



UNIVERSITY OF COPENHAGEN  
FACULTY OF SCIENCE

Ph.D. Thesis

**Uncertainty Quantification in Seismic Subsurface  
Modelling by Informed Proposal Monte Carlo**

Author: Sarouyeh Khoshkholgh

This thesis has been submitted to the PhD School of The Faculty of Science,  
University of Copenhagen, on September 1st, 2021  
Department: Niels Bohr Institute, Physics of Ice, Climate and Earth  
Supervisor: Professor Klaus Mosegaard



## **Doctor of Philosophy Thesis in Computational Geophysics**

**Title:** Uncertainty Quantification in Seismic Subsurface Modelling by Informed Proposal Monte Carlo

**Danish title:** Kvantificering af usikkerhed i seismisk undergrundsmodellering ved hjælp af Informed Proposal Monte Carlo

**Institute:** Niels Bohr Institute

**Group:** Physics of Ice, Climate and Earth

**Thesis delivered on:** September 1, 2021

**Author:** Ph.D. candidate, M.Sc. Sarouyeh Khoshkholgh, University of Copenhagen

**E-mail:** sarouyeh@nbi.ku.dk

**Supervisor:** Professor Klaus Mosegaard, University of Copenhagen



## **Abstract**

A key feature of optimizing subsurface resource exploration is accurate geophysical modelling. In this regard, associating, combining and integrating petrophysical, geological and geophysical data play a crucial role. To do so, solving probabilistic inverse problems of desired model parameters in the subsurface is used to describe reservoir properties. Subsurface uncertainty quantification is obtained through probabilistic solutions to corresponding geo-statistical and geophysical inverse problems. In geoscience, Monte Carlo sampling methods are widely used in producing solutions for nonlinear inverse problems. A fundamental problem when using Monte Carlo search or sampling algorithms is the inefficiency due to the high computational cost of forward calculations, particularly when dealing with large scale inverse problems. This thesis addresses this issue and describes a new methodology that significantly improves the performance of MCMC algorithms, resulting in more effective uncertainty analysis.

There are a number of algorithms that attempt to guide Monte Carlo sampling by exploring the target distribution while it is being performed. However, many of them are limited by the No-Free-Lunch theorem. According to the No-Free-Lunch theorem, the more information about the problem we add, the more efficient algorithm could potentially be. This study presents a new methodology for the Markov Chain Monte Carlo (MCMC) algorithm designed for highly nonlinear problems with computationally expensive forward calculations and a large number of model parameters.

In this thesis, we first explain Informed Proposal Monte Carlo, in which information about the target distribution is introduced to the sampling procedure by using a global proposal distribution. This can be achieved by finding an approximate posterior distribution and using it as global proposal in MCMC algorithm. This proposal distribution is problem dependent and typically calculated using simplified physics.

Afterwards we review some of the most recent, blind and informed MCMC algorithms. Then the theoretical and methodological framework of our approach is presented. We introduce our specific strategy for generating prior models in which image warping is used to perturb the subsurface velocity. Finally, we apply our proposed methodology to the probabilistic problem of full-waveform inversion of seismic data with a large number of model parameters. The results indicate that injecting external information in the form of a global proposal can significantly reduce the convergence time and increase efficiency of the algorithm.

## Dansk Resume

Afgørende for optimering af geofysisk ressourceeftersforskning er brugen af nøjagtige og beregningseffektive modelleringsmetoder. I den henseende spiller integration af petrofysiske, geologiske og geofysiske data en afgørende rolle. Til det formål udnyttes sandsynlighedsbaserede inverse metoder til beregning af de ønskede modelparametre i undergrunden. Kvantificering af usikkerhed af undergrundsparemetre opnås ved at samle sandsynlige løsninger til geostatistiske og geofysiske inverse problemer. Inden for geovidenskab anvendes Monte Carlo samplingmetoder i vid udstrækning til generering af løsninger til ikke-lineære inverse problemer. Et grundlæggende problem ved disse metoder er ineffektivitet på grund af høje beregningsomkostninger ved fremadberegninger (beregning af data ud fra modelparametre), især når der er tale om store inverse problemer.

Denne afhandling behandler dette problem og beskriver en ny metode, der forbedrer ydeevnen for MCMC-algoritmer betydeligt, hvilket resulterer i en mere effektiv usikkerhedsanalyse.

Der findes en række algoritmer, der forsøger at styre Monte Carlo-sampling ved at indsamle information om sandsynlighedsfordelingen mens den samples. Mange af dem er dog begrænset af No-Free-Lunch-teoremet. Ifølge No-Free-Lunch-teoremet er det sådan, at jo flere oplysninger om problemet vi tilføjer, jo mere effektiv vil algoritmen være. Denne undersøgelse præsenterer en ny Markov Chain Monte Carlo (MCMC)-algoritme designet til meget ikke-lineære problemer med beregningsdyre fremadberegninger og et stort antal

modelparametre. I denne afhandling forklarer vi først informed-proposal Monte Carlo, hvor information om fordelingen indbygges i samplingproceduren ved hjælp af en global proposal-fordeling. Dette kan opnås ved at finde en omtrentlig posterior fordeling og bruge den som en global proposal i MCMC-algoritmen. Proposalfordelingen er problemafhængig, og den beregnes ved hjælp af en forenklet fysisk model.

Først gennemgår vi nogle af de nyeste "blinde" og informerede MCMC-algoritmer i denne afhandling. Derefter præsenteres den teoretiske og metodiske ramme for vores tilgang. Vi introducerer vores specifikke strategi til generering af a priori model. Endelig anvender vi vores foreslåede metode til sandsynlighedsbaseret, fuld bølgeforminversion med et stort antal modelparametre. Resultaterne indikerer, at brugen af fysisk baseret ekstern information i form af en global proposalfordeling kan reducere konvergenstiden betydeligt og øge effektiviteten af algoritmen.

## **Acknowledgments**

I would like to offer my special thanks to my supervisor Prof. Klaus Mosegaard for the continuous support of my Ph.D. study and related research, for his patience, motivation, encouragement and immense knowledge. Without his precious support it would not be possible to conduct this research.

I would like to thank Dr Andrea Zunino for his help and advice throughout the project and all the current and former members of Physics of Ice, Climate and Earth group at Niels Bohr Institute for collaboration, encouragement and such great working environment. I would also like to thank Lloyd's Register for a good partnership during this Ph.D. project.

And my biggest thanks to my family for all the love, support and encouragement they have shown through this research.

## **Preface**

This PhD thesis entitled “Uncertainty Quantification in Subsurface Modelling by Informed Proposal Monte Carlo“ was submitted at the University of Copenhagen, Niels Bohr Institute, Physics of Ice, Climate and Earth group in partial fulfillment of PhD requirements. This research project has been supervised by Professor Klaus Mosegaard. This PhD project has been funded by Innovation Fund Denmark through the OPTION project, grant number: 5184-00025B. This project (OPTION-Optimizing Oil Production by Novel Technology Integration) is a joint industry project with Lloyd’s Register, Welltec, Technical University of Denmark (DTU) and University of Copenhagen. This work is part of OPTION project (Work Package 3) named “Reservoir Uncertainty Analysis”.

# *Contents*

<b>1</b>	<b>Introduction .....</b>	<b>1</b>
1.1	Inverse problem .....	2
1.2	Monte Carlo methods .....	6
1.3	Blind algorithms.....	10
1.4	Our direction and contribution .....	12
<b>2</b>	<b>Monte Carlo with problem dependent proposal.....</b>	<b>14</b>
2.1	Sampling probability density.....	14
2.2	Extended Metropolis .....	16
2.3	Proposal distribution.....	16
2.4	Informed proposal Monte Carlo.....	19
<b>3</b>	<b>Forward modelling and prior computation.....</b>	<b>24</b>
3.1	Forward modelling.....	24
3.2	Prior Computation.....	25
3.2.1	Image Warping.....	26
3.2.2	Producing samples from the prior.....	27
<b>4</b>	<b>Full-waveform inversion using Informed proposal MCMC.....</b>	<b>32</b>
4.1	Sampling with an Informed proposal.....	33
4.2	IPMC procedure for sampling the posterior probability.....	
	in full-waveform inversion .....	34

<b>5 Discussion.....</b>	<b>37</b>
<b>6 Conclusion.....</b>	<b>42</b>
<b>References.....</b>	<b>45</b>
<b>List of Figures.....</b>	<b>51</b>
<b>Appendices</b>	
<b>A Ricker Wavelet.....</b>	<b>53</b>
<b>B Gaussian Distribution.....</b>	<b>55</b>
<b>C Exploding reflectors.....</b>	<b>56</b>
<b>D Full waveform modelling.....</b>	<b>62</b>
<b>E Scientific Work .....</b>	<b>63</b>
<b>E.1 Informed proposal Monte Carlo</b>	
Sarouyeh Khoshkholgh, Andrea Zunino, Klaus Mosegaard	
<b>E.2 Full-waveform inversion by Informed-Proposal Monte Carlo</b>	
Sarouyeh Khoshkholgh, Andrea Zunino, Klaus Mosegaard	
<b>E.3 Evolution of the Stress and Strain field in the Tyra field during the Post-Chalk Deposition and Seismic Inversion of fault zone using Informed-Proposal Monte Carlo</b>	
Sarouyeh Khoshkholgh, Ivanka Orozova-Bekkevold, Klaus Mosegaard	
<b>E.4 Probabilistic Approach for Risk Assessment of CO<sub>2</sub> Storage</b>	
Ivanka Orozova-Bekkevold, Sarouyeh Khoshkholgh, Klaus Mosegaard	

## **E.5 Coupled multi-scale well and surface seismic data inversion using wavelet decomposition**

Sarouyeh Khoshkholgh, Andrea Zunino, Thomas Mejer Hansen, Klaus Mosegaard

# *Chapter 1*

## **Introduction**

Earth scientists have always been seeking new methods for investigating and exploring the Earth structure in order to obtain useful and practical information and produce reliable models. One main application is in oil and gas industry where a proper reservoir description is of key importance. This involves determination of spatial variability of reservoir rock properties and behaviour of the fluids in the reservoir. Seismic data inversion and seismic data integration has an important role in this matter. Earth model reconstruction from observed seismic data requires solving inverse problems and uncertainty analysis in subsurface modelling. Inverse problem methods are among the most widely used in geophysics and reservoir characterization studies for creating geological models. In an attempt to integrate all the knowledge available about the earth parameters including indirect or direct information, we built probabilistic inverse problems to perform an uncertainty analysis in desired areas. Complexity and non-linearity of forward models, high-dimensionality of model space and uncertainty related to prior knowledge or data are some of the challenges that probabilistic modelling methods face.

