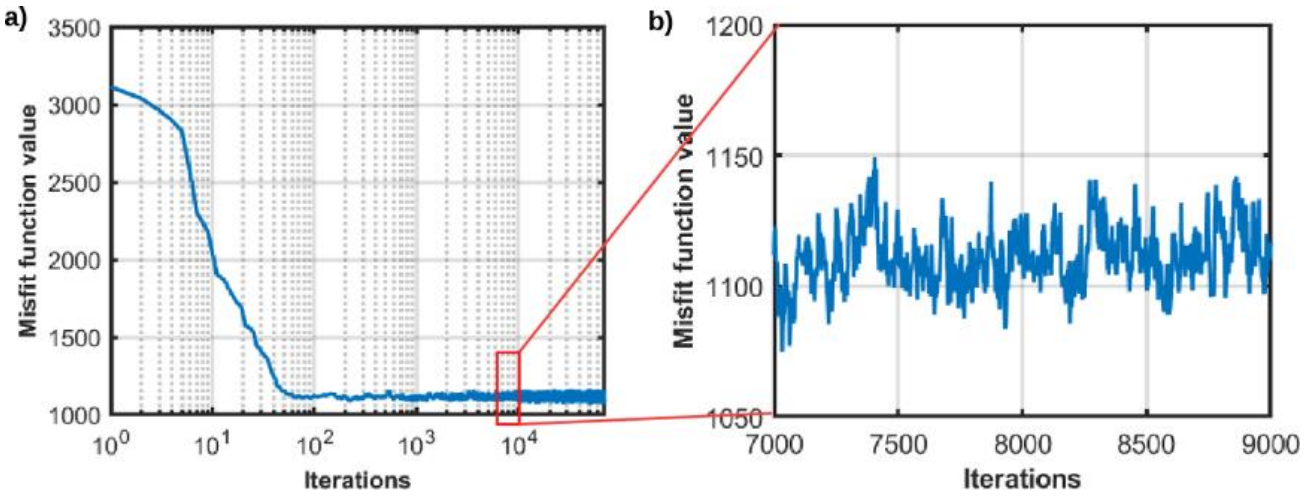


*Figure 74: a) Evolution of the misfit function during HMC sampling. b) Close-up of a) showing the highly variable misfit value versus the iteration number.*

The total computational cost is approximately 40 minutes on the same hardware configuration as that used for the synthetic tests.

### 5.5.2 Interval-Oriented Inversion

Now, we apply the 1D interval-oriented inversion along the in-line section shown in Figure (75). In this case, we employ a single HMC chain running for 5,000 iterations. The low-frequency elastic model that acts as the starting point for the HMC sampling is derived from a geostatistical interpolation of well log data (Figures 75.a-c).

Figures (75.d-f) show the estimated posterior mean values for *Vp*, *Vs*, and density. Again, significant decreases in Vp, Vs and density mark the transition between the encasing shale and the target gas sand interval. We observe that the lateral continuity of the results decreases from Vp, to Vs and is particularly low for density. This can be explained, similar to the explanation for the target-oriented inversion, by considering the relative influence played of each elastic parameter on the observed reflected amplitudes. Since the density cannot be reliably estimated with realistic noise levels (Buland and Omre, 2003), we are less confident of the estimated density model.

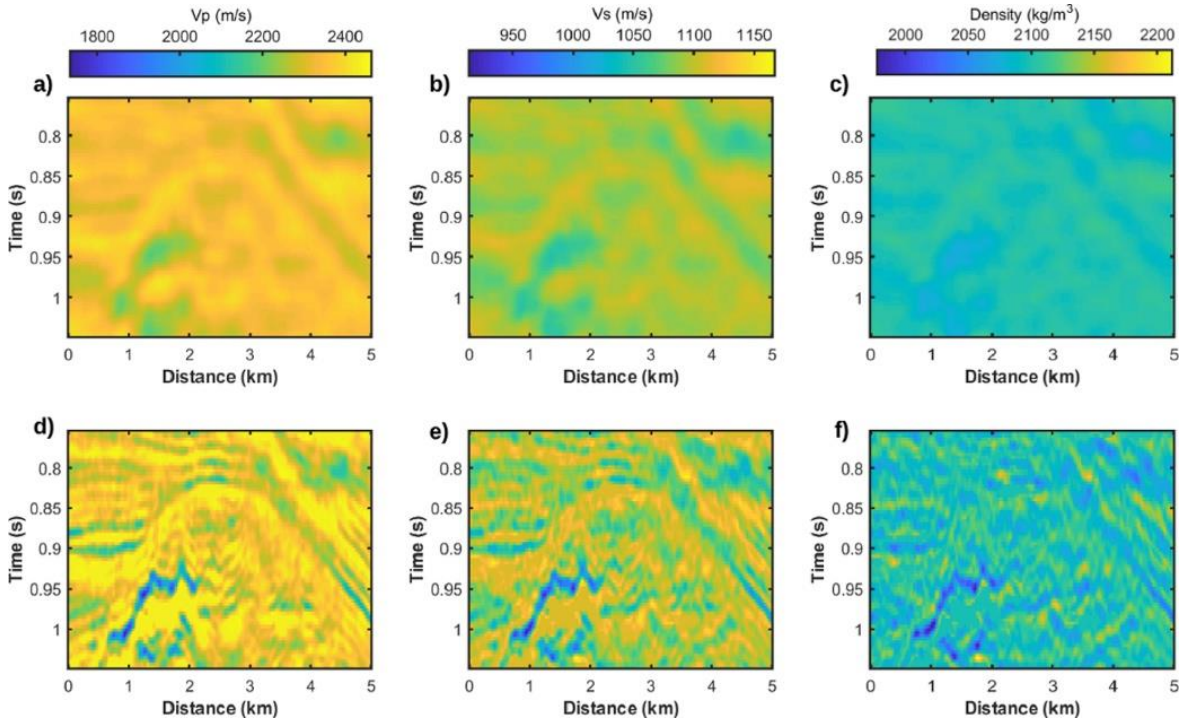


*Figure 75: a), b, and c) represent the low-frequency Vp, Vs and density models, respectively, used as starting points for the HMC inversion. d), e), and f) show the estimated posterior Vp, Vs, and density mean models, respectively.*

Figure (76) presents the 1D marginal distributions estimated for the CMP located at the horizontal coordinate of 1 km (see Figure 63). Again, we note that the uncertainties increase from *Vp*, to *Vs* and are particularly high for density.

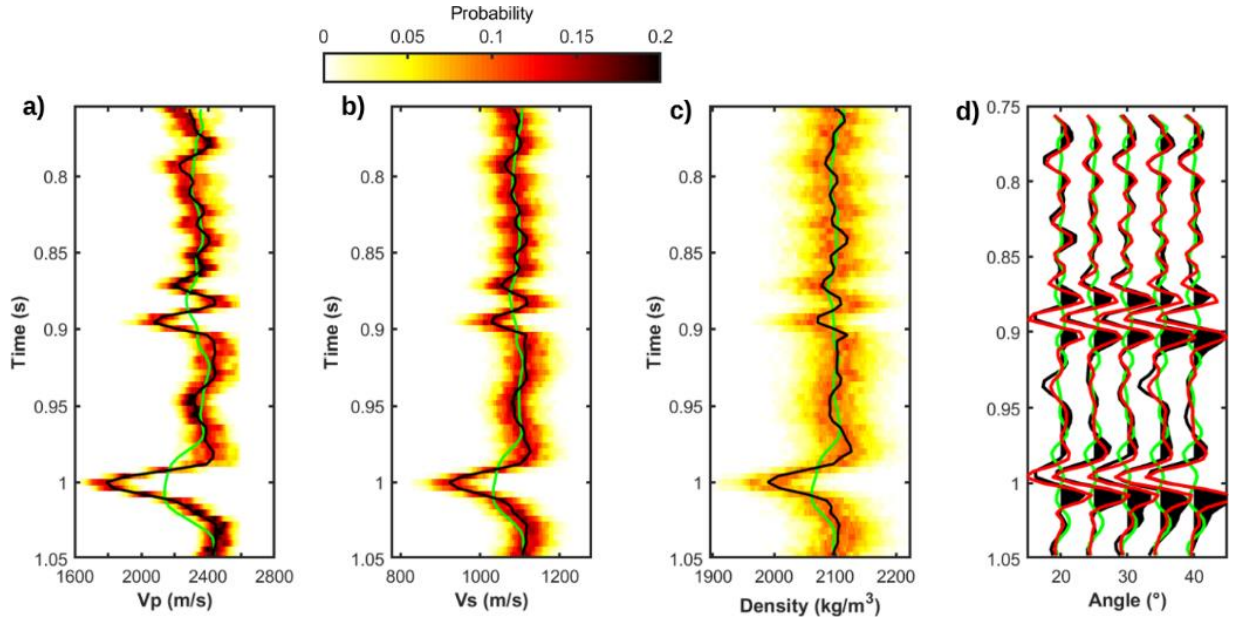


*Figure 76: a)-c) Comparisons between the estimated marginal distribution for Vp, Vs and density (colour scale), the starting models (green lines), and the posterior means (black lines). d) Comparison between the observed data (black), the predicted data computed on the posterior mean (red) and the data computed on the starting model (green). These results pertain to the CMP gather located at the horizontal coordinate of 1 km (see Figure 75).*

The estimated standard deviations along the considered in-line section are represented in Figure (77). The lateral scattering that affects the estimated standard deviation maps is probably related to the different signal-to-noise ratios of adjacent CMPs.

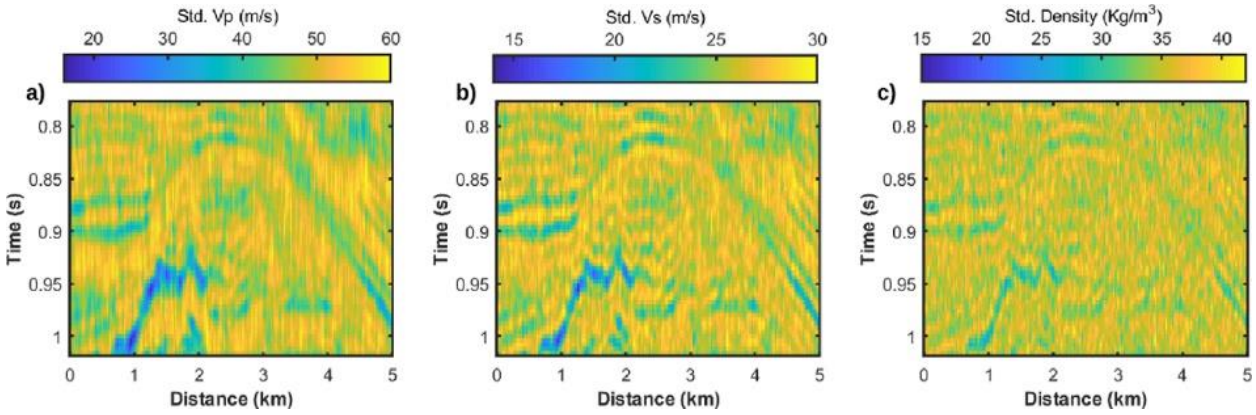


*Figure 77: Estimated standard deviations for Vp (a), Vs (b), and density (c).*

Analogous to that for the target-oriented inversion, the evolution of the misfit value for a single CMP gather inversion shows very fast convergence towards the stationary regime that is reached in 60 iterations (Figure 78.a). Figure (78.b) shows that the interval-oriented HMC algorithm is able to perform long jumps in the model space, thus producing very low correlation coefficients between successively sampled models.

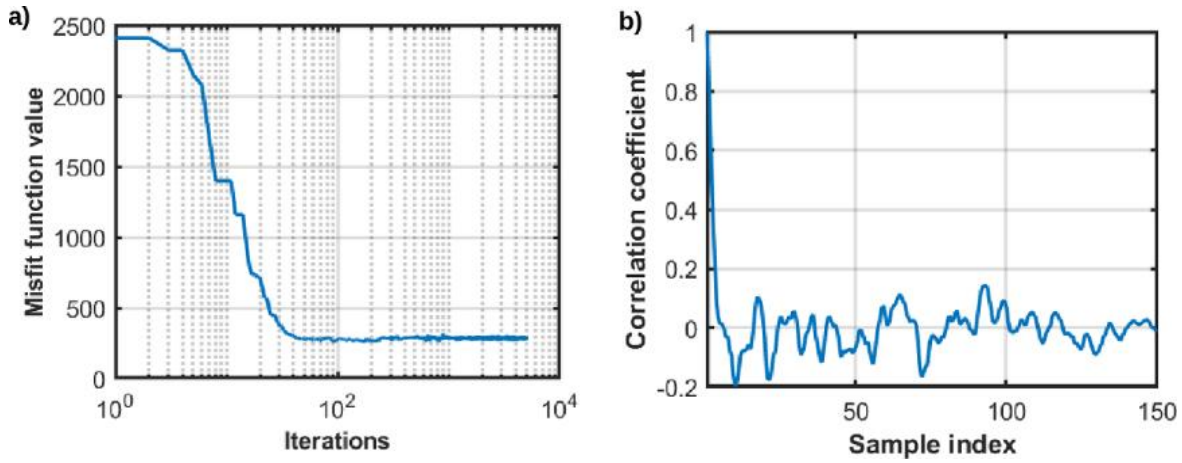


*Figure 78: a) Example of evolution of the misfit function during the HMC sampling for a given CMP position. b) Example of autocorrelation for the Vs parameter.*

Finally, we compare the mean Vp model provided by the HMC inversion and the logged Vp values measured along three wells investigating the target reservoir, as illustrated in Figure (79). The good match between the sonic log and the Vp field provided by the HMC algorithm illustrates the reliability and applicability of the implemented HMC algorithm.

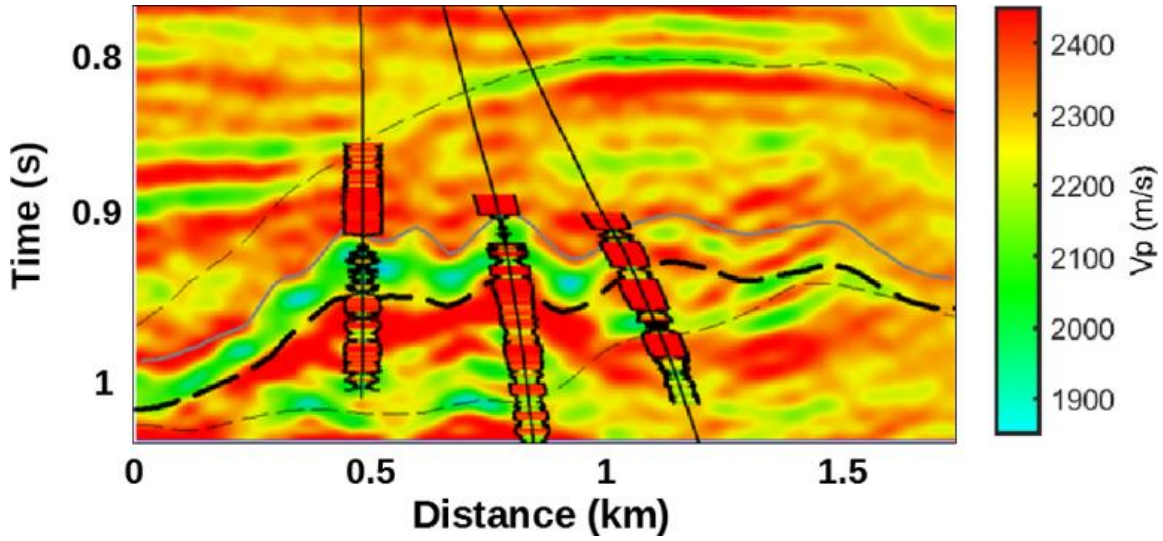


*Figure 79: A Comparison of the mean Vp solution provided by the HMC inversion and the logged Vp values recorded in three wells. The dotted and continuous black and grey lines delineate the target zone.*

The acceptance rate during the post-burn-in stage oscillates around 50-60%, demonstrating an optimal setting for the HMC user-defined parameters. The total computational cost for inverting the entire 2D section is approximately 13 hours on the same hardware setting as previously described.

## 5.6 Conclusions

The application of the proposed HMC inversion algorithm for target-oriented and interval-oriented inversions allowed the derivation of accurate posterior uncertainties for a non-linear forward modelling operator based on the exact Zoeppritz equations. Under the assumption of a Gaussian-distributed prior model, we included spatial constraints to further decrease the null space and ill-conditioning of the inverse problem at hand.