This thesis seeks to address some of the mentioned challenges. In this research, we seek to find a methodology for solving high-dimensional highly non-linear seismic inverse problems in an efficient way. In our research, we look at Markov Chain Monte Carlo algorithm when an informed proposal distribution is used to guide the search or sampling process and to improve the algorithm's efficiency. We'll begin with a brief overview of the general inverse problem and some common methods used for solving it.

## 1.1 Inverse Problem

Parametrization or describing the Earth as numbers is the very first step of inversion in geophysics. This process includes obtaining a group of model parameters  $\mathbf{m}$  through a mapping function  $f$  applied on Earth structure  $m$ . Here  $m$  is an abstract model and not yet parameterized.

$$\mathbf{m} = f(m) \tag{1.1}$$

Data can be defined as a set of numbers which is the result of interaction between our physical system and measuring tools and sometimes a third system have the task of measurement digitization. The forward mathematical relationship between observed data and physical parameters in a particular system is shown in the formula below. The inverse problem is the problem of achieving unmeasurable models  $\mathbf{m}$  from observed and measured data  $\mathbf{d}$  (Tarantola, 2005; Mosegaard and Tarantola, 1995; Mosegaard and Tarantola 2002).

$$\mathbf{d} = g(\mathbf{m}) \tag{1.2}$$

In deterministic approaches the mismatch between observed data and result from theoretical calculation is minimized and one final model is obtained that describes the best achievable model. By minimizing a misfit function some data fitting methods such as least square attempt to estimate the best match. The possible limitation of these methods are non-uniqueness of the solution and sensitivity to errors. The inverse problem is usually ill-posed in sense of Hadamard since it's dependence between the observable and the model may have lack of continuity, causing the solution to be unstable, or it happens that the solution may not exist or be non-unique (Hadamard, 1902; 1923). Methods such as regularization techniques estimate a well-posed problem and try to find the closest solution that fits the data by adding additional constraints. Tikhonov regularization is a typical regularization technique in which an extra expression is added to the least square mismatch function that controls the prior knowledge with a regularization parameter (Tikhonov, 1963).

For a proper analysis of an inverse problem, it is necessary to take into account possible measurement errors and uncertainty related to modelization or lack of data and information. In a probabilistic strategy a probability density is defined in data space that characterizes observed or measured information, and another probability density is described in model space that refers to our prior knowledge about model parameters which is completely independent from observed data.

Linear Gaussian inverse problem is the simplest form of Inverse problem where the forward function is linear.

$$\mathbf{d} = \mathbf{Gm} \tag{1.3}$$

The posterior probability density can be shown to be a Gaussian with the following mean and covariance (Tarantola 2005):

$$\mathbf{m}_{post} = \mathbf{m}_0 + (\mathbf{G}^T \mathbf{C}_n^{-1} \mathbf{G} + \mathbf{C}_m^{-1})^{-1} \mathbf{G}^T \mathbf{C}_n^{-1} (\mathbf{d} - \mathbf{G} \mathbf{m}_0) \quad (1.4)$$

$$\mathbf{C}_{post} = (\mathbf{G}^T \mathbf{C}_n^{-1} \mathbf{G} + \mathbf{C}_m^{-1})^{-1} \quad (1.5)$$

It is quite common in reality that we encounter nonlinear problems where the exact form of the posterior distribution is unknown and prior distribution is complicated. Bayes's theorem is an outcome of the definition of conditional probabilities described in the Kolmogoroff space (Box and Tiao 1992). According to Bayes's theorem the probabilities of an event can be updated by giving the occurrence of a relevant event. For instance the probability of event A giving event B is written as:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} \quad (1.6)$$

Where  $p(B|A)$  represents the probability of B happening when A is true, and  $p(A)$  and  $p(B)$  are marginal probabilities.

In Bayesian statistics the prior probability density is our approximate belief showing the probable parameter values before inspecting the data. The measured data is used in the likelihood function for computing probable parameter values that fit the data within the errors. These two distributions are integrated to create the posterior probability which does not always have an analytical form. Generating an ensemble of models that contain independent realizations from the posterior is one way of dealing with this issue (Mosegaard and Tarantola, 1995). In a probabilistic framework where  $d = g(\mathbf{m})$  defines our nonlinear forward problem, the joint prior probability density can be written:

$$\rho(\mathbf{d}, \mathbf{m}) = \rho_D(\mathbf{d})\rho_M(\mathbf{m}) \quad (1.7)$$

If  $\theta(\mathbf{d}, \mathbf{m})$  is a distribution in the joint data-model space describing the (exact or uncertain) correlation between  $\mathbf{d}$  and  $\mathbf{m}$  according to the physical laws, the conjunction of theoretical and experimental information according to (Tarantola, 2005) is:

$$\sigma(\mathbf{d}, \mathbf{m}) = k \frac{\rho(\mathbf{d}, \mathbf{m})\Theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})} \quad (1.8)$$

Where  $\mu$  is the homogenous probability density. The homogeneous probability distribution appoints a probability proportional to the volume for each region (Mosegaard, 2002). Considering the following assumption:

$$\mu(\mathbf{d}, \mathbf{m}) = \mu(\mathbf{d})\mu(\mathbf{m}) \quad (1.9)$$

The marginal posterior probability density in the model space can be shown to be:

$$\sigma_M(\mathbf{m}) = k\rho_M(\mathbf{m}) \int_D d\mathbf{d} \frac{\rho_D(\mathbf{d})\Theta(\mathbf{d}|\mathbf{m})}{\mu_D(\mathbf{d})} \quad (1.10)$$

Hence, the solution to the inverse problem is defined as a probability density function which is related to the product of the prior probability density and an integral named the *likelihood function*  $L(\mathbf{m})$ .

$$\sigma(\mathbf{m}) = k \rho_M(\mathbf{m})L(\mathbf{m}) \quad (1.11)$$

$L(\mathbf{m})$  represents the physical principles that connect model parameters and observed data. Figure 1 shows a one-dimensional example in which the conjunction of prior information and Likelihood function produce a posterior probability distribution which is the solution to the problem and represents the combination of the two states of information (Tarantola, 2005).

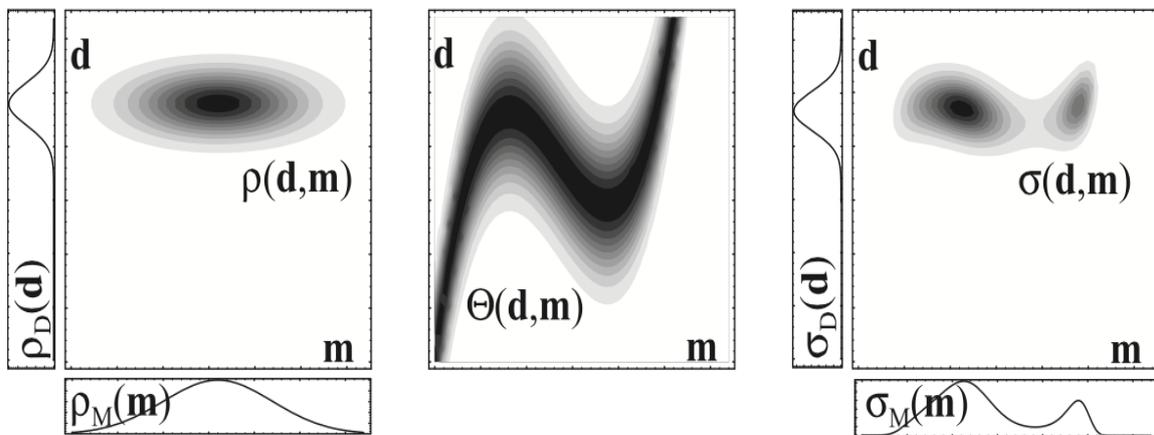


Figure 1.1.  $\rho_D(\mathbf{d})$  and  $\rho_M(\mathbf{m})$  are probability densities showing our knowledge about data and prior information respectively.  $\rho(\mathbf{d}, \mathbf{m})$  is the joint probability density and  $\Theta(\mathbf{d}, \mathbf{m})$  represents the physical theory and connection between  $\mathbf{d}$  and  $\mathbf{m}$ .  $\sigma(\mathbf{d}, \mathbf{m})$  is the solution of inverse problem and shows the conjunction of the two states of information (Tarantola, 2005).

## 1.2 Monte Carlo methods

When the physical relation between data space and model space is highly nonlinear, the posterior probability density function is far from Gaussian, and linearization methods are inadequate for achieving satisfactory results. The non-Gaussianity of the posterior could also come from non-Gaussianity of the noise

or the prior probability density (Mosegaard, 2006). As previously stated, obtaining the analytic expression of a probability density is not always feasible and practical, so we can instead sample it. Monte Carlo is recognized as being one of the most important methods for sampling a probability density by performing a random walk in the space. It has grown in popularity as computational technology has developed over time (starting with Yanovskaya, 1967; Press 1968). One major benefit of using Monte Carlo methods is that the analytical representation of the probability density is not required and all we need is the ability to calculate the function at a particular location (Mosegaard et al 2002). One of the first applications of Monte Carlo for investigating nonlinear problems was done by Koren et al. 1991 in which, for the first time, they introduced the movie strategy where a number of sample models are created and displayed to describe the posterior (Mosegaard, 2011). Monte Carlo methods can be categorized as two different groups:

1. Monte Carlo algorithms used for sampling purposes such as the Metropolis algorithm and Gibbs sampler.
2. Monte Carlo algorithms used for optimization purposes such as simulated annealing and genetic algorithms.

One of the first studies of Markov chain Monte Carlo simulation was conducted by Metropolis et al. (1953) where they used this sampling method to simulate thermodynamic equilibrium. Hastings (1970) made a generalization of Metropolis's previous method, later named the Metropolis-Hastings's algorithm. The Metropolis-Hastings algorithm is a well-known algorithm for sampling any kind of probability density function. In the Metropolis-Hastings algorithm we need to be able to compute the likelihood  $L(\mathbf{m})$  and the prior  $\rho_M(\mathbf{m})$  at any selected point in the model space. Then the algorithm will run as follows:

- I. Beginning from the current model  $\mathbf{m}_c$ , a new model  $\mathbf{m}_t$  is proposed using a uniform sampler A (the proposal generator).
- II. The proposed model is accepted with the following probability:

$$p_{acc} = \min\left(1, \frac{L(\mathbf{m}_t)\rho_M(\mathbf{m}_t)}{L(\mathbf{m}_c)\rho_M(\mathbf{m}_c)}\right)$$

- III. If the model is rejected we re-sample  $\mathbf{m}_c$
- IV. If it is accepted we replace  $\mathbf{m}_c$  by  $\mathbf{m}_t$  in the next step.

In the literature, there are many examples of using Metropolis algorithm for obtaining the posterior probability distribution (Pedersen and Knudsen, 1990; Koren et al, 1991; Gouveia and Scales, 1998; Dahl-Jensen et al, 1998; Khan et al, 2000; Khan and Mosegaard, 2001). When the prior is complex or far from uniform, the Metropolis algorithm can be computationally expensive (Hansen et al 2012). Mosegaard and Tarantola introduced extended Metropolis, a more powerful algorithm for non-linear inverse problems with complex priors. In chapter 2 the Extended Metropolis is explained.

Geman and Geman (1984) proposed a particular form of Metropolis-Hastings algorithm which was known as the Gibbs sampler (Gelfand et al 1990) and closely similar to the data augmentation algorithm by Tanner (1987). As explained by Mosegaard and Sambridge 2002, in the Gibbs sampler operates in the following way:

- I. The proposal distribution

$$U_k(x_i|x_j) = \begin{cases} \frac{p(x_i)}{\sum_{x_k \in \mathcal{N}_j^k} p(x_k)} & x_i \in \mathcal{N}_j^k \\ 0, & \text{otherwise,} \end{cases}$$

$\mathcal{N}_j^k$  is the group of points that deviates from  $x_j$

- II. The acceptance probability  $p_{acc} = 1$  for all proposed models.

We will not have any rejected models using this method; however, when the computational cost of calculating the target distribution  $p(x)$  is expensive, the Gibbs sampler is not a good option for solving the problem. Detecting essential (high-probability) locations of the probability density in a high dimensional space is the main challenge of the Monte Carlo sampling approach and any other inversion methods. There are several Monte Carlo-based methods that have been commonly used for optimization purposes: Simulated annealing, genetic algorithms and particle swarm methods are some examples of evolutionary algorithms used for optimization. Implementing ideas from statistical mechanics, simulated annealing is used to enhance the sampling procedure and detecting the global optimum of the posterior function (Kirkpatrick 1983). This is accomplished by increasing a system's temperature to a high level and gradually lowering it until it reaches the lowest energy state. If the energy difference between the current and the proposed states is negative in Monte Carlo simulated annealing, the move is accepted, otherwise, the acceptance probability is given by  $\exp\left(-\frac{\Delta E}{k_B T}\right)$ , where  $T$  is the temperature and  $k_B$  is the Boltzmann constant (Metropolis 1953).

Another example is Genetic Algorithms, which find the global minimum using ideas from biological evolution. Genetic Algorithms can be seen as a group of techniques rather than a single straightforward algorithm. This approach has been used to solve a variety of problems in earth science (Stoffa and Sen, 1991; Gallagher et al, 1991; Wilson and Vasudevan 1991; Sambridge and Drijkoningen, 1992; Scales et al, 1992; Sen and Stoffa 1992; Smith et al 1992).

As Markov Chain Monte Carlo became more and more popular in sampling posterior distributions and uncertainty analysis, new applications emerged with

larger computational challenges, such as inverse problems with computationally expensive forward calculations and complicated prior models. The main source of computational challenges in MCMC is the curse of dimensionality (Curtis and Lomax, 2001) which is related to the long time it takes for the algorithm to achieve convergence. There are plenty of different algorithms proposed for making MCMC sampling more efficient and practical (Liu 2002, MacKay 2003, Brooks et al 2005).

A newly developed method for dealing with computationally expensive Monte Carlo likelihood calculations is neural-network-based algorithms. Trained neural networks can estimate the hidden mapping function that connects two model spaces given a set of inputs and outputs to the algorithm. In order to train the algorithm a great number of randomly drawn realizations from the prior needs to be subject to forward calculations. By computing the forward response we will have the required training pairs. (Devilee et al. 1999 ;de Wit et al., 2013; Laloy et al. (2017); Laloy et al. (2018); Holm-Jensen and Hansen (2020); Earp & Curtis, 2020; Earp et al., 2020).

### **1.3 Blind algorithms**

These methods (Classical Monte Carlo, Simulated annealing, Genetic algorithms and Neural Networks) are categorized as blind algorithms where the characteristics of the target probability distribution are not considered in the sampling process. In blind algorithms just the input models and the output data from an optimization black box are considered in searching and sampling, and no additional information of the target distribution is used. Some blind algorithms try to detect the properties of the posterior distribution while in operation and sampling, but they are still bounded by the amount of information in the sample points they have visited.

According to No-Free-Lunch Theorem introduced by Wolpert et al 1997, the average efficiency of all blind optimization algorithms is exactly the same for all optimization problems. In other words, if a certain algorithm is more effective for a specific problem then it should not be true for other problems. Mosegaard 2012 specifies the limitation of blind inversions for large-scale nonlinear sampling or optimization problems and the essential role of tuning for each specific problem.

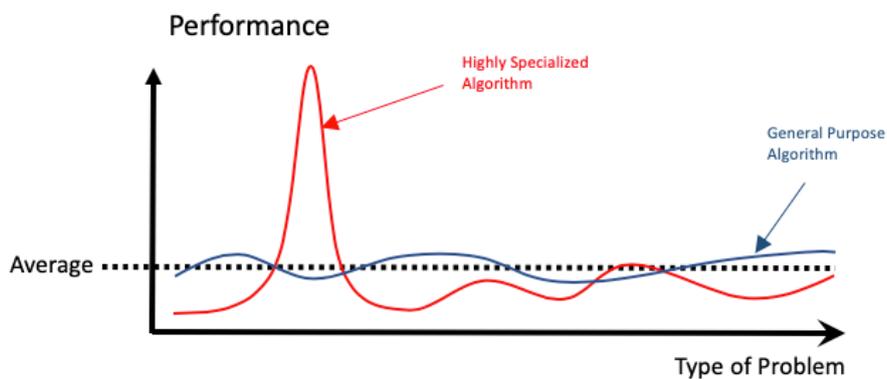


Figure1.2. The average efficiency of all optimization algorithms is similar when dealing with all available problems.

If external characteristics of an posterior probability density is taken into account as additional information guiding the searching process for a specific problem, then we will have an informed algorithm. According to No-Free-Lunch Theorem informed algorithms that consider known features of the selected target distribution in sampling should be more effective than blind algorithms. Hamiltonian Monte Carlo algorithm is a well-known example of informed algorithms (Fitchner et al., 2018; 2019; Gebraad et al., 2020). By obtaining information form the geometry of the model space and finding directions aligned with the iso-surfaces of high probability areas the algorithm aims at performing an efficient sampling. In Hamiltonian Monte Carlo the model is considered as a mechanical particle moving along a trajectory in the joint position-momentum

space which will be later projected back onto the position space, which is identical to our model space. The moving direction is controlled by the gradient of the misfit function which is described as the potential energy of the particle. Examples of applying Hamiltonian Monte Carlo in geophysics can be found in Fitchner et al. (2018, 2019) and Gebraad et al., (2020).

## **1.4 Our direction and contribution**

This Ph.D. thesis investigates a new probabilistic approach for solving large scale inverse problems that are strongly non-linear. In chapter 2 we explain and discuss the theory of Informed proposal Monte Carlo (IPMC), which will later be applied to a real case with the aim of increasing the speed of sampling and the efficiency of the algorithm. In IPMC we insert information obtained from a simplified physical theory of our problem into the algorithm in form of a global proposal distribution. The global proposal is gained through a simplified classical inversion and interpretation and it will only have an effect on the speed of convergence and the sampling process.

The forward modelling of our problem is described in chapter 3 along with the strategy for generating the prior model.

In chapter 4 we demonstrate how IPMC can be applied on a large-scale full waveform seismic problem. This methodology can be adopted for any large-scale non-linear problem when classical MCMC is extremely inefficient. However it should be noted that the efficiency depends on the quality of our proposal, or in the other words, it is problem dependent. In our case the proposal distribution is associated to an approximate posterior obtained via a simplified deterministic inversion, which determines the degree of efficiency as it approaches the true posterior.

Appendix E (E1-E5) consists of our scientific work associated with this PhD research, either published or under review. Appendix E1 is a research paper reviewed and published in Geophysical Journal International. Appendix E2 is submitted to Geophysical Journal International and currently under review. Appendix E3 is submitted to Applied Computing and Geosciences Journal and currently under review. Appendix E4 is a project report and the basis for Appendix E3. Appendix E5 is a conference abstract for the GEOSTATS conference in 2016.

# Chapter 2

## Monte Carlo with problem dependent proposal

### 2.1 Sampling the probability density

A direct analytical technique for solving inverse problems is not always applicable. In this case sampling the posterior probability density is the only way to describe the target distribution. In the probabilistic/Bayesian framework, information about model parameters is usually described by the posterior probability density with an unknown structure. For small inverse problems, exhaustive sampling can be implemented to visit every point in the model space. For larger problems, however, sampling is based on random model selection in the model space. This is known as importance sampling, and only models sampled from the prior density function and with an acceptable data fit will be considered in the sampling (Mosegaard 1998).