The application of the algorithm to both synthetic and field data demonstrated the applicability, reliability, and efficiency of the implemented algorithm. In particular, the HMC algorithm showed extremely fast convergence rates and retrieved accurate posterior assessments, together with reliable model parameter estimations.

The replacement of the Jacobian matrix with a matrix operator analytically derived from a linear approximation of the full Zoeppritz equations allowed us to further decrease the computational time needed to reach the stationary regime and increase the correlation value between successively

sampled models. The applicability of this reformulation should indeed be evaluated case by case because the validity of the linearization of the Zoeppritz equations depends on the angle range considered and on the elastic contrasts at the reflective interfaces.

This work demonstrates that the HMC method is a very promising approach for non-linear AVO inversion that could constitute a valid alternative to the more popular McMC strategies. For this reason, the HMC is surely one of the most promising McMC algorithms adopted.

# CHAPTER 6

# A machine learning approach for AVO inversion.

The inclusion of lateral constraints, as previously discussed in detail, is the most effective approach to reduce the null space exploration and reduce the ill-conditioning of the AVO inversion. The main limitation of all approaches described to this point is that these geostatistical constraints lead to increased dimensional matrices, which require increased computational efforts to be inverted. This obviously represents an issue limiting the introduction of potential multi-dimensional constraints into the inversion framework. We therefore propose an alternative approach to include lateral constraints, potentially multi-dimensional, into the inversion framework through a machine learning approach that combines a discrete cosine transform (DCT) with a convolutional neural network (CNN). To propagate the uncertainties of the network approximation and the noise contamination into the estimated models, we build a Monte Carlo simulation framework.

## 6.1 Introduction

As previously demonstrated, the application of Markov chain Monte Carlo approaches is strongly recommended for non-linear forward modelling operators and strongly ill-conditioned inverse problems. The major issue related to all these methods concerns the considerable computational time needed to properly sample the posterior PDF and attain stable results. Moreover, to further reduce the null space exploration and increase the algorithm performances, geostatistical constraints are often used but need a further increased computational effort for accurate uncertainty appraisals.

Over the last few years, incredible breakthroughs have been made in computer applications; extremely high-speed multi-core GPUs have been developed that greatly advanced the frontiers of scientific possibilities. In particular, we refer to the development and application of machine learning (ML), among which we are particularly interested in convolutional neural networks (CNNs); these applications compose a supervised learning task that maps a general input signal into an associated output response, extracting essential features from a training dataset and mapping them into an associated output.
The potentialities of this approach are almost infinite, spanning from image recognition to recommender systems, natural language elaboration, bioinformatics, finance and many other fields.

To be properly set, such an approach requires a large training dataset over which all the internal CNN parameters need to be updated and refined. This is a computationally extensive training, but once it is completed and the network is trained, it can easily be used to convert any input dataset into the corresponding output response in nearly real time.
The application of CNNs to geophysical problems actually spans seismic texture identification (Xiong *et al.* 2018; Waldeland *et al.* 2018), velocity estimation (Araya-Polo *et al.* 2018), impedance inversion (Das *et al.* 2019), full-waveform inversion (Lewis and Vigh 2017; Richardson, 2018), electromagnetic inversion (Puzyrev 2019) or first-break picking (Yuan *et al.* 2018).

In this work, we combine a CNN with a discrete cosine transform (DCT; Lochbühler *et al.* 2014) reparameterization to solve the AVO inverse problem, in which the subsurface elastic properties of P-wave velocity, S-wave velocity and density are inferred from partially stacked seismic data at different incidence angles.

The DCT generates orthogonal base functions to expand a signal expressing the subsurface model vector into a series of cosine functions oscillating at different frequencies. Usually, most of the variability of the original signal is expressed by the first low-order DCT coefficients, and for this reason, this mathematical transformation can be used for model and/or data compression. The DCT base functions also allow the preservation of the spatial and temporal variability of the elastic parameters in the recovered solution because the order of the retained DCT coefficients directly determines the wavelength of the recovered elastic model along the different spatial/temporal directions. Such mutual and spatial/temporal dependencies of the elastic properties are also imposed on the training and validation datasets and to this end, we employ the direct sequential co-

simulation method with joint probability distribution (Co-DSSj; Horta and Soares, 2010), which also allows us to draw random simulations from a non-parametric elastic prior model.

In a DCT-based inversion, the unknown parameters become the numerical values of the coefficients and no longer the elastic variables associated with the AVO responses. In our application, the DCT constitutes an additional feature extraction technique that reduces both the number of unknown parameters in the elastic inversion and the number of pixels in the image input and output of the CNN.

Therefore, the discrete cosine transform model reparameterization technique is used here to reduce the dimensionality of inverse problems and provides an alternative approach to introduce the lateral constraints that preserve the lateral continuity of the elastic properties.

The CNN inversion is also combined with a Monte Carlo approach to properly estimate the uncertainties affecting the retrieved solution. Indeed, note that the trained network learns an approximated function that maps the observed data into the associated model. However, this approximation introduces an additional source of uncertainty, usually called modelling error, which must be properly accounted for and propagated into the final solution. We focus on synthetic inversion experiments in which the reference (true) model mimics a real gas-saturated reservoir hosted within a turbiditic sequence. Different inversion tests that realistically simulate errors in the estimated wavelet and in the assumed noise distribution are carried out to properly assess the robustness and applicability of the implemented approach.

The predictions of the implemented CNN inversion approach are benchmarked with those yielded by the popular trace-by-trace linear inversion algorithm proposed by Buland and Omre (2003) and previously described.

## 6.2 The Adopted Method

### 6.2.1 The Discrete Cosine Transform

The DCT is a Fourier-related transform that expresses a finite signal in terms of a sum of cosine functions oscillating at different frequencies. Generally, the DCT transformation of a 1-D signal *x* of length *N* can be written as follows:

$$y(k) = \sqrt{\frac{2}{N}} \sum_{n=1}^{N} x(n) \frac{1}{\sqrt{1+\delta_{k1}}} \cos\left(\frac{\pi}{2N}(2n-1)(k-1)\right) \tag{112}$$

where $\delta_{k1}$ represents the Kroeneker delta, *y* are the *N* coefficients of the transformation that fully describe the original signal *x* in the transformed DCT space and *k* represents the order of each DCT coefficient.

This linear transformation is extremely useful since the most information of the input signal is concentrated in the first coefficients, so that $q < N$ coefficients can be used to approximate the original signal within a previously fixed level of confidence.

Figure (80) shows the DCT ability to reconstruct an input 1D signal considering different ranges of DCT coefficients. From this figure, we note (Figure 80.d) that the first coefficients contain a low-frequency trend of the signal and that further increasing the coefficients considered in the DCT allows us to reconstruct the oscillations at higher frequencies always more confidently. The DCT can be indiscriminately applied to multidimensional input signals, where the multi-dimensional DCT is given by a separable product of DCTs along each dimension.

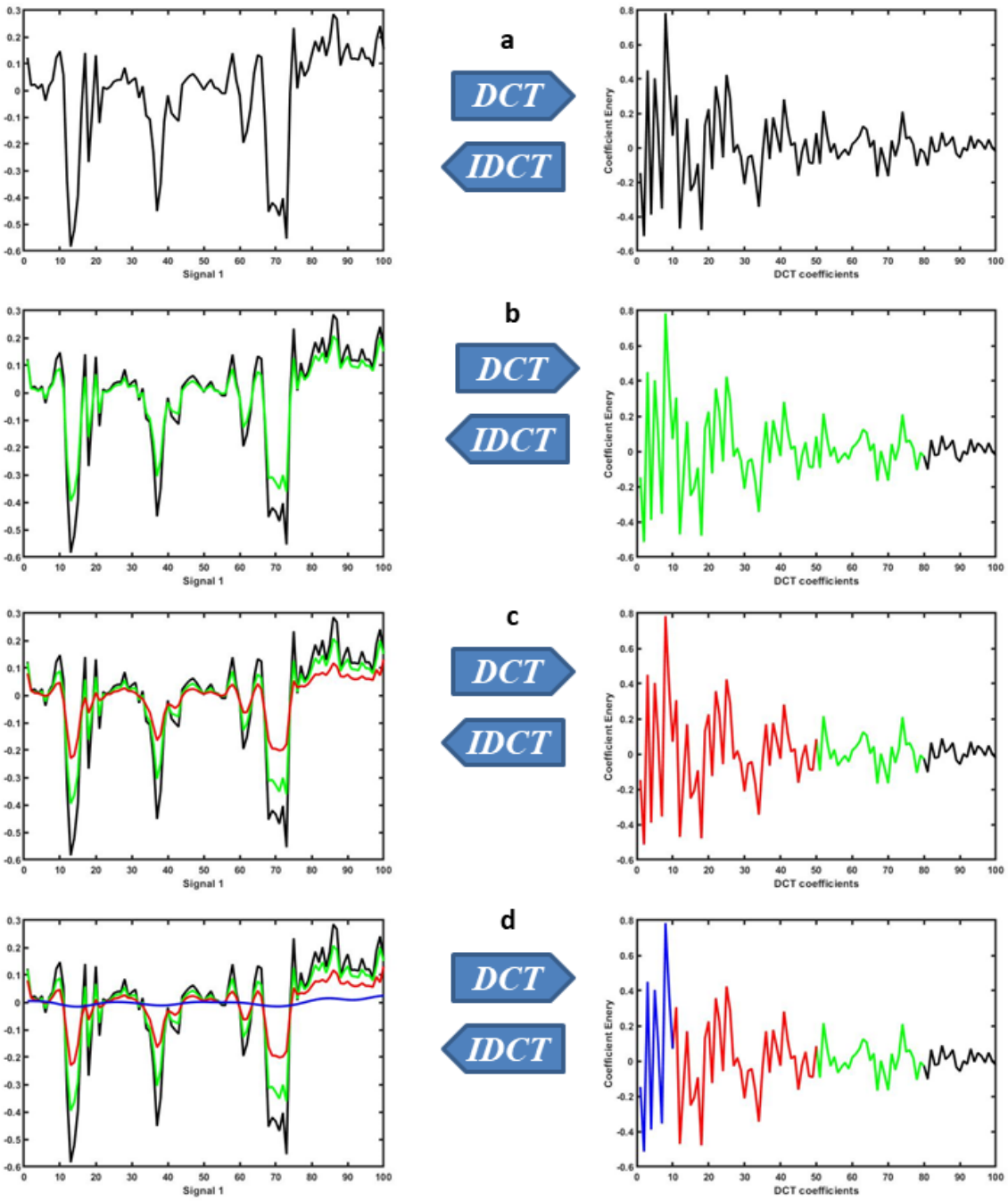


*Figure 80: The DCT and its inverse applied to a generic signal. From a) to d), we observe that the progressively lower number of coefficients used in the DCT (100, 80, 50 and 10) allows us to reconstruct the original signal always with lower resolution. Note that d) contains the low-frequency trend of the input signal.*

Using a matrix formalism, equation (112) becomes:

$$\mathbf{y} = \mathbf{B}\mathbf{x} \tag{113}$$

where the vectors **x** and **y** represent the original and the transformed signal, respectively, and **B** is an *N*-by-*N* matrix in which the *i*-th column contains the *i*-th order cosine function (base function) spanning the DCT space. An approximation of the signal **x** can be obtained by considering only the first *q* DCT base functions:

$$\tilde{\mathbf{x}} = \mathbf{B}_q^T \mathbf{y}_q \tag{114}$$

where $\tilde{\mathbf{x}}$ is the approximated signal, **B***qT* is a partition of the matrix **B** with *N* rows and *q* columns representing the first *q* DCT base functions and the vector $\mathbf{y}_q$ contains the first *q* coefficients that multiply the base functions.

In the context of geophysical inversion, this means that the numerical values of these *q* coefficients become the unknowns to be inferred from the data. The DCT has been applied both to the data and model space to reduce the dimension of the CNN input and output.

### 6.2.2 The Convolutional Neural Networks

Neural networks are computational architectures inspired by biological neuronal implants. These architectures are usually made up of interconnected and interactive components called nodes or neurons. The single nodes simulate biological neurons by receiving information and performing simple operations on them, at the end of which they pass the results to the other neurons.

Just as in biological cells where the information arrives through the dendrites, is processed in the soma and is passed through the axon, the neural networks have an input layer, a hidden layer (that can be composed of several layers) in which they are processed, and an output layer.

The inputs of a given neuron are the products of the outputs of the previous neurons associated with weights. During forward propagation, in which all the connections proceed from the input layer to the output layer, the inputs are summed and passed through an activation function before being sent

to another node. A constant bias is applied to the activation function, which shifts its position. Typical biological and mathematical neural networks are shown in Figure (81).

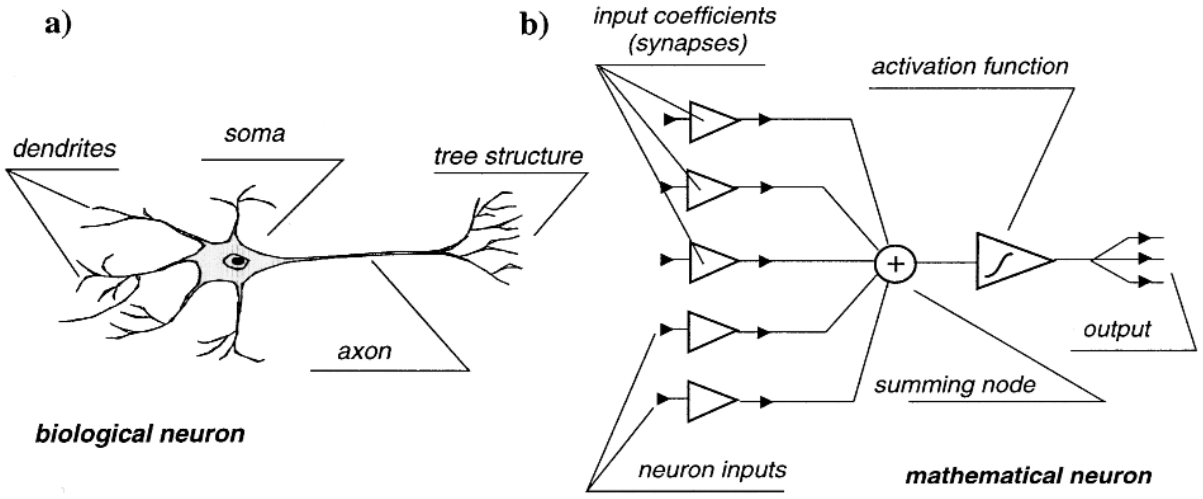


*Figure 81: a) A biological neuron and b) the corresponding mathematical neuron that approximates the biological neuron in the neural networks.*

Convolutional networks are simply neural networks that use convolution in at least one of their layers. CNNs use blocks of convolutional layers, subsampling layers, and fully connected layers to extract features from 1D, 2D, or 3D input maps treated as grids of pixels. The extracted features form the so-called feature maps.

The process can be briefly written as:

$$O_j^p = f\left(b_j + \sum_{i=1}^{I} O_i^{p-1} * W_j\right), \quad j = 1, 2, \dots, J \tag{115}$$

where $b_j$ is a scalar value representing the bias, $f()$ is the activation function used to include non-linearity in the mapping process, $O_j^p$ is the $j$-th feature map in the $p$-th layer, $O_i^{p-1}$ represents the $i$-th feature map in the ($p$-1)-th layer, and $\mathbf{w}_{i,j}$ denotes the $j$-th filter associated with the $p$-th layer, which is the weight matrix connecting $O_j^p$ with $O_i^{p-1}$. This filter has a user-specified size and slides over the input map with a specified stride. $I$ represents the number of feature maps in the ($p$-1)-th layer, whereas $J$ is the total number of feature maps in the $p$-th layer, which is equal to the number of filters considered in the layer.

The internal CNN parameters to be updated are $\mathbf{w}_i$ associated within each layer. Their optimization is an iterative procedure where an initial configuration is tested and is successively updated based on the difference between the desired and computed output. This updating process can be written as:

$$\mathbf{W}_i = \mathbf{W}_{i-1} - \gamma \frac{\partial \varepsilon}{\partial \mathbf{W}_{i-1}} \tag{116}$$

where $i$ represents the iteration number, $\varepsilon$ is the loss function value, and $\gamma$ is the so-called learning rate. Therefore we proceed by minimizing the error prediction between the net performances and the desired data.

In the architecture of a CNN several operations can be performed over the various features; among others, the two main important and most commonly used are listed below:

- Pooling layers are used to reduce the dimensions of the data by combining the outputs of neuron clusters in one layer into a single neuron in the next layer. The most common pooling strategies are the maximum pooling or the average pooling (Scherer *et al.* 2010).

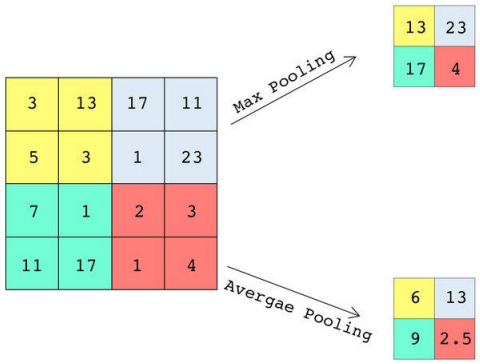


*Figure 82: The different effects of maximum pooling and averaged pooling over an input grid.*

- Fully connected layers, have every neuron in one layer fully connected to every neuron in another layer. These fully connected layers are generally appended at the end of the last convolutional block, and their outputs can be either continuous values (regression problems) or discrete values (classification problems).

To define the CNN architecture, some hyper-parameters must be set: number of hidden layers, number of filters, kernel width and stride of the convolution operators, activation function, a method for weight initialization, an optimization algorithm to minimize the loss function and to update the filter weights, number of epochs and learning rate. There are no rigid rules to set these hyper-parameters, and the final choice is often dictated by personal preference and experience. However, some very general rules exist; for example, the dimension of the convolution filters must not be greater than the dimension of the feature map to which the filter is applied. Further, more filters are used, and more features can be extracted. However, the number of features is limited and specific for the problem at hand. In contrast, a too low number of filters can produce underfitting to the observed data.

As previously stated, net training is a computationally expensive operation. The total amount of available data is divided into three different categories:

- *Training* is used for net creation, gradient calculation and weight updating.
- *Validation* is used to stop the iterations and the training process before overfitting.
- *Testing* in this case is the reference model over which the net performance is tested.

The optimal setting in our case is found through a trial-and-error procedure in which we change different hyper-parameters, such as the number of filters, their size, the learning rate, the batch size, and the type of activation function. The final architecture is determined based on the net performances on the validation dataset.

## 6.3 The Implemented CNN Inversion

We focus our efforts on the same previously defined area used in both in Chapter 3 and Chapter 5. Here, the seismic dataset refers to a geological area in which a turbiditic sequence hosts a gas-saturated reservoir (see Figure 83). This represents our reference model constituted by 51 time samples and 71 *cross-lines*.

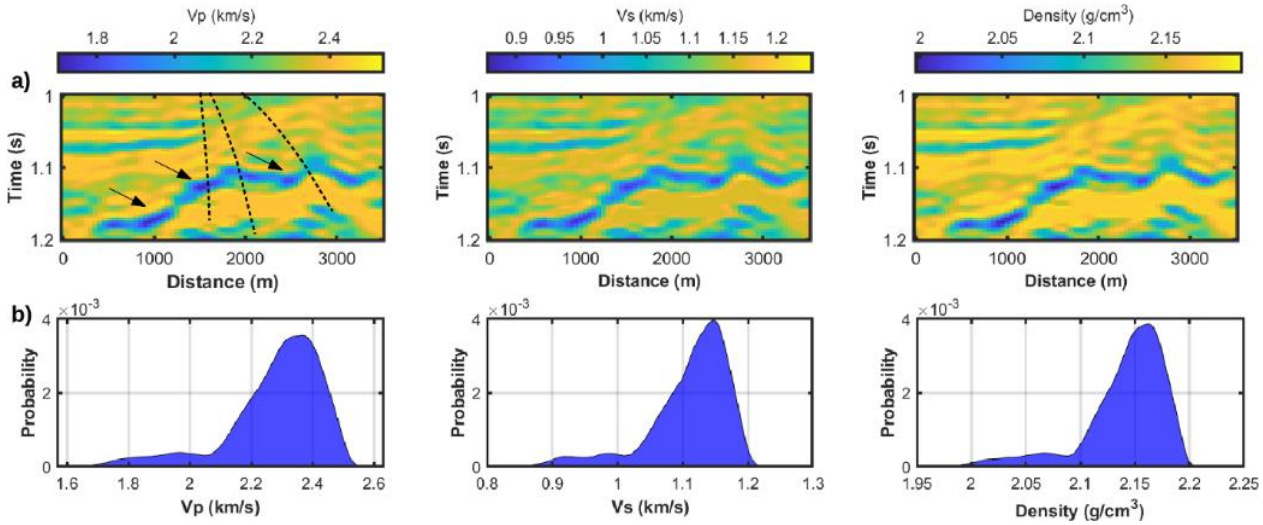


*Figure 83: a) The elastic properties of Vp, Vs, and density of the reference model. In a) the black arrows point towards the main sand reservoir body, whereas the dotted black lines depict the trajectories of 3 wells that are used to build the model and to define the prior elastic distribution. b) The marginal non-parametric prior distributions for the three elastic properties derived from the five available wells.*

The prior data is again extracted from all 7 available wells penetrating the area of interest, which we again assume to be stationary over the entire study area. The non-parametric distribution of the elastic variables is derived by the application of the kernel density estimator, as reported in Figure (83).

A stationary 2D Gaussian variogram model is again adopted, in which the vertical and lateral ranges are inferred from the vertical variability of the upscaled well log data and from the lateral variability of the observed seismic data. The variogram is again derived by the Kroeneker product between a general 2D correlation function (2D extension of equation 15) and the prior elastic covariance matrix. The ranges of the variogram are equal to 0.004 [ms] and 160 [m] along the temporal (vertical) and spatial (lateral) directions, respectively. We consider the prior information as being stationary over the entire study area, and given the structural complexity of the reference model this assumption constitutes a simplification of the actual variability of the elastic properties.

Since we need numerous models to train the CNN, 20000 2D elastic models are drawn from the prior distribution using the direct-sequential simulation method. Thee full Zoeppritz equations and a 30-Hz zero-phase Ricker wavelet are used to compute observed data that are contaminated with Gaussian uncorrelated noise with a standard deviation corresponding to 35% of the standard deviation of the noise-free dataset. We consider three partial angle stacks corresponding to incidence angles of 0 (near), 15 (mid), and 30 (far) degrees.

Using this incidence angle range each model is composed of 51 time samples, 71 cross-lines and 3 elastic parameters for a total of 10863 parameters; each related seismic dataset is composed of 50 time samples, 71 cross-lines and 3 incidence angles for a total of 10650 data points.

Our next step involves the estimation of the number of DCT coefficients needed to optimally reproduce our 20,000 elastic models and relative seismic responses. Figure (84) presents the data and the DCT-transformed model and graphically shows how we can approximate *Vp*, *Vs*, and density using the first 24 rows and columns of the associated DCT matrices, while in the seismic domain, we consider the first 30 rows and columns of the first two slices forming the DCT-transformed data cube. Therefore, the use of the DCT allows us to compress the 10863-D full elastic space into a 24 x 24 x 3 = 1728-D space tensor and the 10650-D data domain into a 30 x 30 x 2 = 1800-D space tensor.

This strong parameter-space reduction by the DCT involves translating into a less complex network architecture, hyper-parameter setting and faster training phase.

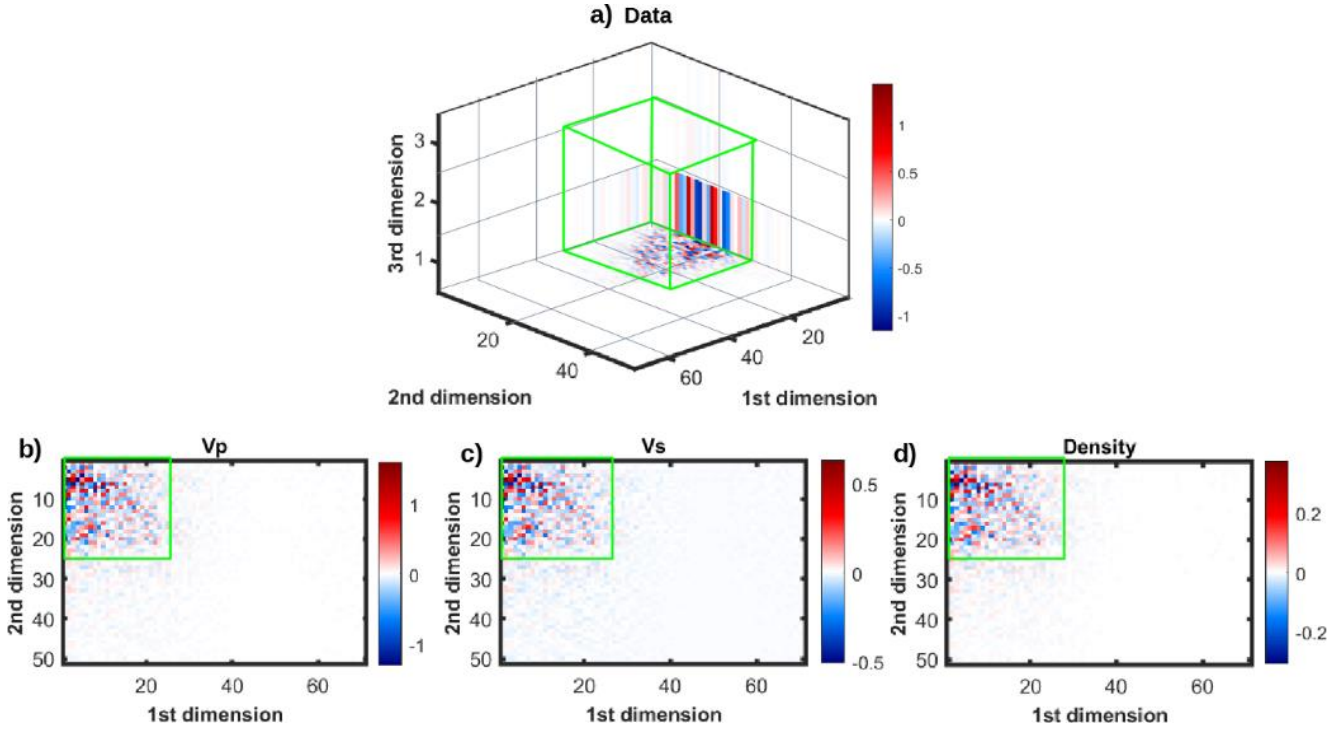


*Figure 84: Some examples showing three sections from DCT-transformed seismic data (a) and the associated DCT-transformed Vp (b), Vs (c), and density (d) models. The green boxes and green rectangles enclose the retained DCT coefficients in the data and model space, respectively. In a) the 1st, 2nd, and 3rd dimensions refer to the cross-line, time, and incidence angle directions, respectively. In b)-d), the 1st and 2nd dimensions refer to the cross-line and time axes, respectively.*

To assess a probabilistic approach to the CNNs, we are now interested in deriving some procedures to address the uncertainties affecting the estimated solutions. In particular, we must propagate in the model space both the noise affecting the seismic data and the error that is introduced by the CNN application.
Under the assumption that both these errors follow Gaussian distributions, we adopt the following Monte Carlo approach:

1. The trained CNN is used to compute the predicted elastic model $\mathbf{m}$ from the observed data vector $\mathbf{d}$.
2. Forward modeling is run to compute the noise-free $\mathbf{d}'$ data associated with $\mathbf{m}$.
3. The noise $\mathbf{n}$ extracted by a Gaussian distribution $(0, \mathbf{C}_n)$ is added, $\mathbf{d}'' = \mathbf{d}' + \mathbf{n}$. The covariance matrix $\mathbf{C}_n$ is assumed to be diagonal with noise uncorrelated along the incidence angles.
4. The trained CNN is used to compute the predicted model $\mathbf{m}''$ from $\mathbf{d}''$.
5. The $\mathbf{e}$ extracted by a Gaussian distribution $N(0, \mathbf{C}_e)$ is added, $\mathbf{m}''' = \mathbf{m}'' + \mathbf{e}$. The covariance matrix $\mathbf{C}_e$ is obtained as the covariance of the difference between the ensemble of elastic properties contained in the training dataset and the related ensemble of models predicted by the trained CNN (following Hansen and Cordua, 2017).
6. The $\mathbf{m}'''$is stored and steps from 3) to 6) are repeated for *n* Monte Carlo iterations.

Each generated vector $\mathbf{m}'''$can be considered a possible subsurface scenario in accordance with the observed data, the trained CNN, and the assumed distributions for the noise and modelling errors.

The final workflow of the implemented CNN inversion is illustrated in Figure (85).

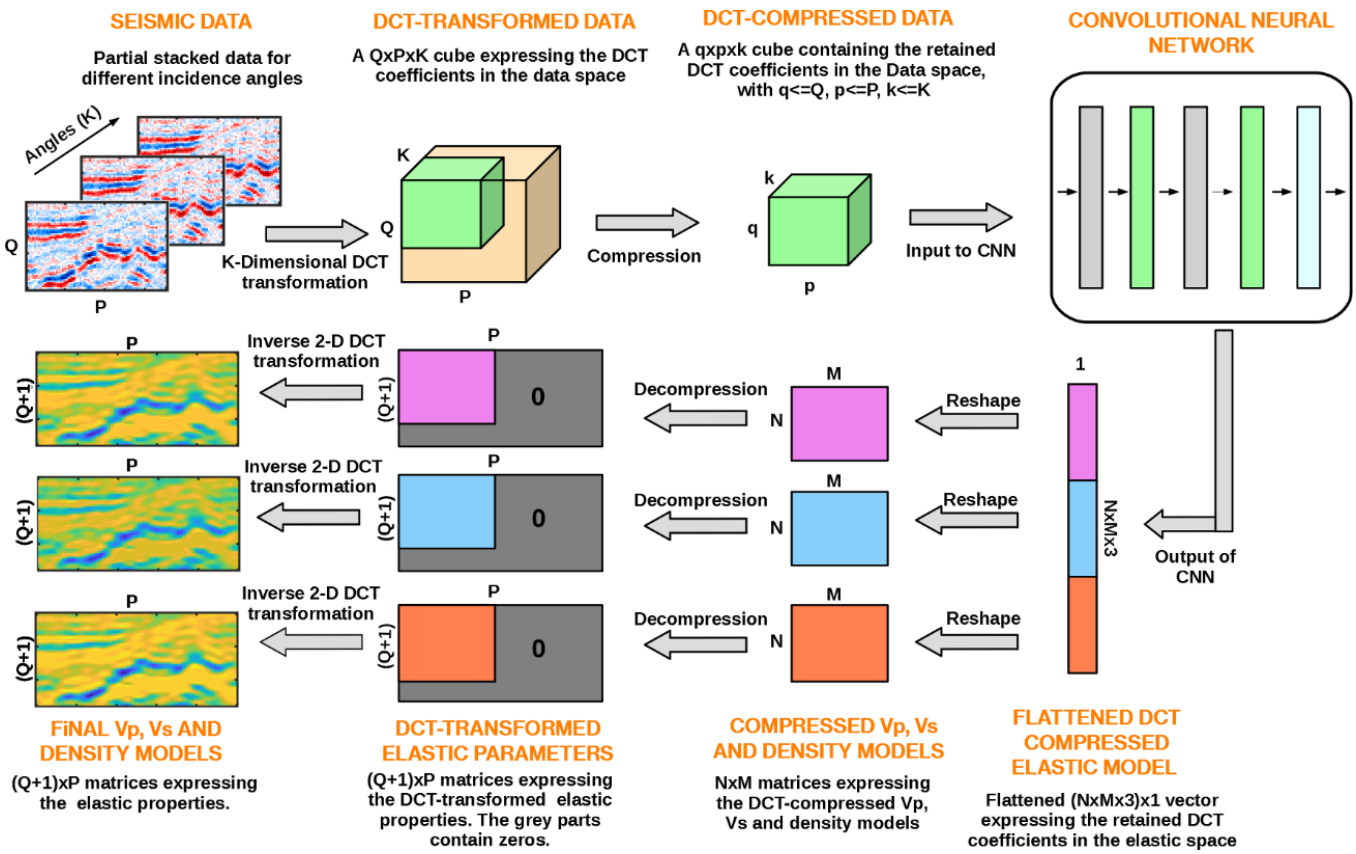


*Figure 85: Workflow of the implemented CNN inversion approach. The Monte Carlo method for uncertainty propagation is not represented in this figure.*

## 6.4 The Network Architecture

From the 20,000 available simulations extracted through the direct-sequential simulation method, we set 18,000 for the *training* and the remaining 2,000 for the *validation*.