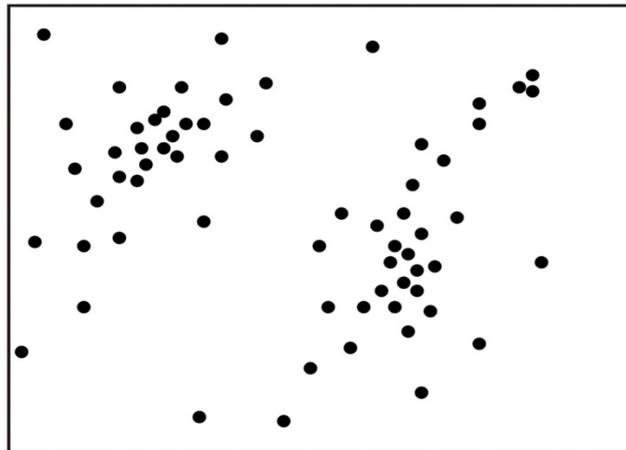
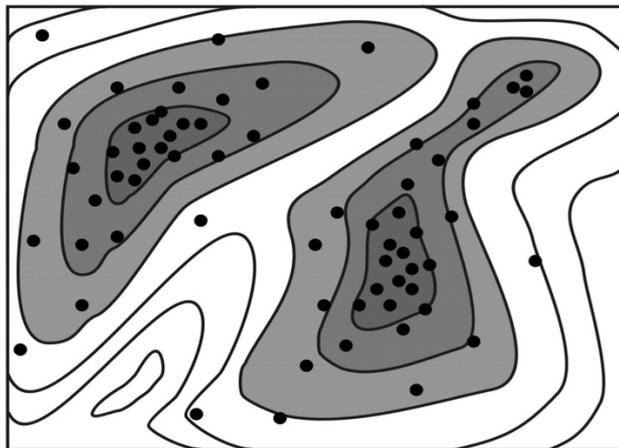
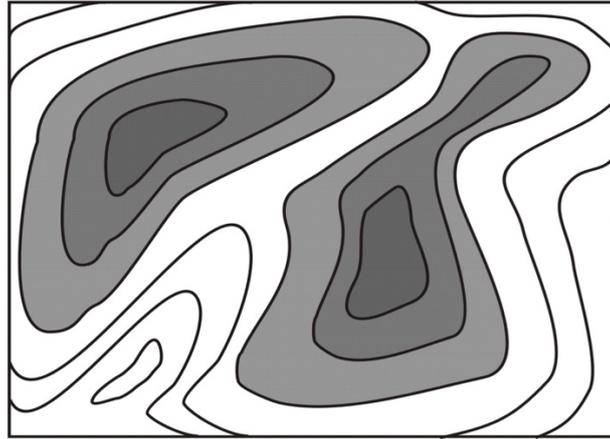


Figure 2.1. Complex probability densities can be represented by a collection of samples in the model space (Mosegaard, 2006)

## 2.2 Extended Metropolis

Markov chain Monte Carlo (MCMC) methods describe a stochastic sampling technique based on Markov processes where the prediction of future sample models is conditionally independent of the past and only depends on the current sample model. Extended metropolis is an MCMC-based algorithm that can be used for sampling of the posterior and consists of two separate algorithms  $V$  and  $U$ .  $V$  is a random iterative algorithm  $\mathbf{m}^{(n+1)} = V(\mathbf{m}^{(n)})$  that samples the prior probability density, whereas  $U(0,1)$  generates uniformly distributed random numbers (Mosegaard et al., 1995; Mosegaard, 2006). The updating algorithm  $W$  is defined as:

$$\mathbf{m}^{(n+1)} = W(\mathbf{m}^{(n)}) = \begin{cases} V(\mathbf{m}^{(n)}) & \text{if } U(0,1) \leq \min \left[ 1, \frac{L(V(\mathbf{m}^{(n)}))}{L(\mathbf{m}^{(n)})} \right] \\ \mathbf{m}^{(n)} & \text{else} \end{cases} \quad (2.1)$$

Which will sample the posterior probability density  $\sigma(\mathbf{m}) \propto L(\mathbf{m}) \rho(\mathbf{m})$ .

In this regard, the choice of  $V$  has a significant impact on the algorithm's performance. But because there is usually a large difference between the prior and the likelihood function, suggested models chosen by  $V$  often have a small likelihood value. Therefore, for large-scale problems, a blind extended Metropolis algorithm will not be so efficient (Mosegaard and Tarantola, 1995 ).

## 2.3 Proposal distribution

As previously stated, the Metropolis algorithm works by rejecting or accepting proposed models from the prior distribution. If  $f(x)$  is a posterior probability density and  $x$  is the current value, a proposed new model is offered from a

probability distribution known as proposal distribution  $q(x'|x)$ . This model will be accepted with the following probability (Metropolis et al., 1953; Hastings, 1970; Khoshkholgh et al., 2021):

$$P_{acc}^{x \rightarrow x'} = \min\left(\frac{f(x')q(x|x')}{f(x)q(x'|x)}, 1\right) \quad (2.2)$$

As mentioned by Mosegaard et al., 2002 any random walk should satisfy two conditions: *microscopic reversibility* meaning that the probability of entering a microscopic neighbourhood should be equal to the probability of leaving it, and *detailed balance* meaning that the probability of a transition from  $x_i$  to  $x_j$  is equal to the probability of transition from  $x_j$  to  $x_i$ . That is,

$$P_{acc}^{x \rightarrow x'} q(x'|x) f(x) = P_{acc}^{x' \rightarrow x} q(x|x') f(x') \quad (2.3)$$

The proposal distribution has no effect on the posterior distribution being sampled; but it is essential because it influences convergence time. The algorithm's search procedure is determined by the proposal probability density, and a proper proposal distribution should result in minimum settling time (burn-in time). The interval between sampling an initial low probability area and sampling regions with high probability model parameters is known as the burn-in time (Mosegaard, 2006). The proposal distribution can be divided into two categories: Local proposal distributions and global proposal distributions.

A local proposal distribution  $q(x'|x)$  is reliant on the starting point  $x$ . Under the two assumptions that  $q(x'|x) = q(x' + a|x + a)$  and  $q(x'|x) = q(x|x')$ , we can simplify the acceptance probability as follows:

$$P_{acc}^{x \rightarrow x'} = \min\left(\frac{f(x')}{f(x)}, 1\right) \quad (2.4)$$

The downside of using a local proposal distribution is that when it is very narrow and only allows small proposed steps, many iterations are needed for searching and sampling, and the ideal situation that the proposal is locally proportional to the target distribution,  $f(x') = C \cdot q(x'|x)$ , where  $C$  is constant, will rarely occur. Khoshkholgh et al. 2021 provides a more detailed explanation.

On the other hand, a *global* proposal distribution has no connection with the current starting point  $x$ , and can be defined as a constant probability distribution where  $q(x|x') = h(x)$ . If  $q(x'|x) = h(x')$  is closely similar to  $f(x')$  the Metropolis algorithm will be very efficient. If  $h(x')$  and  $f(x')$  are equal there will be no rejections and we will have the perfect situation. There are two different methods for finding global proposal distributions: The first approach is to interpolate the visited points to generate a local approximate proposal  $h(x)$  similar to  $f(x)$  in the vicinity of  $x$  (Christen 2005). However, such interpolation methods are constrained by the No-Free-Lunch Theorem because no external information is added and only already generated models are used in the sampling improvement.

The second approach is to use external information to generate a global proposal distribution which is not obtained from previously explored regions. This can be done by computing a rough version of the target distribution. The global proposal should be very much alike the target distribution everywhere even far from generated samples. In this thesis we propose a new approach to find a global approximation by using external information from simplified physics. By estimating a simple forward model and acquiring an approximate posterior, a global proposal distribution is obtained and used in the sampling process. In this

approach we keep the prior distribution and the likelihood the same as in the probabilistic/Bayesian framework.

## 2.4 Informed proposal Monte Carlo

If the problem is linear or Gaussian, the analytical form of posterior is known and can be used as a proposal. If the problem is nonlinear, there are two general situations. In the first scenario, we have external information about derivatives of the misfit and it is possible to calculate the derivative at each point. A known example that can be placed in this category is the Hamiltonian Monte Carlo method, where external information is used for coordinate transformation (Fitchner et al., 2018). A transformation in the parameter space is made where the transformed global proposal is close to the transformed target distribution. More details regarding how HMC should be understood in this framework can be found in Khoshkholgh et al., 2021.

In the second scenario, our external information is obtained from knowledge of the forward relationship between model parameters and data. The idea behind MCMC with a problem dependent proposal is to find an approximate posterior  $\tilde{f}(x)$  using the physics of the problem. In this way the approximate and the true posterior will, to some degree, be similar everywhere in the space or at least in most places where probabilities are significant. Figure 2 shows a simple example of such situation where we have a) a true posterior distribution and b) an approximate posterior which is used as a global proposal distribution. The approximate posterior is used for guiding and accelerating the sampling process therefore increasing the acceptance rate and having a more efficient algorithm.

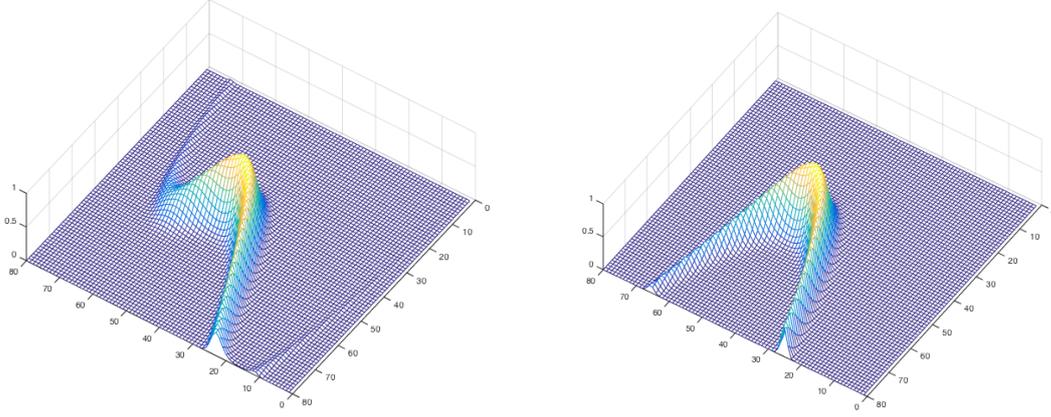


Figure 2.2. a) The true posterior distribution b) approximate posterior

Assume now that, in equation 1.1, the homogeneous probability density and the marginal prior are assumed to be constant. This gives

$$\sigma(\mathbf{d}, \mathbf{m}) = k \cdot \rho(\mathbf{d}) \theta(\mathbf{d}, \mathbf{m}) \quad (2.5)$$

We assume that  $\rho(\mathbf{d})$  is only nonzero close to  $\mathbf{d}_{obs}$  or, in the other words, modelization errors, described by  $\theta(\mathbf{d}, \mathbf{m})$ , are much larger than observational data errors. Then we can rewrite the equation as follows:

$$\sigma_m(\mathbf{m}) \propto \sigma(\mathbf{d}_{obs}, \mathbf{m}) \approx \theta(\mathbf{d}_{obs}, \mathbf{m}) \quad (2.6)$$

If we find an approximate solution  $\tilde{\mathbf{m}}$  by performing a simplified deterministic inversion, the true modelization error could be obtained if we had the true solution. But since we don't have the true solution, we create a new inverse problem, similar to the original one, that has  $\tilde{\mathbf{m}}$  as the true solution.