We randomly select 18,000 out of 20,000 models and associated seismic datasets to train the CNN, while the remaining 2,000 examples constitute the validation dataset. We use the MATLAB implementation of CNN that we run on a quad-core Intel(R) Core(TM) i-7 7700HQ CPU at 2.80 GHz with 32 Gb RAM.

We perform different experiments to optimally set the main CNN hyper-parameters, and the final CNN architecture (represented in Figure 86) is chosen according to the best fit on the validation set.

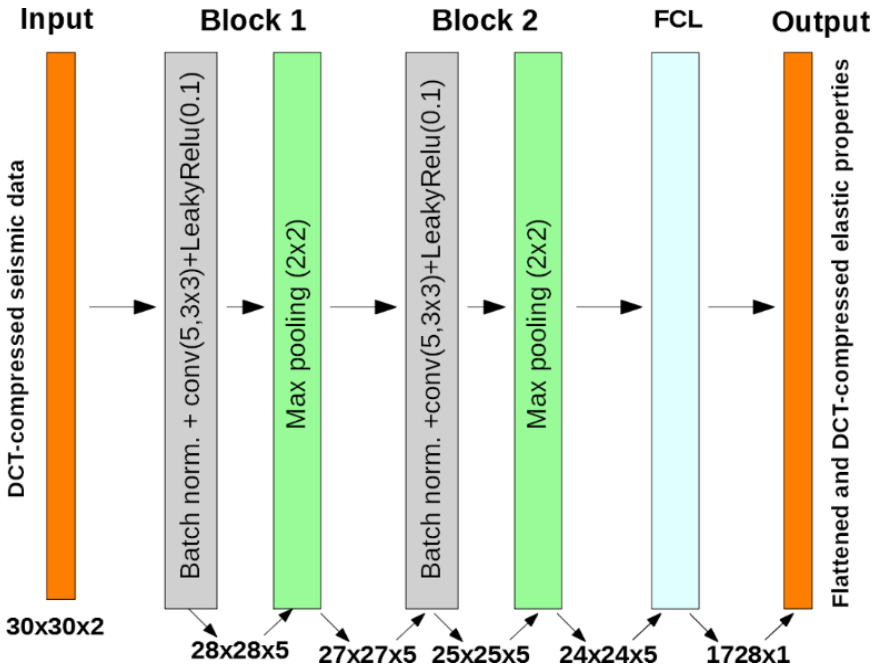


*Figure 86: Schematic representation of the adopted CNN architecture. The image on the far left is the input of the network, which is a 3D cube expressing the 1800 DCT coefficients used for data compression. The image on the far right is the output of the network, which is a vector containing the 1728 DCT coefficients expressing the elastic properties. The intervening rectangles represent convolutional layers and are annotated with key parameters. In the grey rectangles, the initial value in brackets (e.g., 5) indicates the number of filters. This is followed by the filter size (i.e., 3 × 3). The green rectangles also indicate the dimension of the maximum pooling filter (2×2). The cyan rectangle represents a fully connected layer (FCL). The numbers at the bottom of each rectangle indicates the dimensions of the input and output to each layer.*

This architecture consists of two convolution blocks and a fully connected layer. The two convolution blocks use 5 convolution filters with a size of 3 × 3 and a stride of 1. After each convolutional layer, we use the Leaky Relu activation function with a slope of 0.1 (Krizhevsky *et al*., 2012). In addition, batch normalization is used within each convolution block. After the convolution blocks, maximum pooling with a size of 2 × 2 and a stride of 1 is applied for subsampling, whereas a dropout of 0.1 is used to prevent overfitting.

We adopt the RMSprop optimizer (i.e., an unpublished, adaptive learning rate method) running for 20 epochs to minimize the root-mean-square error (RMSE) between the expected and the predicted outputs. The *He* method is used to initialize the network weights (He *et al*., 2015). We use a batch size of 64 and an initial learning rate of 0.001 that is multiplied by 0.9 every epoch. We choose this batch size and initial learning rate because they guarantee the best performances in the experiments we carried out.

From the evolution of the RMSE for the training and validation datasets, we observe that the network successfully converges after approximately 15-16 epochs (Figure 87).

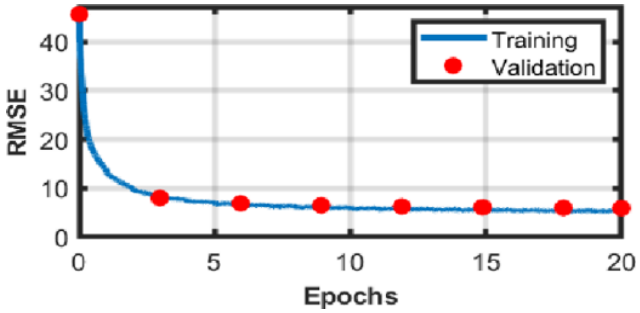


*Figure 87: Evolution of the RMSE error for the training and validation sets during the learning process.*

Figure (88) shows some comparisons between two elastic models extracted from the validation set and the corresponding CNN predictions. The quality of the results seems to demonstrate the reliability of the implemented CNN inversion. Figure (89) shows an example of the seismic data used to train the net.

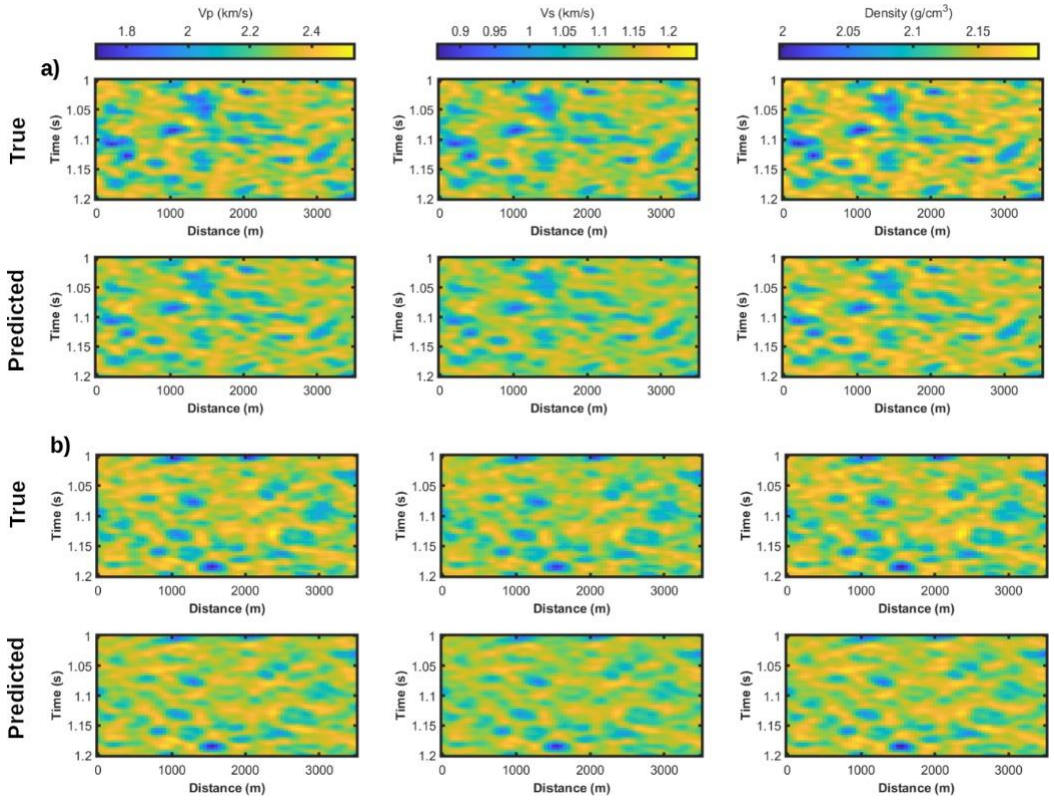


*Figure 88: a) and b) show a comparison between two elastic models extracted from the validation set and the corresponding CNN prediction. In a) and b) from left to right, we represent Vp, Vs, and density.*

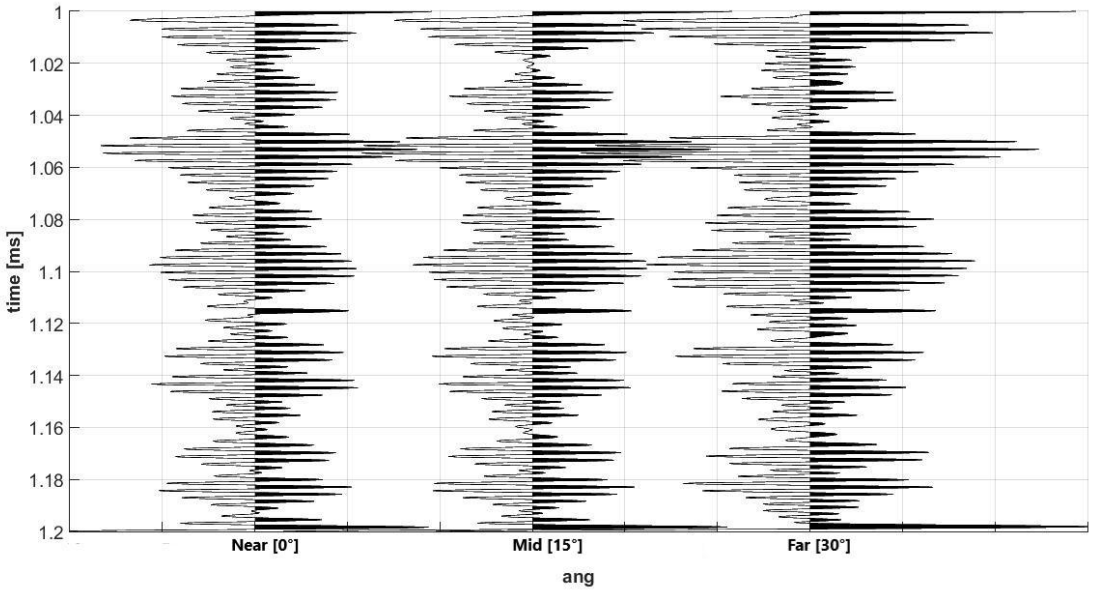


*Figure 89: Example of noise-free seismic data extracted from the training dataset used as input to the described algorithm.*

For comparison, Figure (90) represents the elastic models predicted by a trained CNN with the same architecture as previously described (Figure 86) but when the DCT has not been applied to compress the input and output responses. The network parameters, the training, and the validation dataset are the same as described previously. In other words, in this case, the CNN has been trained to predict the 10,863 model parameters from the 10,650 observations. The unphysical and scattered elastic predictions prove that a much more sophisticated network is now needed to reliably map the seismic response into the associated elastic properties.

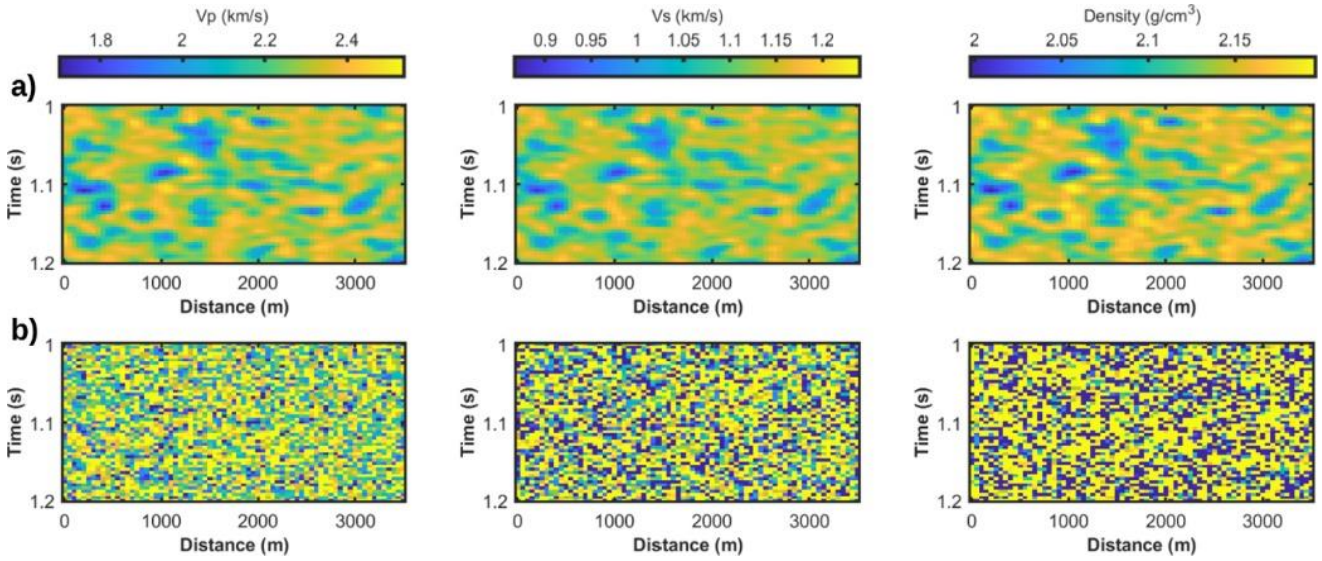


*Figure 90: a) An elastic model extracted from validation set 0. b) The elastic model predicted by a trained CNN with the same architecture as previously described but in this case, the DCT has not been used to compress the input and output responses.*

## 6.5 Inversion Examples

In this section, we describe the application of the trained CNN to invert synthetic seismic data computed on the reference model shown in Figure (83). As previously mentioned, we perform seven inversion tests in which we assess the robustness of the trained CNN inversion. The details of these seven inversion experiments are listed below:

- Test 1: The noise standard deviation and the source wavelet used to generate the observed data input to the CNN are the same as those previously used to generate the training and validation example (see Table 5). This test is aimed at assessing the network performance in the best case.

- Test 2: The standard deviation of the Gaussian and uncorrelated noise that contaminates the data is higher than that assumed in the training phase. The numerical value of the standard deviation is 0.05, which corresponds to 88% of the standard deviation of the observed noise-free data. This test illustrates the network performances in case of an underestimation of the actual noise affecting the data.

- Test 3: In this case, we add to the observed data both uncorrelated and spatially/temporarily correlated Gaussian noise. Both types of noise have a standard deviation equal to 0.03. This

test illustrates the network performances in the case of correlated noise affecting the data. Indeed, the popular assumption of uncorrelated noise may constitute an oversimplification, and correlated noise in the data can be ascribed to the residuals of multiples that have not been successfully removed during the processing phase. Note that in this case, we expect biased predictions because Gaussian and uncorrelated noise in the data was assumed for the training phase.

- Test 4: In this case, the source wavelet used to generate the data has a peak frequency of 25 [Hz], which is smaller than that used to train the network. The noise statistics and the source wavelet phase are the same as those used in the training examples.

- Test 5: In this case, the source wavelet is computed by applying a phase rotation of 30 degrees to the wavelet used in the training process.

- Test 6: The source wavelet is characterized by a peak frequency of 25 Hz and a phase rotation of 30 degrees with respect to the zero-phase wavelet used in the training phase. Tests 4, 5, and 6 investigate the network stability in case of errors in the wavelet estimation process.

- Test 7: The last example is a combination of tests 3 and 6. We generate the observed data with a source wavelet with a peak frequency of 25 Hz and a 30° phase rotation. The observed data are contaminated with both uncorrelated and coherent noise with standard deviations of 0.03.

Figure (91) shows a comparison between the source wavelet used to generate the training examples and those employed in inversion tests 4, 5, 6, and 7 to generate the observed dataset. In all 7 tests, the CNN outcomes are compared with those provided by the linear, trace-by-trace inversion approach of Buland and Omre (2003), for brevity referred to as the standard approach (SA) in the following discussion. In both cases, the structural similarity index (SSI; Wang *et al*. 2004) is used to quantitatively assess the quality of the predictions. Note that the linear inversion assumes a log-Gaussian distribution for the elastic properties and imposes only vertical constraints on the recovered model while it neglects their lateral correlation.

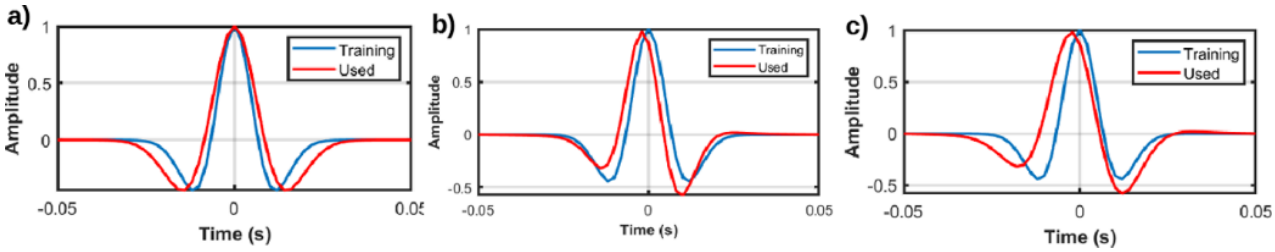


*Figure 91: a) Comparison between the source wavelet used to generate the training examples and that used to compute the observed data in Test 4. b) As in a) but for Test 5. c) As in a) but for Tests 6 and 7.*

In all the linear inversions, the data covariance matrix is computed assuming the noise statistic used to generate the training examples (i.e., Gaussian uncorrelated noise with a standard deviation of 0.02). The vertical correlation range included in the prior model covariance matrix is equal to that used to generate the training and validation sets (4 [ms]).

| **Parameters** | **Training** | **#1** | **#2** | **#3** | **#4** | **#5** | **#6** | **#7** |
|---|---|---|---|---|---|---|---|---|
| ***Uncorrelated noise standard deviation*** | 0.02 | 0.02 | 0.05 | 0.03 | 0.02 | 0.02 | 0.02 | 0.03 |
| ***Correlated noise standard deviation*** | 0 | 0 | 0 | 0.03 | 0 | 0 | 0 | 0.03 |
| ***Source wavelet phase*** | 0° | 0° | 0° | 0° | 0° | 30° | 30° | 30° |
| ***Source wavelet peak frequency*** | 30 Hz | 30 Hz | 30 Hz | 30 Hz | 25 Hz | 30 Hz | 25 Hz | 25 Hz |
| ***Vertical range of correlated noise*** | / | / | / | 12 ms | / | / | / | 12 ms |
| ***Lateral range of correlated noise*** | / | / | / | 150 m | / | / | / | 150 m |

*Table 5: Characteristics of the noise distributions and assumed source wavelets in the training datasets and in the seven different inversion tests.*

For the first inversion test, both the CNN and the linear inversion retrieve similar results that accurately reproduce the lateral and vertical variability of the true Vp, Vs and density models (Figures 92.a, 93.a). However, note that the noise contamination generates some lateral scattering in the SA solutions, while the combined use of DCT and CNN preserves the lateral continuity in the recovered models shown in Figure 91.a.

The lateral scattering affecting the SA results is aggravated in the second test, in which we underestimate the noise affecting the data (Figures 92.b and 93.b). It is expected that this underprediction generates overfitting with the observed data that significantly deteriorates the overall quality of both the CNN and SA predictions. The inclusion of coherent noise in the data results in biased predictions, which are particularly visible around the reservoir zone where the elastic contrasts between the encasing shale and the reservoir sand are overpredicted (Figures 92.c and 93.c).

As expected, errors in the source wavelet still yield biased predictions (Figures 92.d-f and 93.d-f ) characterized by both overpredicted or underpredicted elastic contrasts (e.g., see the results of Test 6 in which we simulate errors in both the peak frequency and the phase of the source wavelet).

The combined presence of errors in the wavelet estimation and coherent noise in the data significantly deteriorates the quality of the retrieved elastic models (Figures 92.g and 93.g). However, in this unfavourable scenario, the use of CNN inversion satisfactorily preserves the lateral continuity of the *Vp*, *Vs*, and density models, thereby facilitating the mapping of the formation boundaries.

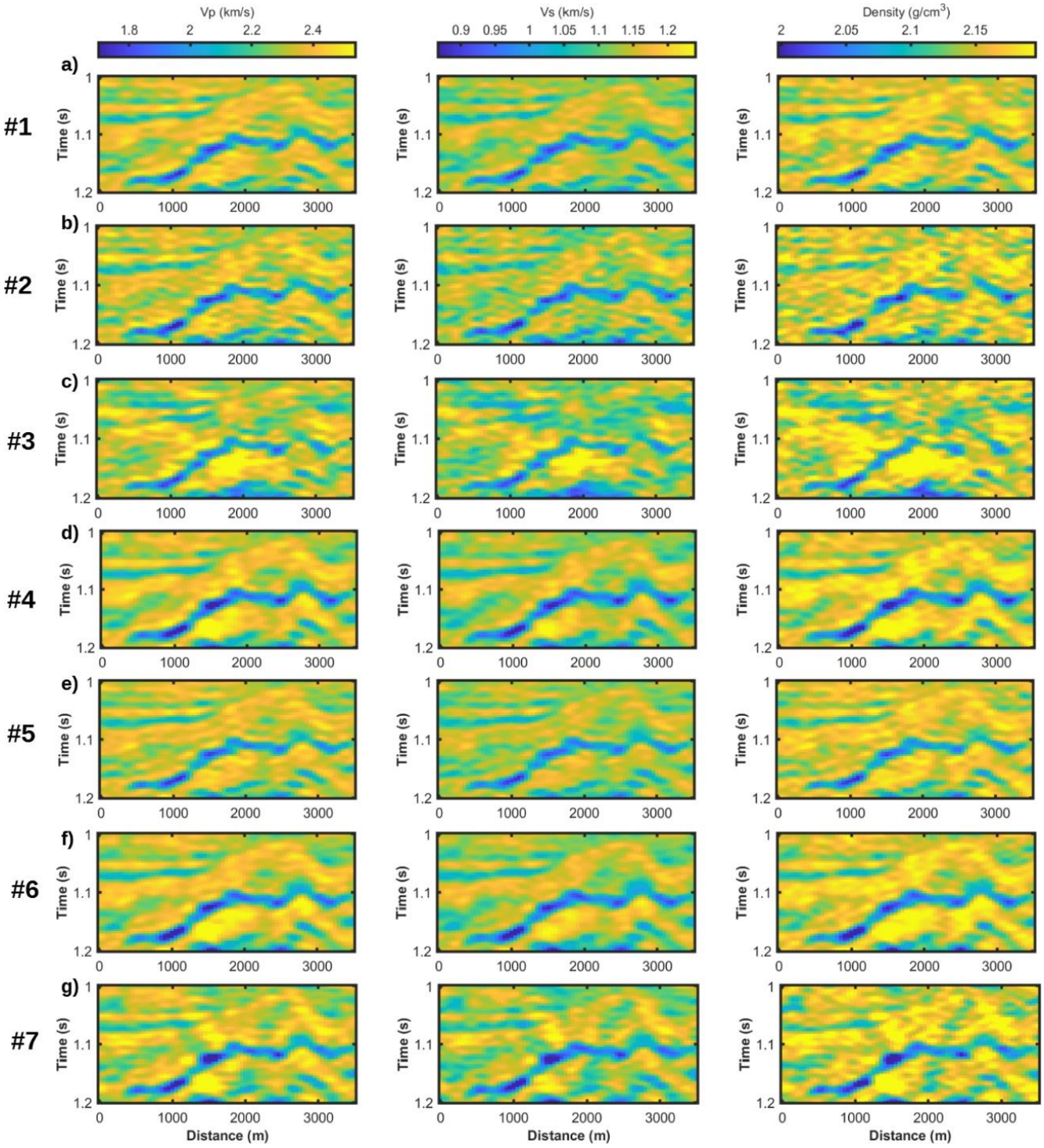


*Figure 92: Vp (left), Vs (centre), and density (right) models estimated by the CNN approach for the seven inversion tests.*

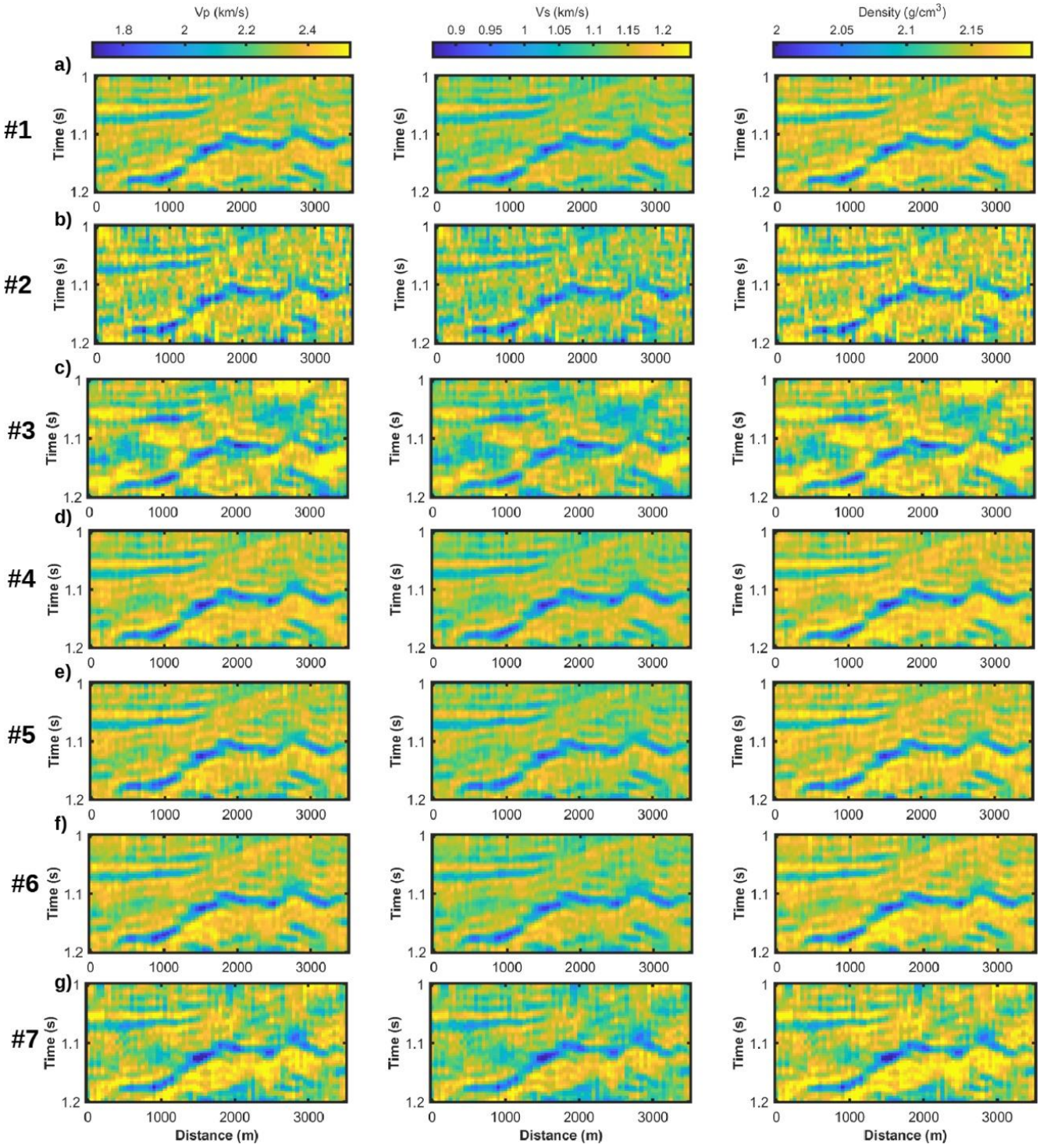


*Figure 93: Vp (left), Vs (centre), and density (right) estimated by the linear inversion for the seven tests.*

The direct comparison of the SSIs yielded by the CNN and SA inversions demonstrates that the implemented approach in all the considered cases guarantees more reliable predictions that are closer to the true elastic model (Figure 94). As expected, for both inversion approaches, we observe

that the quality of the final results decreases passing from Vp to Vs and density in accordance with the different sensitivities of the observed data to the three elastic properties.

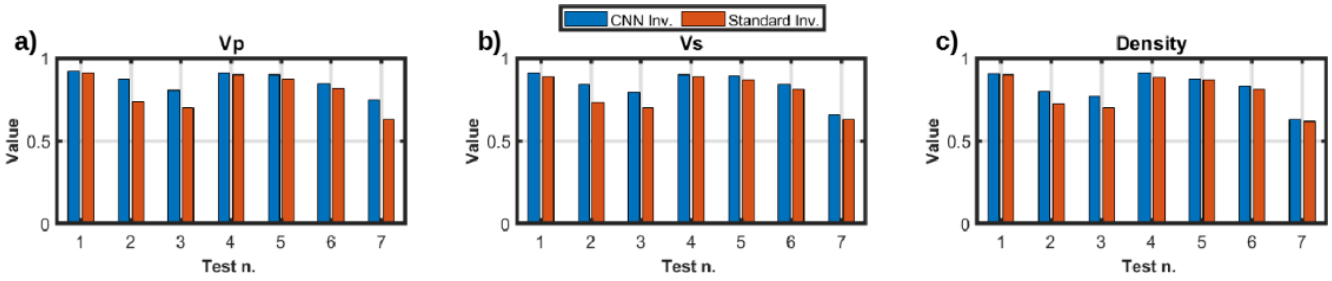


*Figure 94: Comparison of normalized correlation coefficients associated with the CNN and SA results for the seven inversion tests.*

In Figures 95-98 we compare for the Tests 2 and 3 the observed and predicted data resulting from the CNN and SA inversions. These results confirm the applicability of the CNN inversion and its suitability for inverting seismic data in realistic scenarios of heavy noise contamination and errors in the assumed noise statistics and source wavelet estimation.

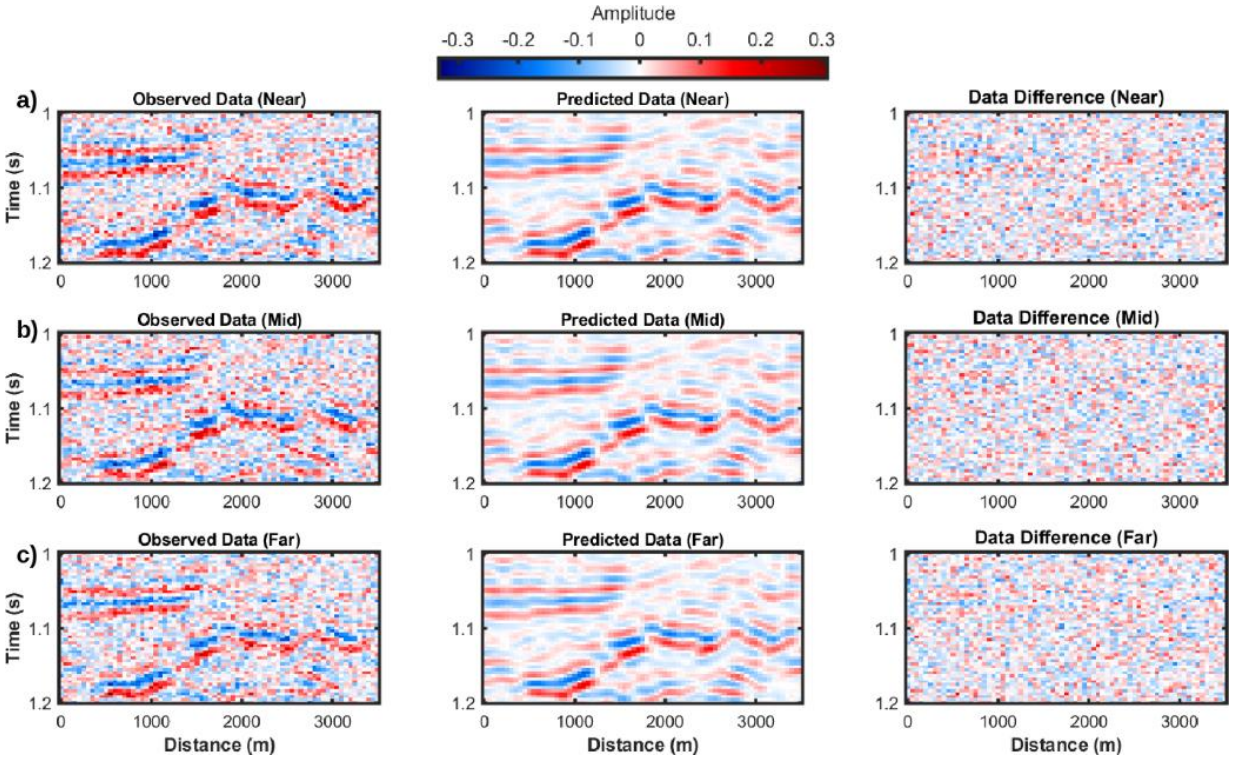


*Figura 95: Comparison between observed data (left column), predicted data (central column), and their sample-by-sample difference (right column) in Test 2 and for the CNN inversion.*