We propose the following method: First a simplified version of the physics behind the problem is determined and used to create an approximate forward  $\tilde{g}(x)$  which would provide a simple, but inaccurate inversion. Second, an

approximate solution  $\tilde{\mathbf{m}} = h(\mathbf{d}_{obs})$  to this simplified inversion is obtained. Third, we evaluate the modelization error by replacing the true model with the obtained approximate solution  $\tilde{\mathbf{m}}$ . Now, instead of computing the true modelization error

$$\delta\mathbf{m}_{true} = \tilde{\mathbf{m}} - \mathbf{m}_{true}, \quad (2.7)$$

which is impossible because we do not have  $\mathbf{m}_{true}$ , we compute the approximation

$$\delta\mathbf{m}_{approx} = \tilde{\mathbf{m}} - h(g(\tilde{\mathbf{m}})). \quad (2.8)$$

The more similar  $\tilde{\mathbf{m}}$  and  $\mathbf{m}_{true}$  are, the closer  $\delta\mathbf{m}_{approx}$  and  $\delta\mathbf{m}_{true}$ . The approximate solution  $\tilde{\mathbf{m}}$  and the components of the approximate modelization error vector  $\delta\mathbf{m}_{approx}$  are used to create a modelization error distribution  $\tilde{\theta}(\mathbf{d}_{obs}, \mathbf{m})$  in a functional form centered at  $\tilde{\mathbf{m}}$ . One simple possibility is to use an isotropic Gaussian with  $\tilde{\mathbf{m}}$  as the mean, and the components of  $\delta\mathbf{m}_{approx}$  as standard deviations. The approximate modelization error distribution can be used as the proposal distribution:

$$q(\mathbf{m}'|\mathbf{m}) = \tilde{\theta}(\mathbf{d}_{obs}, \mathbf{m}) \quad (2.9)$$

Figure 5 shows the true acoustic impedance and the approximated acoustic impedance for a 1D, 1000-parameter inverse scattering problem. Figure 6 compares the true model envelope and the approximate model error envelope calculated from (2.8).

For this one-dimensional example the proposal distribution can be built using an isotropic Gaussian with  $\tilde{\mathbf{m}}$  as the mean, and squared elements of the envelope function as the diagonal of a covariance matrix. A complete example of applying IPMC to a 1D case can be found in Khoshkholgh et al. 2021. The essential role

of using a proposal distribution is to increase the frequency of independent models being accepted without affecting the prior information used in the calculations.

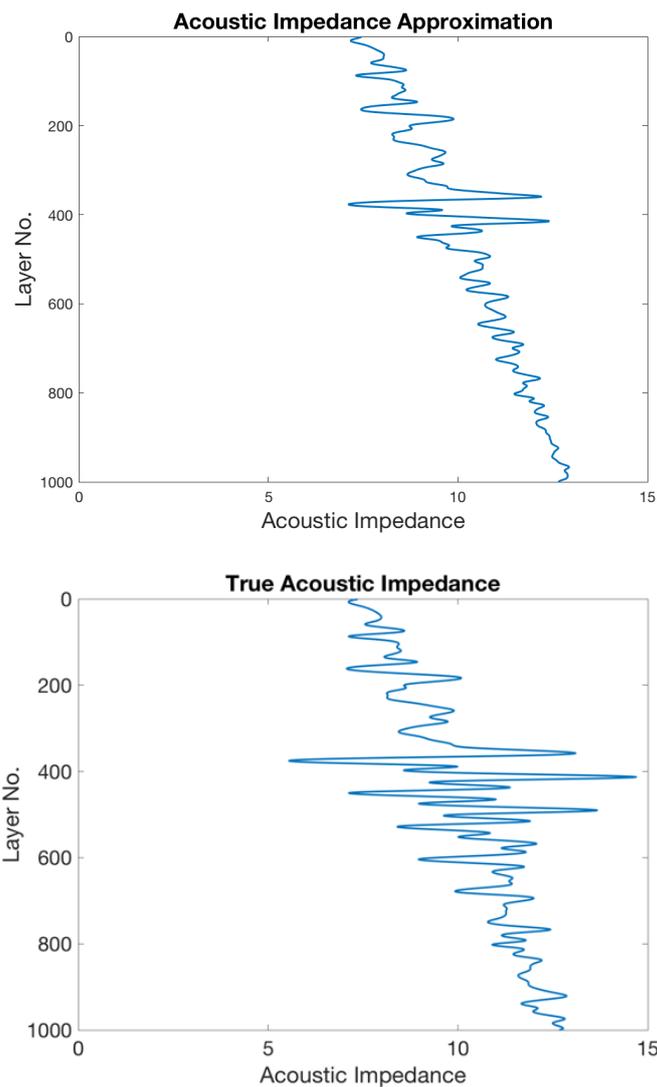


Figure 2.3. The true acoustic impedance and approximated acoustic impedance

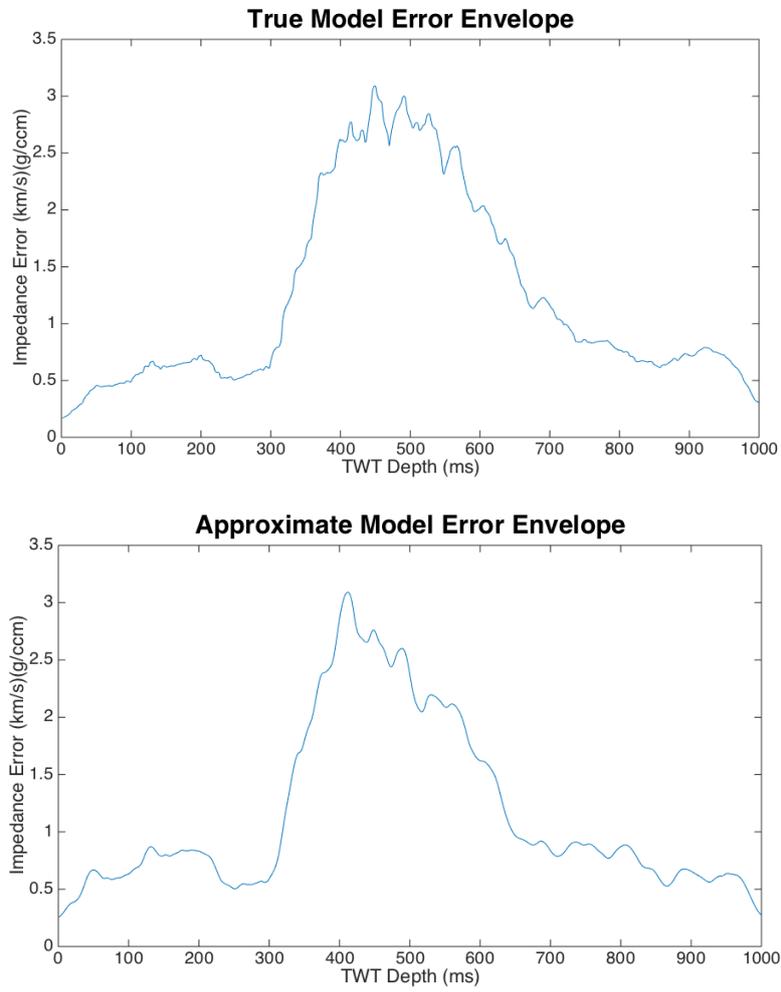


Figure 2.4. The true model error envelope and the approximate model error envelope

# Chapter 3

## Forward modelling and prior computation

### 3.1 Forward modelling

In reality, it often happens that certain variables cannot be directly measured or observed. Data is the information we observe and can be collected as the outcome of some physical experiments. The physical laws that show the connection between measured data and model parameters are known as forward models. When dealing with geophysical inverse problems, understanding and implementing forward calculations is necessary. As stated previously, the misfit is calculated by subtracting observed and predicted data. Physical theory provides the forward operator  $g$  or a computer modelling that must be applied to the model parameters in order to produce predicted values (Tarantola 2005):

$$d = g(m) \tag{3.1}$$

Seismic data is generated by sending pressure waves originating from a seismic source into the subsurface, causing energy to propagate. Energy is reflected at boundaries where there is contrast in rock properties, and the arrival time is

recorded by receivers at the surface (Sheriff, 1995). The most accurate and complete way to generate a seismic data set is through full waveform modelling. A short introduction to full waveform modelling including elastic and acoustic wave equations and the mathematical procedure for finite difference modelling of acoustic waves is provided in appendix D.

### **3.2 Prior computation**

In a probabilistic/Bayesian framework, our previous knowledge and information about the model parameters that is completely independent from observed data, and is called prior information. This information is formed as a probability distribution and shows expected models of the subsurface from previous experience or professional's estimation. In the Bayesian approach, choosing the prior model has been a heavily debated topic (Jaynes, 1985; Scales, 1997; Mosegaard, 2011). Some believe that the prior information is even more important than information from observed data (Journel, 1994; Jaynes, 1984). Some believe in the probabilistic point of view proposed by Tarantola and Valette but they suggest that the prior information should be noninformative in order to avoid bias in the solution (Buland and More, 2003; Khan and Mosegaard, 2002). Some support the idea that the prior information has its own value and importance and should be considered as an independent piece of information (Hansen et al, 2012; Hansen et al, 2016).

In this section we use a piecewise velocity model of the subsurface as an approximate solution to the inverse problem. The aim is to adopt a type of prior probability that would preserve the main topology of the model and at the same time permit large variations in the model. Our prior probability density appoints non-zero probability to layered velocity models where each layer is homogenous

with constant velocity values. In other words, the prior model has the topology of the layering found in the approximate solution. After randomly selecting a window inside the image, a warping strategy is used for perturbing the selected area. In the following, the mathematics behind the warping procedure is explained (Khoshkholgh et al., 2021).

### 3.2.1 Image Warping

A topology-preserving, continuous transformation of an image is called image warping and it can be defined as a redefinition in coordinates of each point in the image. There are two different procedures in image warping: A transformation that defined the new positions for each pixel and interpolation that calculates their new values. We have a two-dimensional function  $f(x_1, x_2)$  that describes our image where  $f$  is the velocity value and  $x_1, x_2$  are horizontal and vertical directions.

A displacement field is an image with vectors at each pixel and each vector defines the displacement between the original image and the warped image. Displacement field show how individual pixels in the image will be moved to a new point. Warping is described by a vector displacement field  $u$ ,  $\left| \frac{\partial u_i}{\partial x_j} \right| < 1$  where  $x_1, x_2$  defines image coordinate position. If a random square window is chosen inside the image, the displacement coordinates can, e.g., be calculated in two directions using cosine variations. The displacements are zero at the window boundaries. The displacement size is random such that the maximum displacement is chosen from the modelization error. The direction of the displacement field is random. The size of the window is pre-defined and zero at the boundaries (Khoshkholgh et al., 2021).