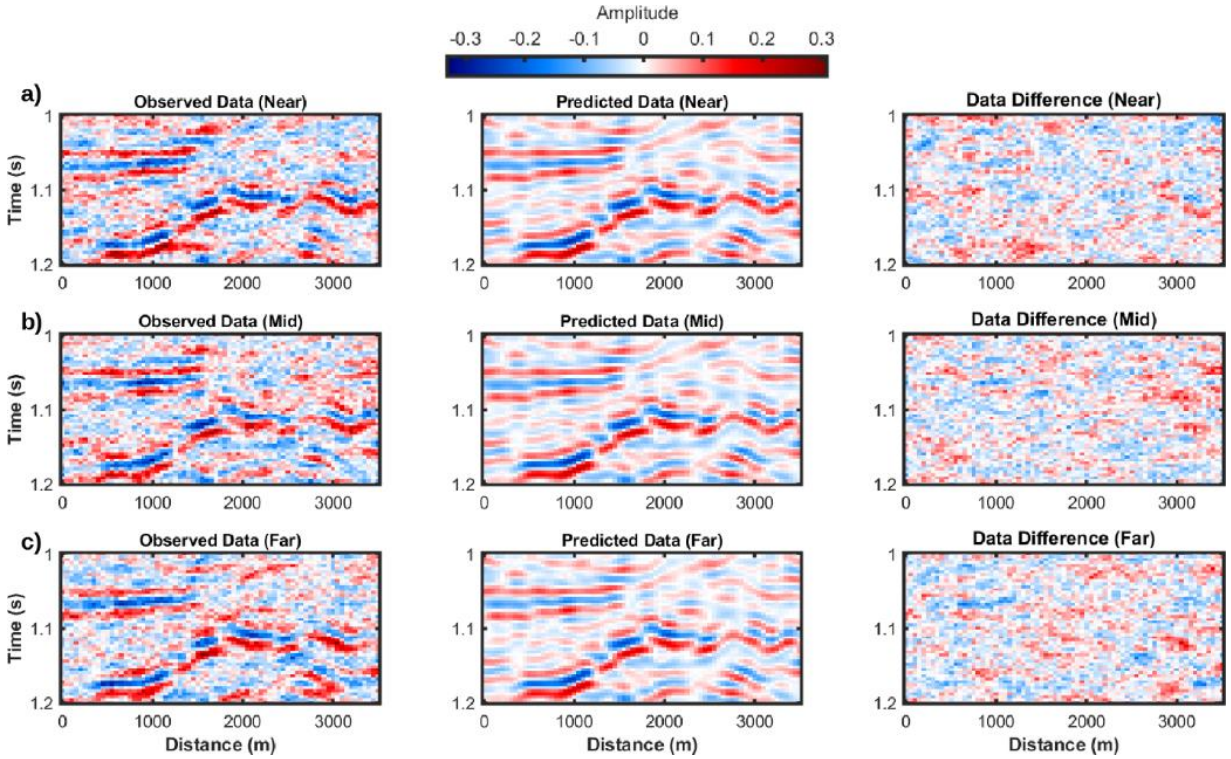


*Figure 96: As in Figure 94 but for Test 3.*

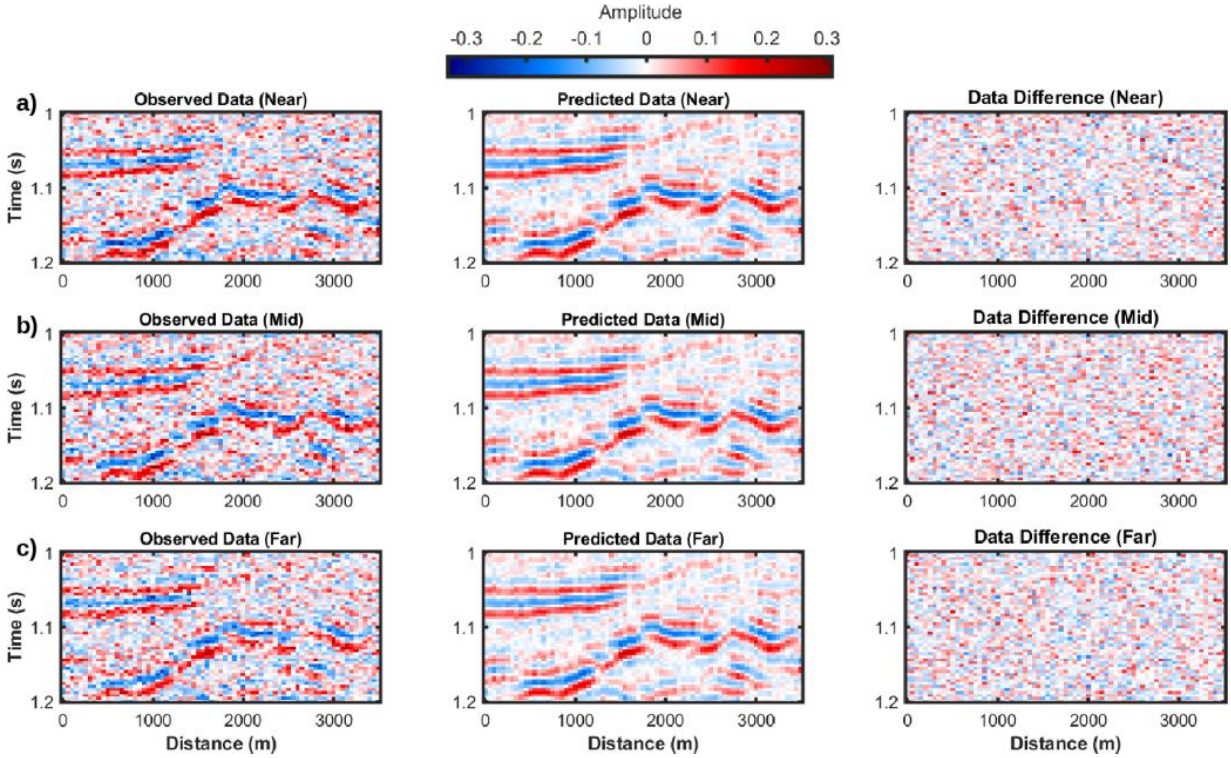


*Figure 97: Comparison between observed data (left column), predicted data (central column), and their sample-by-sample difference (right column) in Test 2 and for the SA inversion.*

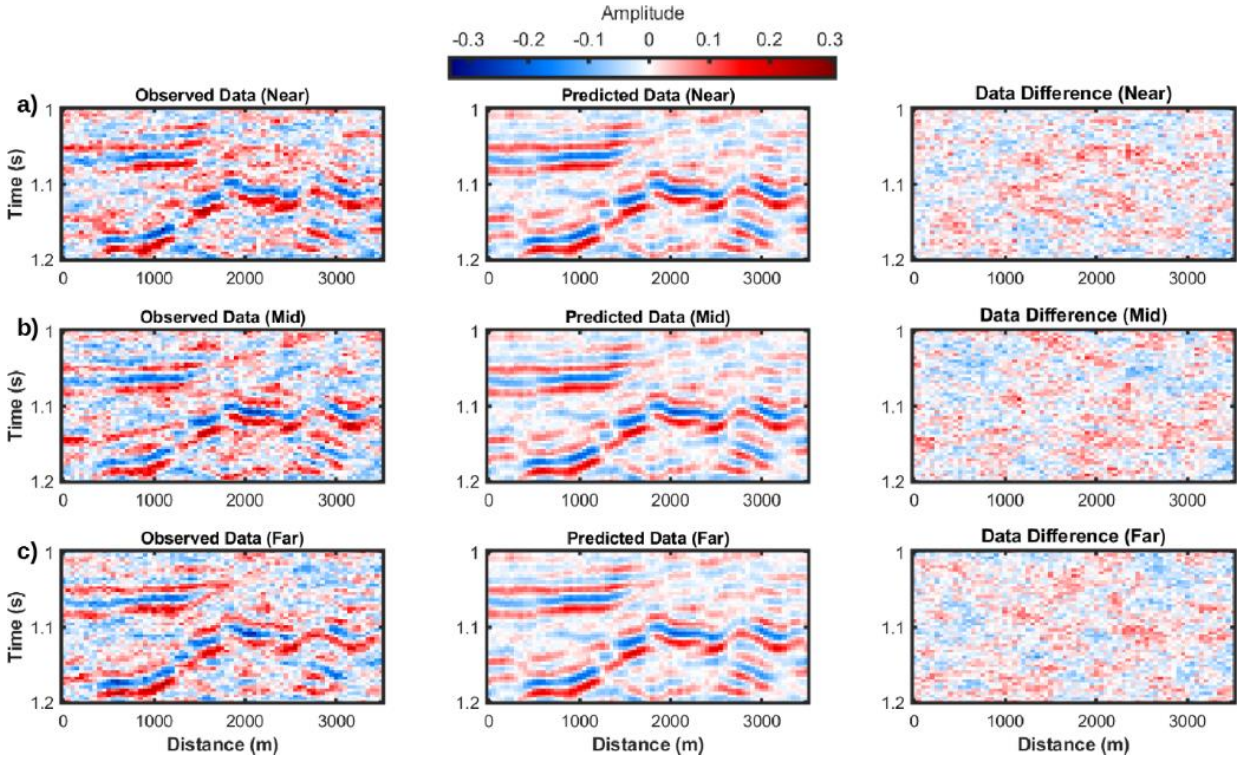


*Figure 98: As in Figure 96 but for Test 3.*

Figure (99) compares the marginal *Vp*, *Vs* and density distributions directly computed from the true model and those derived from the CNN predictions for Tests 2 and 3. When the noise is correctly estimated (Figure 99.a), the marginal distributions computed from the final predictions show good agreement with those derived from the true model. Otherwise, when the noise is underestimated (Figure 99.b), the predicted distributions exhibit longer tails than the true distribution and, in particular, the retrieved density model shows more dispersed values. This degradation of the density prediction is expected since this is the elastic property that exerts a minor influence on the observed seismic data. However, in this case, important statistical properties of the true and predicted model (i.e., the mode) are still very similar.

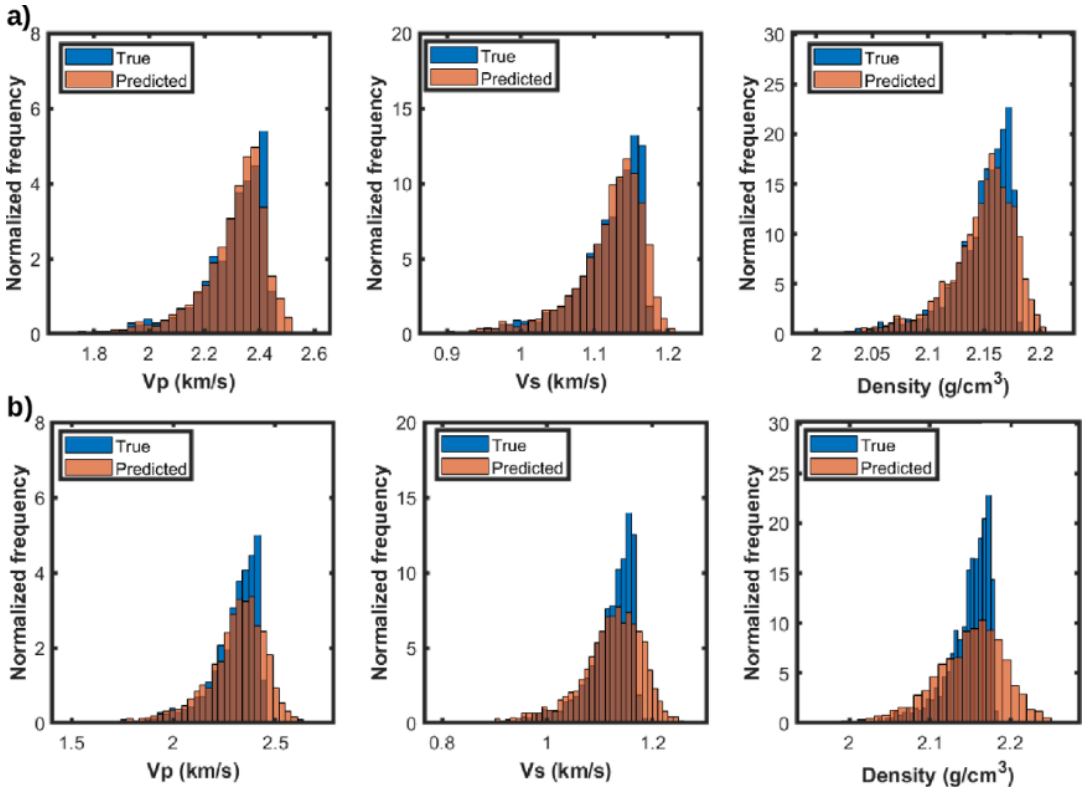


*Figure 99: Comparison between the marginal distributions computed on the true Vp, Vs and density model, and those associated with the CNN predictions. a) for Test 1. b) for Test 2.*

Note that the statistical properties of prior model used to generate the training examples are not derived directly from the true model, but they are realistically inferred from the available well log data measured in the investigated area. This fact could explain the minor deviations between the true and predicted distributions, for example, the different range of variability associated with the true and recovered models. This comparison illustrates that the implemented inversion approach can satisfactorily preserve the actual distribution of the elastic parameters, given that the statistical properties of the prior distribution used to generate the training examples accurately capture the distribution of the *Vp*, *Vs*, and density values in the investigated zone.

Finally, Figures 100-102 illustrate nine MC realizations of the Vp, Vs, and density model drawn for Test 2. All these realizations are in accordance with the assumed noise statistics and with the modelling errors associated with the CNN predictions. As expected, the differences in the nine realizations increase from Vp to Vs and density.