### 3.2.2 Producing samples from the prior

As explained in chapter 2, in IPMC our algorithm, a global proposal could be an approximate posterior that is obtained by simplified physics. For a full-waveform inversion problem, a layered approximate velocity model can be generated by classical signal processing, deterministic approaches, and/or interpretation. If we generate accurate synthetic data from this approximate velocity model, and invert the synthetic data with a simplified inverse (or an interpretation procedure), we will have a modelization error matrix containing a velocity difference at each point in the model. An envelope of these velocity differences can now be generated by using a Hilbert transformation, and the matrix of envelope values determines the maximum of (or standard deviation of) the velocity perturbations at each point, and hence also the maximum warping displacement. In order to generate samples from the prior we can randomly choose between perturbing the velocity value at each layer or the warping procedure. Figure 6 shows seven different realizations from the prior.

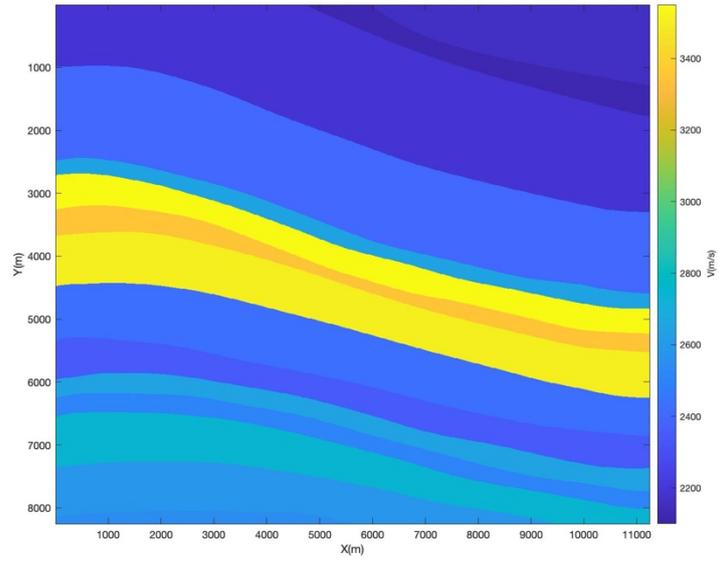
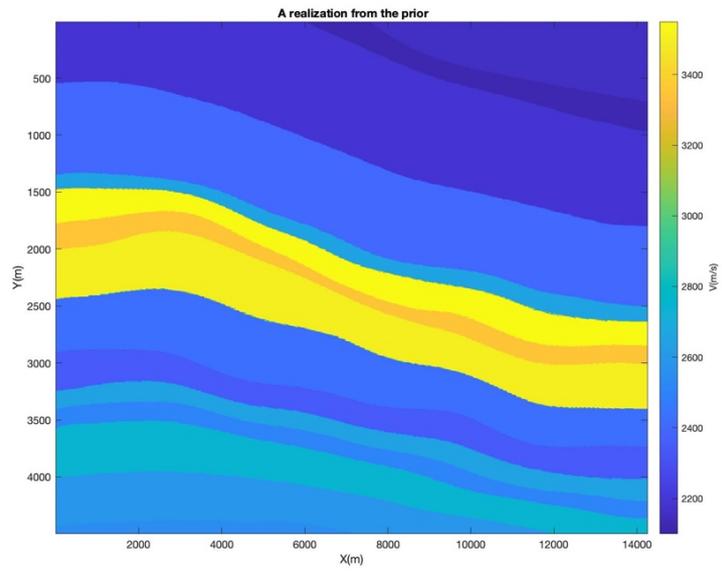
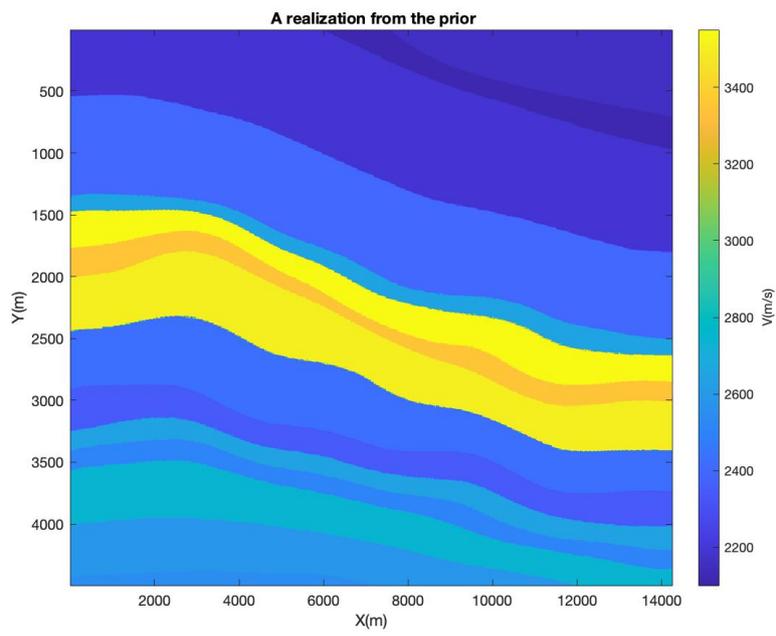
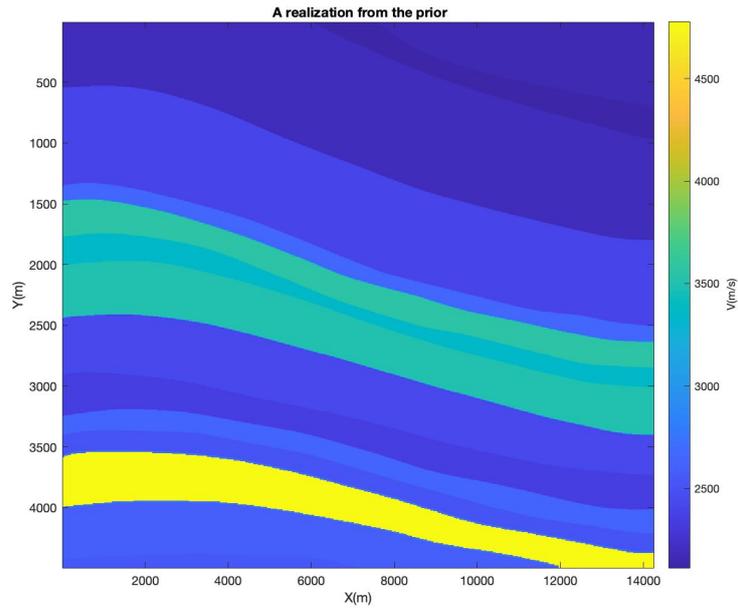
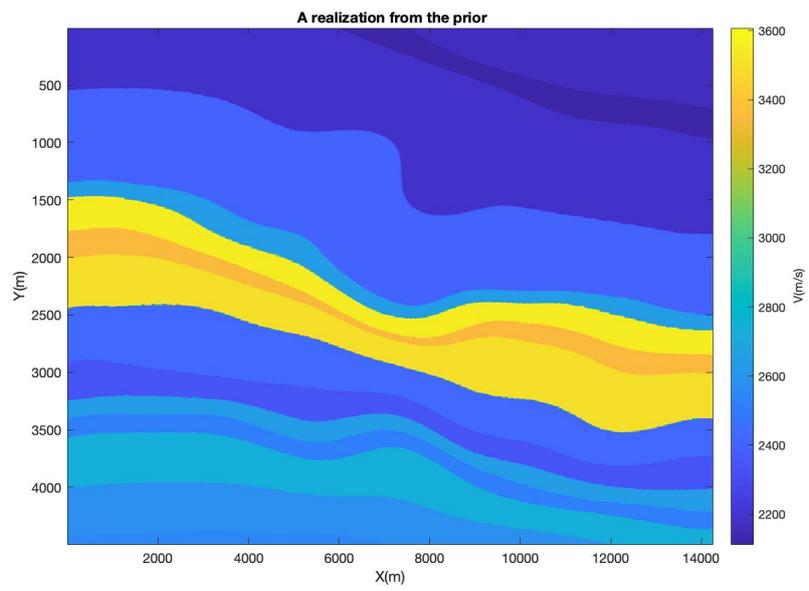
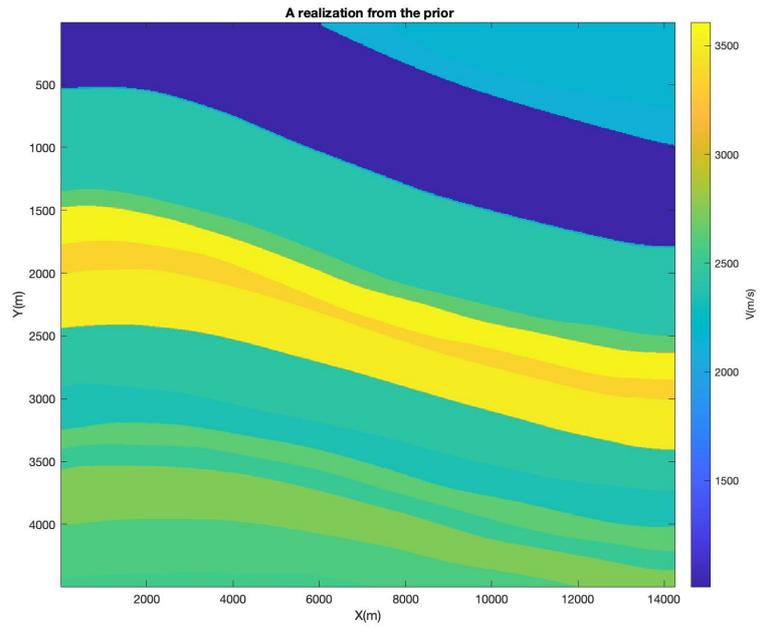


Figure 3.1. Approximate velocity model







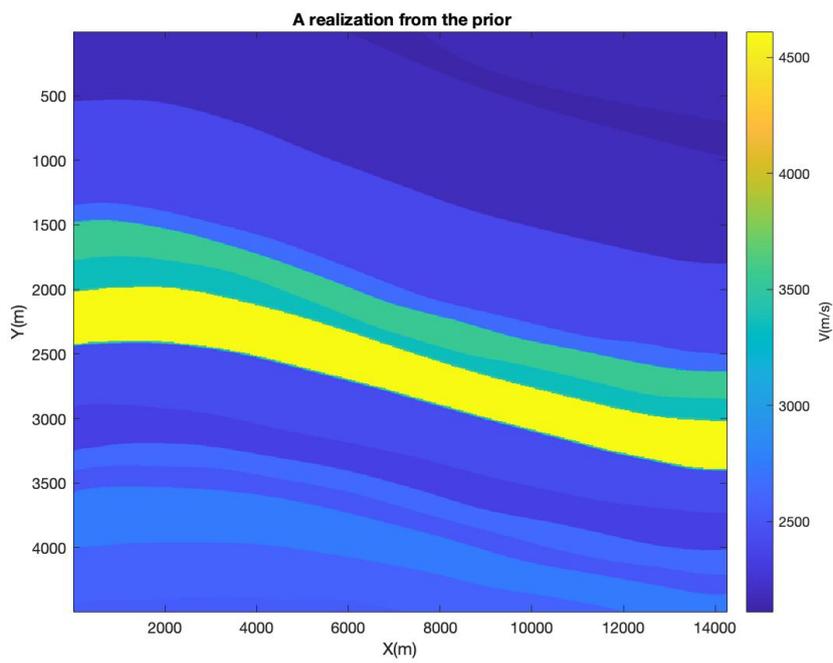
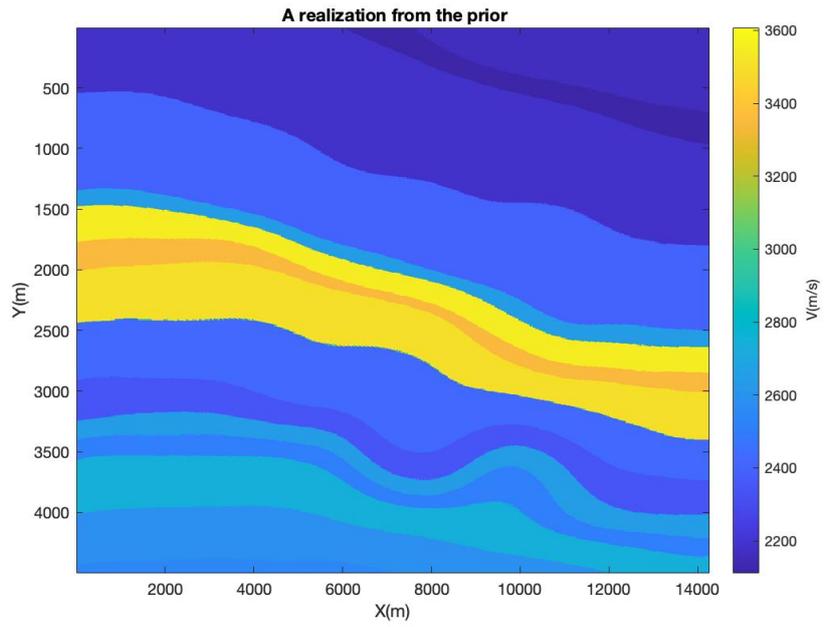


Figure 3.2. Some realization from the prior

## Chapter 4

### **Full-waveform inversion using Informed proposal MCMC**

Full-waveform inversion is a well-known technique for producing high-resolution models that present physical properties of the subsurface. In FWI The full content of seismic data is used to create models that are able to fit observed data in a chosen area. Initial studies in FWI tried to look at this problem as an optimization procedure and they aim at solving it by minimizing the misfit in an iterative approach (Lailly 1983, Tarantola 1984). A full review of FWI that highlights least-squares optimization with elastic and acoustic examples is given by Virieus and Operto 2009. Since this problem is highly non-linear with extreme multimodality in the posterior, optimization approaches cannot supply proper uncertainty analysis for evaluating and understanding all aspects of the outcome of FWI problem.

Applying Markov chain Monte Carlo methods for FWI provides uncertainty analysis from the samples generated from posterior distribution. In MCMC the form and mathematical description of the posterior distribution is not needed and it is, in principle, easy to deal with complex multi-modal posterior distributions. However, in a high-dimensional space and for such computationally expensive forward calculations as we have in FWI, regular MCMC is not efficient and practical. Here we explore a couple of methods available for handling full-waveform inversion where we have large scale, highly nonlinear probabilistic problems.

Variational inference for full-waveform inversion is a non-sampling method first proposed by Nawaz and Curtis 2018. Zhang et al 2020 applied the method to full-waveform inversion and named it VFWI. In this approach, a family of variational functions is determined and through optimization the algorithm searches for the best approximate posterior. This can be done by minimizing the Kullback-Leibler divergence (Kullback & Leibler 1951) iteratively to find the best approximate posterior.

Hamiltonian Monte Carlo is another successful approach applied to FWI problem (Fichtner et al. 2018, Gebraad et al. 2020). In these methods our knowledge and information about the characteristics of the problem plays an important role.

#### **4.1 Sampling with an Informed proposal**

In this section we will describe our application of IPMC to a 2D full-waveform, seismic inverse problem in the acoustic approximation (Khoshkholgh et al., 2021). The problem has  $\sim 10^6$  parameters, but it illustrates how a combination of an informed proposal and prior information can make such a problem computationally tractable.

We use an Extended Metropolis-Hastings algorithm (Mosegaard and Tarantola, 1995) where we sample the posterior in the following way:

1. A model from proposal distribution  $q(\mathbf{m}'|\mathbf{m})$  is chosen and proposed to perturb the current model  $\mathbf{m}$ .
2. This new model  $\mathbf{m}'$  will be either accepted or rejected with the following probability:

$$P_{acc}^{\rho} = \min\left(\frac{\rho(\mathbf{m}')q(\mathbf{m}|\mathbf{m}')}{\rho(\mathbf{m})q(\mathbf{m}'|\mathbf{m})}, 1\right) \quad (4.2)$$

If  $\mathbf{m}'$  gets rejected, we repeat  $\mathbf{m}$  (the current model).

3. If  $\mathbf{m}'$  is accepted in step 2, another acceptance probability generated from the misfit function (likelihood) determines if  $\mathbf{m}'$  should be accepted or rejected.

$$P_{acc}^L = \min\left(\frac{L(\mathbf{m}')}{L(\mathbf{m})}, 1\right) \quad (4.3)$$

If  $\mathbf{m}'$  gets rejected,  $\mathbf{m}$  will be repeated again.

## 4.2 IPMC procedure for sampling the posterior probability in full-waveform inversion

We use 2D data from part of the Marmousi velocity model (Versteeg, 1994). The overall procedure can be expressed as follows:

- 2D synthetic seismic data is generated by using the exploding reflector model that approximates zero offset data.

- Depth migration and deconvolution is done for obtaining a reflectivity model.
- From the reflectivity, a high resolution velocity model is produced and added to the background velocity.
- Interpreting the visible reflectors, a homogeneous-layer approximate velocity model is obtained that is an approximation to the true velocity model.
- Synthetic data are calculated from the approximate model, using accurate finite-difference wave simulation.
- The synthetic data is now inverted and a second approximation is obtained.
- The difference between the first and the second approximation is an estimation of the modeling error.
- Using a Hilbert transform, we can find the model error envelope and the modelization error distribution that informs us about a reasonable standard deviation of the perturbation at each point.
- IPMC now starts from the approximate model, using the modelization error distribution as global proposal.
- In 30 percent of the perturbations, a random layer in the model is chosen and the velocity at all the points inside the layer are perturbed.
- In the rest of the perturbations, the layer boundaries are perturbed by warping the model, whereas the velocity value in each layer is constant.

Looking at the results from IPMC inversion, we notice a considerable gain in efficiency and reduction in convergence time. This is noticeable when we compare result from the IPMC algorithm to those of the classical Extended Metropolis algorithm. It should be noted that noise variance has effect on convergence time in both of the algorithms, and also the fact that more data would influence the sampling speed and results in more constraint solution.



# Chapter 5

## **Discussion**

Inversion of seismic data for providing reasonable models of the subsurface has been widely used in geophysics. A probabilistic framework for solving an inverse problem is usually more practical and trustworthy than deterministic approaches due to ill-posedness and bias. Monte Carlo sampling methods are widely used for obtaining solutions to nonlinear probabilistic problems, but the challenges of using MCMC in highly-nonlinear high-dimensional problems have been a topic of research in different research areas, and many algorithms have been proposed for efficient sampling through MCMC. We should keep in mind that for a fair comparison between different algorithms we should consider all the information provided by the problem and all numerical operations performed in the algorithm.

In the following the challenges and problems in applying and comparing MCMC to different problems is discussed and the advantages of using IPMC is explained.

It happens quite often that we see comparisons in the literature where different algorithms are considered without making sure that they are solving the exact same problem. When comparing the efficiency of different algorithms it is important to consider the same problem, and to use the total number of numerical operations as a measure of the computational workload. For example, when comparing Machine learning algorithms to other algorithms, the computational part of the training process must be included when evaluating the efficiency.

Another example is in gradient-based methods when the numerical operations needed to compute the gradients are not considered. It should also be noted that a derivative-based algorithm could be very fast on a smooth problem but meaningless for a non-smooth case. For example in case of a geostatistical problem with categorical parameters. This highlights that we should know there is no general, practical MCMC sampler that could be efficient for all problems.

A consequence of the No Free Lunch Theorem is that an informed algorithm that employs external knowledge about characteristics of the target distribution in order to navigate the search or sampling procedure is more efficient than a blind algorithm without any additional information. For any inversion algorithm the posterior distribution has a fixed amount of information but in addition to this information we might have external knowledge about the characteristics of the posterior such as smoothness, maximum value etc. Using this additional information results in better efficiency, as measured by the number of numerical operations needed to generate a specific number of independent models. This statement is true for both deterministic (optimization based) and sampling based methods.

It is clear that if external information about the target distribution is used then the sampling process will gain efficiency. In our methodology this information is introduced as a global proposal. In the IPMC approach a global proposal is obtained from the approximate physics of the problem. It is important to note that by using this method we are not just focusing to sample a given area of the model space. But the proposal is arranged such that the areas which are considered more probable by the approximation are being searched. This will not generate any bias asymptotically because when a model with high probability is proposed the acceptance probability will be decreased. If the sampler is paused sooner than expected we would notice that it has mostly explored the regions where the proposal distribution has higher value. If we continue sampling for a considerable amount of time we will see that the model space will be explored everywhere. This is typically not the case for simple MCMC because if the algorithm is stopped earlier, then the area around the arbitrary starting point may be over-represented, even if it does not properly represent high-probability areas.