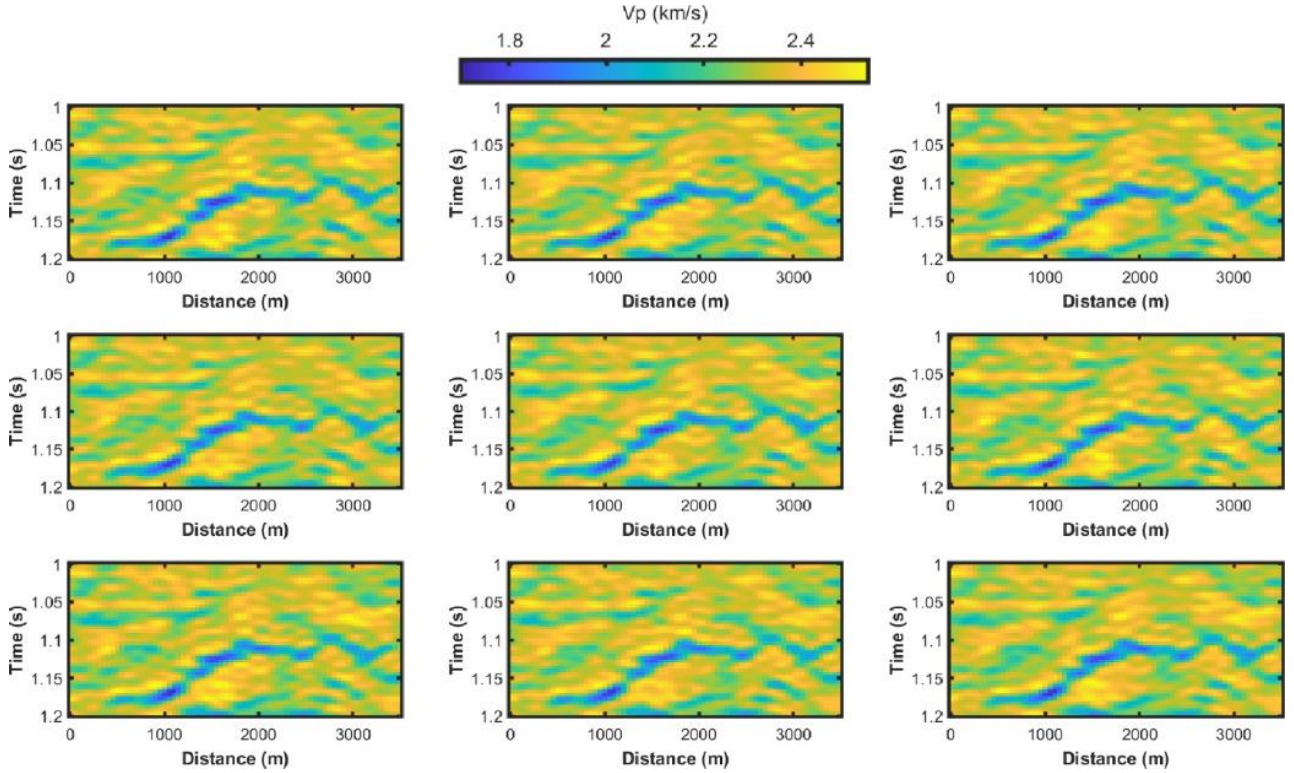


*Figure 100: Nine MC realizations of Vp models associated with Test 2.*

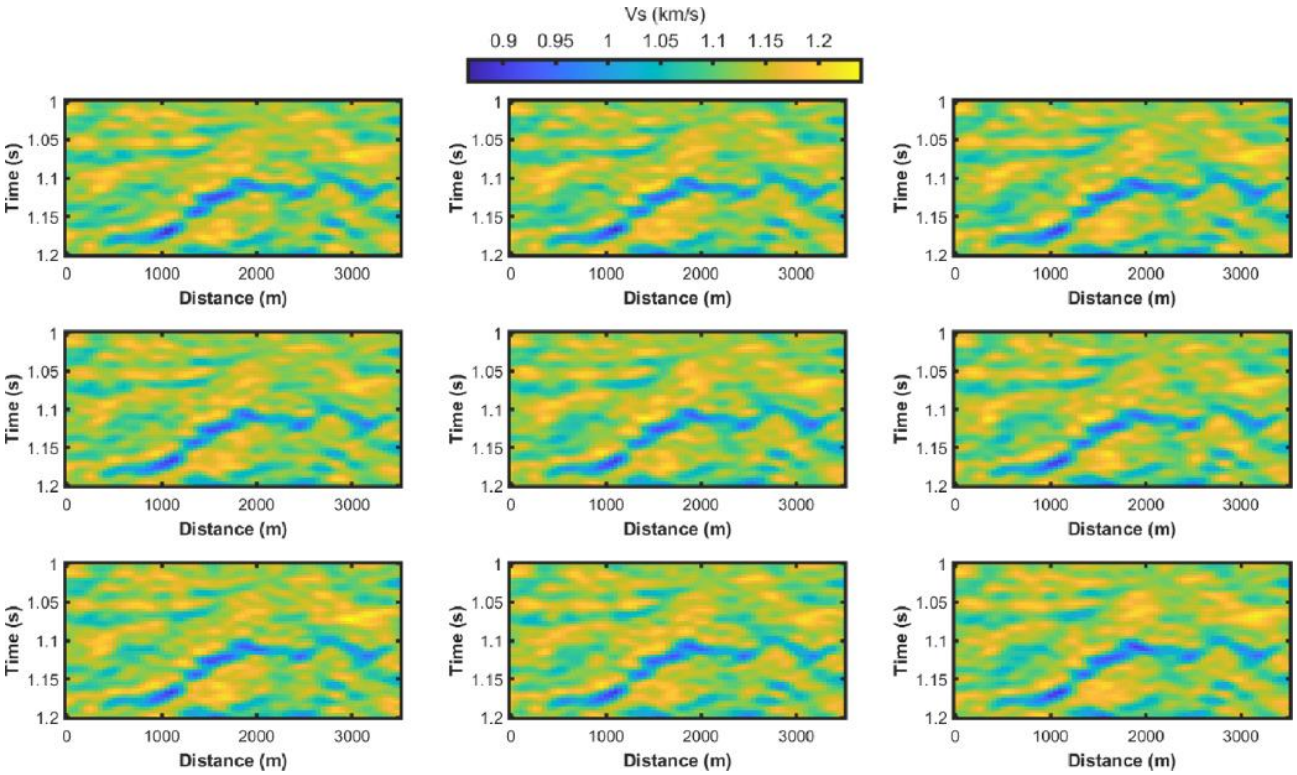


*Figure 101: Nine MC realizations of Vs models associated with Test 2.*

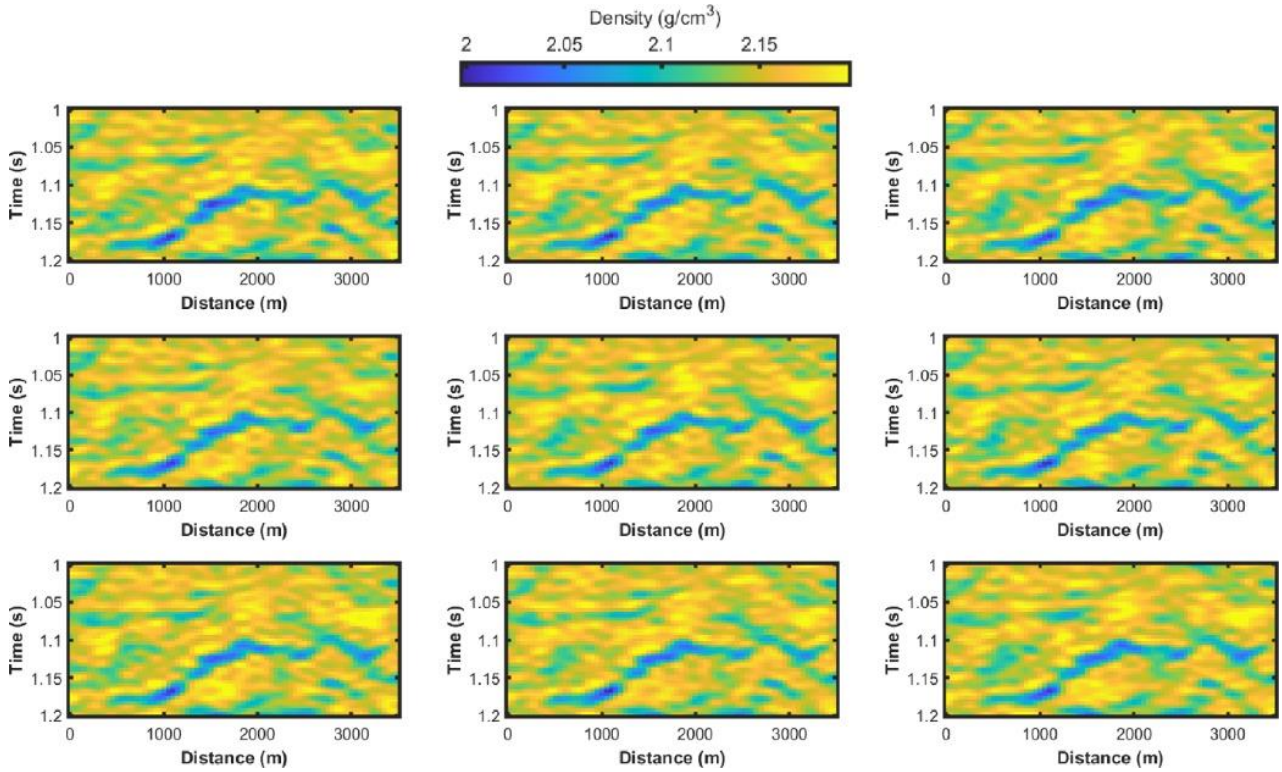


*Figure 102: Nine MC realizations of density models associated with Test 2.*

## 6.6 Conclusions

One of the main practical limitations of inversion approaches with spatial constraints is the high computational cost caused by the considerable dimensionality of the matrices to be inverted. In the same context, the applicability of geostatistical inversions is discouraged by the computational cost of the geostatistical simulation process and by the number of forward modellings needed to converge towards the final solution. Finally, the application of McMC inversions with 2D or 3D spatial constraints is hampered by the considerable number of forward evaluations needed to attain stable posterior estimation and by the usually low acceptance rate of probabilistic sampling.

The popularity of machine learning approaches allows the development of an alternative inversion approach with a modest computational demand, which is robust and also capable of providing a quantification of the uncertainties affecting the final predictions. In particular, the major efforts are devoted to combining the capability of the CNN in determining the function mapping the input data towards its output response and the compression ability of the DCT. Moreover, the use of DCT makes it possible to substantially reduce the dimensionality of the input and output of the network,

thereby reducing the complexity of the network architecture and the computational cost for the training phase; furthermore, this approach does not require any regularization into the inversion framework. Instead, the network is trained on a data set containing realistic models and thus learns to reproduce a similar model that fits the input data. The choice of the number of DCT coefficients to approximate the data and the elastic models should always constitute a compromise between the desired temporal and spatial resolutions and the dimensionality reduction of the parameter space.

The implemented inversion consists of four stages: data generation, learning process, model prediction from a given data input, and Monte Carlo simulation for posterior model generation.
The first stage can be performed in various ways, and here, we use the Co-DSSj geostatistical simulation algorithm that can handle non-parametric prior models while preserving the joint distribution of the elastic properties to be simulated. This stage can be very time consuming, although it is perfectly parallelizable.

The computing time for generating the ensemble of 20,000 training and validation models with the Co-DSSj method is approximately 6 hours, whereas the training phase takes approximately 10 minutes. The CNN inversion gives predictions in real time, while the MC simulation of the nine elastic models runs takes less than 2 s. Therefore, in view of 3D CNN inversions, much effort should be devoted to decreasing the computational effort for the generation of training and validation examples.

The performed synthetic experiments show very promising results and demonstrate that a convolutional neural network can effectively approximate the inverse of a nonlinear operator that is quite difficult and expensive to compute analytically, especially when vertical and lateral constraints are imposed on the recovered solution. The learned network achieves satisfactory predictions when realistic conditions are simulated (e.g., errors in the estimated source wavelet or erroneous assumptions about the noise statistics). In all the tests carried out, the CNN provides more accurate predictions than the standard, laterally unconstrained AVO inversion approach.

Notably, the quality of the CNN predictions seems less affected by erroneous assumptions about the noise statistics and errors in the estimated source wavelet than the outcomes provided by the SA method. In any case, reliable prior model assumptions and high-quality seismic data are essential for the successful application of the implemented inversion approach to field dataset.

# CHAPTER 7

# General conclusions

The inversion procedure is one of the most important processes in reservoir characterization for hydrocarbon exploration, $CO_2$ sequestration and aquifer identification because it allows the extraction of quantitative information about a possible target from seismic data. In particular, we are interested in inverting the amplitude *versus* offset (AVO) response of a single reflection along a previously interpreted 2D stratigraphic horizon (target-oriented inversion) or the seismic responses within a given time interval (interval-oriented inversion). The adopted Bayesian approach allows us to estimate the uncertainties related to the estimated models and to introduce prior information into the inversion framework that can be derived, for example, from well log data.

We focus our efforts on numerical Markov chain Monte Carlo approaches, which are extremely useful when dealing with non-linear, ill-conditioned inverse problems. These iterative approaches basically involve a two-step procedure: in the first step, random models are generated according to some previously defined perturbation method; in the second step, these models are accepted or rejected according to the Metropolis-Hastings acceptance rule. If the newly generated model is accepted, it represents the starting model for the next iteration; otherwise, the algorithm continues to generate new models until one is accepted. This procedure allows the algorithm to move all around the model space and in theoretically infinite iterations reach the stationary regime, where the posterior density function is sampled.

This first algorithm we present is suitable for non-parametric multi-modal prior distributions and non-linear forward models. To increase the algorithm convergence rate, we adopt a multi-chain approach, in which we use both the parallel tempering and the delayed rejection scheme algorithms, and we hybridize the proposed approach with the DEMC method. In addition, to overcome the ill-conditioning of the AVO inversion, we propose the inclusion of geostatistical constraints in the

form of lateral and vertical variogram models that guarantee stable solutions in the case of data strongly contaminated by noise. It is important to remark that the effectiveness of the adopted constraints is subjected to informative prior distributions, here derived from available well log data. The results we obtained show better elastic and litho-fluid facies reconstruction than the results of faster analytical approaches. Nevertheless, the prior non-parametric distributions derived through the kernel density estimator are characterized by small skewness, especially in the target-oriented approach. Future developments will therefore be aimed at assessing the stability of the proposed inversion procedure in the case of a strongly skewed or highly multi-modal prior that could more realistically model the actual distribution of the parameters in the subsurface. In addition, it may be interesting to introduce a rock-physics model into the inversion kernel to estimate the petrophysical properties of the investigated area.

Although the inclusion of geostatistical constraints allows us to strongly reduce the null space of solutions and the ill-conditioning of the problem, it increases the computational effort involved in the sampling procedure. For this reason, in Chapter 3, we propose an alternative approach that reduces the ill-conditioning through a transdimensional framework in which the model parametrization is assumed to be unknown during the inversion procedure. This means that the algorithm is allowed to sample the model parametrization together with the elastic-petrophysical parameters while automatically adjusting the underlying model parametrization. This approach produces solutions with appropriate levels of complexity to fit the data to statistically meaningful levels. Unlike those of previous approaches, the prior distribution is here uninformative (i.e., uniform); this means that no other source of information is used except for the seismic data and that only the transdimensional nature of the algorithm guarantees stable PDFs. Again, the parallel tempering and delayed rejection scheme are employed to accelerate the convergence of the algorithm. The transdimensional approach also theoretically allows us to include the noise spatial correlation and the uncertainty of the source signature in the inversion framework. Clearly, increasing the number of unknowns exponentially increases the computational effort needed for attaining stable PDFs. Future developments will apply the proposed method to invert 3D field data. In this context, it will be crucial to further decrease the computational cost of the inversion procedure, for example, by implementing a more efficient and scalable parallel code.

From the results derived in Chapter 3, we note a possible alternative to externally infused geostatistical constraints. In general, all strategies devoted to attenuating the ill-conditioning of the inverse approaches rely on prior structural knowledge of the investigated area, generally derived

from seismic-independent data, such as well logs or geological information. In the approach described in Chapter 4, we again adopt a transdimensional procedure that uses as model discretization the Voronoi cells, whose number and shape are automatically determined by the algorithm sampling. The algorithm autonomously generates models where each Voronoi cell is composed of similar AVO responses, producing 2D final horizons discretized in cells only on the basis of the lateral variability of data and therefore requiring no regularization operator to force the models to honour any external constraint. Since we average the AVO responses of all CDPs falling into each polygon, as well as the associated elastic properties, we propose an automatic approach that reduces the ill-conditioning of the AVO inversion, thereby guaranteeing stable results in the case of noisy data. The main downside of the proposed inversion algorithm is the extremely long computational time needed to attain stable results even for a limited number of unknowns. For this reason, the application of this algorithm to larger areas or to field data is subjected to a more efficient code.

All numerical approaches described so far have very low acceptance ratios and slow convergence rates. For this reason, in Chapter 5, we explore the potentialities of a Hamiltonian Monte Carlo approach. This approach allows making longer jumps in the model space and generating independent perturbed models, which translates into faster convergence than those for previously described methods. In particular, the higher acceptance ratio is achieved with the introduction of the gradient of the misfit function into the inversion kernel. The main issue here is again related to the computational time required to calculate the derivatives related to the Jacobian matrix. To partially overcome this problem but still produce a very efficient inversion approach, we propose replacing the Jacobian matrix with an operator analytically derived from the linear approximation of the non-linear forward equations, which dramatically reduces the computational effort and assumes a constant curvature of the misfit function over the entire model space. The proposed modification cannot be considered universally valid but needs to be evaluated case by case; nevertheless, it shows promising results. The hyper-parameter tuning here is an extremely delicate process that needs to be executed carefully; however, the extremely fast convergence rate and the results obtained make the HMC one of the most promising approaches adopted to date. Future developments will involve the application of the implemented method to reservoir characterization studies with multi-modal and non-Gaussian target distributions. The issues that all HMC approaches face when the posterior distribution is multi-modal are indeed well known, since the sampling becomes inefficient when the components are separated by high-energy barriers (or low-probability regions). This is an attractive open research field, and in this case, analogous to the first

approach presented here, it could be interesting to analyse the algorithm performances when the priors are converted into Gaussian distributions though the normal score transformation. We are currently applying the HMC approach to other highly non-linear inverse problems, such as full-waveform inversion and surface wave dispersion curves.

Finally, Chapter 6 addresses the ill-conditioning of the AVO inversion using a deep learning approach. Here, we explore the potential of a convolutional neural network, a supervised learning algorithm. The implemented method does not require regularization in its common-sense meaning (i.e., the inclusion of model constraints into the error function), but the network is trained on a data set containing realistic subsurface scenarios and thus learns to reproduce a similar model that fits the input data. The lateral dependencies are imposed in the training and validation datasets through a direct sequential co-simulation method. Therefore, we do not introduce external constraints that translate into larger matrices to invert (and therefore increased computational efforts), but we force the network to generate models that already honour the imposed lateral variability. Since we are dealing with high-dimensional problems, we apply a discrete cosine transform (DCT) to the net input and output to reparametrize the problem in a lower-dimensional space. In our case, the DCT constitutes an additional feature extraction tool that uses orthogonal basis functions to compress the dimensionalities of the input and output of the CNN. On the other hand, DCT reparameterization also acts as a model regularization strategy that reduces the number of unknown parameters to be estimated and satisfactorily preserves the assumed lateral continuity in the recovered solution.
The potentialities of deep learning approaches are almost infinite, as they are a wide open research field, and applications can be found in many different scientific fields. Even if our results are promising, it is worth mentioning that the inversion procedure is here transformed into a mathematical optimization of the neuron weights, which no longer has any geophysical meaning. This easily can turn into a black box methodology that the operator uses almost blindly and that should be evaluated with extreme caution from case to case.
The future applications of the proposed approach are nevertheless countless: the combination of CNN and DCT could be actively used to make the previously described approaches faster, for example, to quickly calculate the exact Jacobian matrix for the HMC. This paves the way to a new branch of inversion approaches that exploit a convolutional neural network to approximate the inverse of either non-linear or linear operators that are very difficult and expensive to compute analytically.

**REFERENCES**

Adriansyah, & McMechan, G. A. (2001). AVA analysis and interpretation of a carbonate reservoir: northwest Java basin, Indonesia. *Geophysics, 66*(3), 744-754.

Agostinetti, N. P., & Malinverno, A. (2010). Receiver function inversion by transdimensional Monte Carlo sampling. *Geophysical Journal International*, *181*(2), 858-872.

Avseth, P., T. Mukerji, and G. Mavko, 2010, Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk: Cambridge university press.

Avseth, P., Flesche, H., & Van Wijngaarden, A. J. (2003). AVO classification of lithology and pore fluids constrained by rock physics depth trends. *The Leading Edge*, *22*(10), 1004-1011.

Richards, P. G., & Aki, K. (1980). *Quantitative Seismology: Theory and Methods* (p. 13). Freeman.

Aleardi, M., Ciabarri, F., & Gukov, T. (2018). A two-step inversion approach for seismic-reservoir characterization and a comparison with a single-loop Markov-chain Monte Carlo algorithm. *Geophysics*, *83*(3), R227-R244.

Aleardi, M., Ciabarri, F., & Gukov, T. (2019). Reservoir characterization through target-oriented AVA-petrophysical inversions with spatial constraints. *Pure and Applied Geophysics*, *176*(2), 901-924.

Aleardi, M., F. Ciabarri, and A. Mazzotti, (2017), Probabilistic estimation of reservoir properties by means of wide-angle AVA inversion and a petrophysical reformulation of the Zoeppritz equations: Journal of Applied Geophysics, 147, 28-41.

Ando, T. (2010). Bayesian model selection and statistical modeling. CRC Press.

Araya-Polo, M., Jennings, J., Adler, A., & Dahlke, T. (2018). Deep-learning tomography. *The Leading Edge*, *37*(1), 58-66.

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

Avseth, P., Mukerji, T., & Mavko, G. (2005). *Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk*. Cambridge university press.

Bancroft, J. C. &Feng, H., (2006). AVO principles, processing and inversion. *CREWES Research Report*, *18*, 1-19.

Barclay, F., Bruun, A., Rasmussen, K. B., Alfaro, J. C., Cooke, A., Cooke, D., ... & Özdemir, H. (2008). Inversión sísmica: Lectura entre líneas. *Shulumberger OilField review*, 1-23.

Batzle, M., & Wang, Z. (1992). Seismic properties of pore fluids. *Geophysics*, *57*(11), 1396-1408.

Beichl, I., & Sullivan, F. (2000). The metropolis algorithm. *Computing in Science & Engineering*, *2*(1), 65-69.

Berryman, J. G. (1999). Origin of Gassmann's equations. *Geophysics*, *64*(5), 1627-1629.

Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*.

Bianco, E. (2016). Tutorial: Wavelet estimation for well ties. *The Leading Edge*, *35*(6), 541-543.

Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*(3), 1411-1436.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., & Rawlinson, N. (2012). Transdimensional inversion of receiver functions and surface wave dispersion. *Journal of Geophysical Research: Solid Earth*, *117*(B2).

Bosch, M., Cara, L., Rodrigues, J., Navarro, A., & Díaz, M. (2007). A Monte Carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes. *Geophysics*, *72*(6), O29-O39.

Buland A., (2002), Bayesian seismic AVO inversión. Norwegian University of Science and Technology. Trondheim, Norway.

Buland, A., & Omre, H. (2003). Bayesian linearized AVO inversion. *Geophysics*, *68*(1), 185-198.

Burianyk, M., & Pickfort, S. (2000). Amplitude-vs-offset and seismic rock property analysis: A primer. *CSEG Recorder*, *25*(9), 6-16.

Castagna, J. P., & Backus, M. M. (1993). AVO analysis—Tutorial and review. *Offset-dependent reflectivity—Theory and practice of AVO analysis: SEG Investigations in Geophysics*, *8*, 3-36.

Castagna, J. P. & Chopra, S., (2014). *AVO*. Society of Exploration Geophysicists.

Cordua, K. S., Hansen, T. M., & Mosegaard, K. (2012). Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information. *Geophysics*, *77*(2), H19-H31.

Das, V., Pollack, A., Wollner, U., & Mukerji, T. (2019). Convolutional neural network for seismic impedance inversion. In *SEG Technical Program Expanded Abstracts 2018* (pp. 2071-2075). Society of Exploration Geophysicists

De Figueiredo, L. P., Grana, D., Santos, M., Figueiredo, W., Roisenberg, M., & Neto, G. S. (2017). Bayesian seismic inversion based on rock-physics prior modeling for the joint estimation of acoustic impedance, porosity and lithofacies. *Journal of Computational Physics*, *336*, 128-142.

De Figueiredo, L. P., Grana, D., Bordignon, F. L., Santos, M., Roisenberg, M., & Rodrigues, B. B. (2018). Joint Bayesian inversion based on rock-physics prior modeling for the estimation of spatially correlated reservoir properties. *Geophysics*, *83*(5), M49-M61.

Dosso, S. E., Dettmer, J., Steininger, G., & Holland, C. W. (2014). Efficient transdimensional Bayesian inversion for geoacoustic profile estimation. *Inverse Problems*, *30*(11), 114018.

Dosso, S. E., Holland, C. W., & Sambridge, M. (2012). Parallel tempering for strongly nonlinear geoacoustic inversion. *The Journal of the Acoustical Society of America*, *132*(5), 3030-3040.

Drufuca, G., & Mazzotti, A. (1995). Ambiguities in AVO inversion of reflections from a gas-sand. *Geophysics*, *60*(1), 134-141.

Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid monte carlo. *Physics letters B*, *195*(2), 216-222.

Duijndam, A. J. W. (1988). Bayesian estimation in seismic inversion. part I: principles 1. *Geophysical Prospecting*, *36*(8), 878-898.

Fichtner, A., Zunino, A., & Gebraad, L. (2019). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, *216*(2), 1344-1363.

Flournoy, N., Tsutakawa, R. K., & Tsutakawa, R. (1991). *Statistical Multiple Integration: Proceedings of a Joint Summer Research Conference Held at Humboldt University, June 17-23, 1989* (Vol. 115). American Mathematical Soc..

Gassmann, F. (1951). Über die elastizität poröser medien: Vierteljahrss-chrift der Naturforschenden Gesellschaft in Zurich 96, 1-23. *Paper translation at http://sepwww. stanford. edu/sep/berryman/PS/gassmann. pdf*.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gosset, W. S. (1909). The brewing of the experimental hop farm hops, 1907 Crop (Part II), Together with a note on the present method of hop analysis. *Laboratory Report*, *10*, 202-220.

Grana, D. (2018). Joint facies and reservoir properties inversion. *Geophysics*, *83*(3), M15-M24.

Grana, D., & Della Rossa, E. (2010). Probabilistic petrophysical-properties estimation integrating statistical rock-physics with seismic inversion. *Geophysics*, *75*(3), O21-O37.

Grana, D., Mukerji, T., Dovera, L., & Della Rossa, E. (2012). Sequential simulations of mixed discrete-continuous properties: Sequential Gaussian mixture simulation. In *Geostatistics Oslo 2012* (pp. 239-250). Springer, Dordrecht.

Gunning, J., & Sams, M. (2018). Joint facies and rock properties Bayesian amplitude-versus-offset inversion using Markov random fields. *Geophysical Prospecting*, *66*(5), 904-919.

Haas, A., & Dubrule, O. (1994). Geostatistical inversion-a sequential method of stochastic reservoir modelling constrained by seismic data. *First break*, *12*(11).

Hansen, T. M., & Cordua, K. S. (2017). Efficient Monte Carlo sampling of inverse problems using a neural network-based forward—applied to GPR crosshole traveltime inversion. *Geophysical Journal International*, *211*(3), 1524-1533.

Hashin, Z., & Shtrikman, S. (1963). A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*, *11*(2), 127-140.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).

Horta, A., & Soares, A. (2010). Direct sequential co-simulation with joint probability distributions. *Mathematical Geosciences*, *42*(3), 269-292.

Jeong, C., Mukerji, T., & Mariethoz, G. (2011, December). Adaptive Spatial Resampling for Seismic Inverse Modeling. In *AGU Fall Meeting Abstracts* (Vol. 2011, pp. NS33B-1595).

Jullum, M., & Kolbjørnsen, O. (2016). A Gaussian-based framework for local Bayesian inversion of geophysical data to rock properties. *Geophysics*, *81*(3), R75-R87.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Lan, S., Streets, J., & Shahbaba, B. (2014, July). Wormhole hamiltonian monte carlo. In *Proceedings of the... AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence* (Vol. 2014, p. 1953). NIH Public Access.

Lange, K., Frydendall, J., Cordua, K. S., Hansen, T. M., Melnikova, Y., & Mosegaard, K. (2012). A frequency matching method: solving inverse problems by use of geologically realistic prior information. *Mathematical geosciences*, *44*(7), 783-803.

Larsen, A. L., Ulvmoen, M., Omre, H., & Buland, A. (2006). Bayesian lithology/fluid prediction and simulation on the basis of a Markov-chain prior model. *Geophysics*, *71*(5), R69-R78.

Le Ravalec, M., Noetinger, B., & Hu, L. Y. (2000). The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations. *Mathematical Geology*, *32*(6), 701-723.

Lewis, W., & Vigh, D. (2017). Deep learning prior models from seismic images for full-waveform inversion. In *SEG Technical Program Expanded Abstracts 2017* (pp. 1512-1517). Society of Exploration Geophysicists.

Lochbühler, T., Breen, S. J., Detwiler, R. L., Vrugt, J. A., & Linde, N. (2014). Probabilistic electrical resistivity tomography of a CO2 sequestration analog. *Journal of Applied Geophysics*, *107*, 80-92.

MacKay, D. J., & Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Mackenze, P. B. (1989). An improved hybrid Monte Carlo method. *Physics Letters B*, *226*(3-4), 369-371.

Madsen, R. B., & Hansen, T. M. (2018). Estimation and accounting for the modeling error in probabilistic linearized amplitude variation with offset inversion. *Geophysics*, *83*(2), N15-N30.

Malinverno, A., & Briggs, V. A. (2004). Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, *69*(4), 1005-1016.

Martin, G. S., Wiley, R., & Marfurt, K. J. (2006). Marmousi2: An elastic upgrade for marmousi: Th e Leading Edge, 25, 156–166.

Matheron, G., Beucher, H., De Fouquet, C., Galli, A., Guerillot, D., & Ravenne, C. (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs: SPE Annual Technical Conference and Exhibition, SPE. *Extended Abstracts, SPE-16753–MS*.

Mavko, G., T. Mukerji, and J. Dvorkin, 2009, *The rock physics handbook: Tools for seismic analysis of porous media*. Cambridge University Press.

Mazzotti, A., & Ravagnan, G. (1995). Impact of processing on the amplitude versus offset response of a marine seismic data set1. *Geophysical prospecting*, *43*(3), 263-281.

Mazzotti, A., & Zamboni, E. (2003). Petrophysical inversion of AVA data: Geophysical Prospecting.

Menke, W. (1989). Geophysical data analysis: Discrete inverse theory, volume 45 of. *International Geophysics Series*.

Menke, W. (2018). *Geophysical data analysis: Discrete inverse theory*. Academic press..

Metropolis, N., & Ulam, S. (1949). The monte carlo method. *Journal of the American statistical association*, *44*(247), 335-341.

Neal, R. M., Brooks, S., Gelman, A., Jones, G. & Meng, X. L. (2011). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 113-162.

Puzyrev, V. (2019). Deep learning electromagnetic inversion with convolutional neural networks. *Geophysical Journal International*, *218*(2), 817-832.

Richardson, A. (2018). Seismic full-waveform inversion using deep learning tools and techniques. *arXiv preprint arXiv:1801.07232*.

Rimstad, K., & Omre, H. (2010). Impact of rock-physics depth trends and Markov random fields on hierarchical Bayesian lithology/fluid prediction. *Geophysics*, *75*(4), R93-R108.

Sabeti H., Moradzadeh A., Ardejani F.D., Azevedo L., Soares A., Pereira P. and Nunes R. 2017. Geostatistical seismic inversion for non-stationary patterns using direct sequential simulation and cosimulation . *Geophysical Prospecting*, *65*, 25-48.

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

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

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

Scales, J. A., & Snieder, R. (2000). The anatomy of inverse problems. *Geophysics*, *65*(6), 1708-1710.

Schatzinger, R. A., & Jordan, J. F. (1999). Reservoir Characterization. *Recent Advances: AAPG Memoir*, *71*.

Sen, M. K., & Biswas, R. (2017). Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm. *Geophysics*, *82*(3), R119-R134.

Shuey, R. T. (1985). A simplification of the Zoeppritz equations. *Geophysics*, *50*(4), 609-614.

Tarantola, A., (1987), Inverse Problem Theory: Methods of Data Fitting and Model Parameter Estimation. New York: Elsevier.

Tetyukhina, D. (2011). High-resolution reservoir characterization by seismic inversion with geological constraints. *Geophysics*, *76*(1), Z2-Z2.

Ursenbach, C. P., & Stewart, R. R. (2008). Two-term AVO inversion: Equivalences and new methods. *Geophysics*, *73*(6), C31-C38.

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273-316.

Waldeland, A. U., Jensen, A. C., Gelius, L. J., & Solberg, A. H. S. (2018). Convolutional neural networks for automated seismic interpretation. *The Leading Edge*, *37*(7), 529-537.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.

Xiang, E., Guo, R., Dosso, S. E., Liu, J., Dong, H., & Ren, Z. (2018). Efficient hierarchical transdimensional Bayesian inversion of magnetotelluric data. *Geophysical Journal International*, *213*(3), 1751-1767.

Xiong, W., Ji, X., Ma, Y., Wang, Y., AlBinHassan, N. M., Ali, M. N., & Luo, Y. (2018). Seismic fault detection with convolutional neural network. *Geophysics*, *83*(5), O97-O103.

Yilmaz, Ö. (2001). *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of exploration geophysicists.

Yuan, S., Liu, J., Wang, S., Wang, T., & Shi, P. (2018). Seismic waveform classification and first-break picking using convolution neural networks. *IEEE Geoscience and Remote Sensing Letters*, *15*(2), 272-276.

Zhu, D., & Gibson, R. (2018). Seismic inversion and uncertainty quantification using transdimensional Markov chain Monte Carlo method. *Geophysics*, *83*(4), R321-R334.

Zhu, X., & McMechan, G. A. (2012). Elastic inversion of near-and postcritical reflections using phase variation with angle. *Geophysics*, *77*(4), R149-R159.

Zoeppritz, K. (1919). Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen: Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. *Mathematisch-Physikalische Klasse*, *1*, 66-84.

Zunino, A., Mosegaard, K., Lange, K., Melnikova, Y., & Mejer Hansen, T. (2015). Monte Carlo reservoir analysis combining seismic reflection data and informed priors. *Geophysics*, *80*(1), R31-R41.