There are some algorithms and practical implementations in the literature where they use approximate calculations in order to gain efficiency. For example in the neighbourhood algorithm (Sambridge 1999) the likelihood value is approximated by using the computed values of adjacent models and thus misfit and forward calculations are replaced by a nearest neighbourhood approximation. Lekic and Romanowicz 2011 proposed a method that uses approximate techniques in waveform tomography for imaging the Earth structure. They develop a hybrid approach that uses spectral element method for modelling waveforms in upper mantle. It should be pointed out that the performance of such algorithms depends directly on the quality of the approximations being made and if an approximate operation is not properly chosen it will produce an inaccurate posterior and false results.

However a new aspect in our approach is that we use the approximation but take it out again to avoid bias. And even if misleading and inaccurate information is injected as the approximate proposal, it will not distort the posterior in the long run. This is not the case for a prior or a likelihood where misinformation will mislead the posterior. Our proposal is obtained from the physics of the problem and represents important structure of the posterior. New additional information is being injected to construct an approximation that makes it different from other algorithms whose information comes from samples of the posterior density alone.

We believe that our methodology can cause significant improvement in any available algorithms by augmenting them with an informed proposal. One could argue that the external information can augment the prior instead of being used as a proposal. The reason we introduce the external information only via the proposal is that the overlap between the external information and information already represented by the posterior will not be a problem and this could have significant practical value. Furthermore, in the worst case, even an oversimplified (poor) external information introduced by a proposal will only make the algorithm slower, but not introduce any bias.

By implementing the methodology described in this thesis we can augment an existing blind sampler with an informed proposal which is built from external knowledge and we can turn the sampler into a specialized algorithm directed towards sampling our particular target distribution. In this way we obtain a single purpose algorithm that will probably function well in our specific problem but most likely fail in other cases. This approach can be applicable to problems with both smooth or non-smooth posterior distribution.



# Chapter 6

## **Conclusion**

This thesis explores some challenges and limitations in probabilistic inverse problems in geoscience. Practical issues in uncertainty quantification of the subsurface, including the lack of efficiency, non-linearity, high-dimensionality and high computational cost have been addressed.

In this study we focus on the case of using Markov Chain Monte Carlo for solving inverse problems and sampling the posterior distribution. We have investigated the effect of using external information in form of a proposal distribution in order to make MCMC algorithm considerably more efficient. This external information presents the properties and structure of the target distribution. The evidence from this study suggest that by using simplified physics of a specific problem we can build an approximate posterior and use it as a global proposal to provide consistency between the proposal mechanism and the posterior distribution.

The suggested methodology enables the MCMC algorithm to complete the convergence process faster and the proposed models in sampling process will be chosen from regions with higher probabilities. Another significance of this method lies in the fact that even a poor approximate proposal will not distort or mislead the algorithm asymptotically, it will just affect the computation time.

In a numerical example provided by this study we use an approximate posterior obtained by processing and interpretation of seismic data. This approximate model is used for evaluating the modelization error and then we use it to make a global proposal distribution. We have managed to significantly increase the efficiency of MCMC sampling strategy by using informed global proposal in our inverse problem.

It should be noted that the algorithm performance using IPMC depends on the external proposal selected for the problem. The more similarity between the target distribution and the proposal, the higher efficiency of the algorithm. The proposed methodology has a potential to be applied to any type of MCMC sampling algorithm as long as we are able to form a proposal probability density that can approximate the posterior properly.



## References

- Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM, 163–169.
- Tarantola, A., and B. Valette, 1982a, Inverse Problems = Quest for Information: *Journal of Geophysics*, 50, 159–170.
- Tarantola, A., and B. Valette, 1982b, Generalized nonlinear inverse problems solved using the least-squares criterion: *Reviews of Geophysics and Space Physics*, 20, 2, 219–232, doi: [10.1029/RG020i002p00219](https://doi.org/10.1029/RG020i002p00219).
- Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: *Journal of Geophysical Research*, 100, no. B7, 12,431–12,447, doi: [10.1029/94JB03097](https://doi.org/10.1029/94JB03097).
- Mosegaard, K., and A. Tarantola, 2002, Probabilistic Approach to Inverse Problems: Chapter for the *International Handbook of Earthquake and Engineering Seismology: Academic Press*, 237–265. 0-12-440652-1. Published for the International Association of Seismology and Physics of the Earth Interior (IASPEI).
- Mosegaard, K., and M. Sambridge, 2002, Monte Carlo analysis of inverse problems: *Inverse Problems*, 18, R29–R54, doi: [10.1088/0266-5611/18/3/201](https://doi.org/10.1088/0266-5611/18/3/201).
- Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: *Inverse Problems*, 14, 405–426, doi: [10.1088/0266-5611/14/3/004](https://doi.org/10.1088/0266-5611/14/3/004).
- Mosegaard, K., 2012. Limits to Nonlinear Inversion. in: *Applied Parallel and Scientific Computing. Proceedings of the PARA 2010 Meeting: Revised Selected Papers, Part I*. Springer, pp. 11-21. *Lecture Notes in Computer Science*, no. Part 1, vol. 7133
- Mosegaard, K., 2011. Quest for consistency, symmetry and simplicity—the legacy of Albert Tarantola. *Geophysics* 76, W51–W61
- Mosegaard, K., 2006. *Monte Carlo Analysis of Inverse Problems*. University of Copenhagen, Copenhagen

- Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2012, Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling: Computational Geosciences, doi: [10.1007/s/10596-011-9271-1](https://doi.org/10.1007/s/10596-011-9271-1).
- N. Tikhonov (1963), 'Solution of incorrectly formulated problems and the regularization method', Soviet Meth. Dokl. 4, 1035–1038.
- Tikhonov, A. N., 1963, Resolution of ill-posed problems and the regularization method (in Russian): Doklady Akademii Nauk SSSR, 151, 501–504.
- Hadamard (1902), 'Sur les problèmes aux dérivées partielles et leur signification physique', Princeton University Bulletin 13, 49–52.
- Hadamard (1923), Lectures on Cauchy's Problem in Linear Partial Differential Equations, Yale University Press.
- Hastings W K 1970 Monte Carlo sampling methods using Markov Chain and their applications *Biometrika* **57** 97–109
- Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984)
- Box, George E. P, and George C Tiao. *Bayesian Inference in Statistical Analysis*. Hoboken: John Wiley & Sons, Incorporated, 1992. Print.
- Metropolis N, Rosenbluth M N, Rosenbluth A W, Teller A H and Teller E 1953 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087–92.
- Kirkpatrick S C, Gelatt D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80.
- Gelfand, A. E., and A. F. M. Smith, Sampling based approaches to calculating marginal densities, *J. Am. Stat. Assoc.*, 85, 398–409, 1990.
- Wolpert D. H., and Macready W. G., 1997. No free lunch theorems for optimization: IEEE Transactions on Evolutionary Computation 1, p 67-82
- Fichtner, A., Zunino, A., and Gebraad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems: Geophysical Journal International, 216, p 1344-1363.

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

Gebraad, L., Boehm, C., and Fichtner, A., 2020. Bayesian elastic fullwaveform inversion using Hamiltonian Monte Carlo. *Journal of Geophysical Research: Solid Earth*, 125, e2019JB018428.  
<https://doi.org/10.1029/2019JB018428>

Liu, J.S., 2002. *Monte Carlo Strategies in Scientific Computing*: Springer. ISBN-13: 978-0387952307.

MacKay, D., 2003. *Information Theory, Inference and Learning Algorithms*: Cambridge University Press, ISBN-13: 978-0521642989.

Brooks, S., Gelman, A., Jones, G. and Meng, X.-L., 2011. *Handbook of Markov chain Monte Carlo*: CRC Press.

Laloy, E., Hraut, R., Lee, J., Jacques, D., and Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network: *Advances in Water Resources*, 110, p 387-405.

Laloy, E., Hraut, R., Jacques, D., and Linde, N., 2018. Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network: *Water Resources Research* 54, p 381-406,  
[doi.org/10.1002/2017WR022148](https://doi.org/10.1002/2017WR022148)

de Wit, R. W., Valentine, A. P., & Trampert, J. (2013). Bayesian inference of earth's radial seismic structure from body-wave traveltimes using neural networks. *Geophysical Journal International*, 195(1), 408–422.

Devilee, R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems: Inverting surface wave velocities for eurasian crustal thickness. *Journal of Geophysical Research: Solid Earth*, 104(B12), 28841–28857.

Earp, S., & Curtis, A. (2020). Probabilistic neural network-based 2d travel-time tomography. *Neural Computing and Applications*, 32(22), 17077–17095.

Earp, S., Curtis, A., Zhang, X., & Hansteen, F. (2020). Probabilistic neural network tomography across grane field (north sea) from surface wave dispersion data. *Geophysical Journal International*, 223(3), 1741–1757.

Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: *Journal of Geophysical Research*, 96, no. B12, doi: [10.1029/91JB02278](https://doi.org/10.1029/91JB02278).

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

Sheriff, R., Geldart, L., 1995. *Exploration Seismology* (2nd). Cambridge University Press.

Wang, Y., 2015, Frequencies of the Ricker wavelet, *GEOPHYSICS*, VOL. 80, NO. ;P. A31–A37.

Ricker, N., 1943, Further developments in the wavelet theory of seismogram structure: *Bulletin of the Seismological Society of America*, 33, 197–228.

Ricker, N., 1944, Wavelet functions and their polynomials: *Geophysics*, 9, 314–323, doi: [10.1190/1.1445082](https://doi.org/10.1190/1.1445082).

Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 6th edition, 2007. ISBN 978-0-13-187715-3.

Jaynes, E.: Highly informative priors. In: Bernardo, J., DeGroot, M., Lindley, D., Smith, A. (eds.) *Bayesian Statistics*, vol. 2, pp. 329–360. Elsevier, Amsterdam (1985)

Scales, J., Sneider, R.: To Bayes or not to Bayes? *Geophysics* **62**(4), 1045–1046 (1997)

Journel, A. G. (1994), Modeling uncertainty: Some conceptual thoughts, in *Geostatistics for the next Century*, Springer, New York, pp. 30–43.

Jaynes, E. T. (1984), Prior information and ambiguity in inverse problems, *Inverse Problems*, 14, 151–166.

Khan, A., Mosegaard, K.: An inquiry into the lunar interior: a nonlinear inversion of the Apollo lunar seismic data. *J. Geophys. Res* **107**(E6), 19–44 (2002)

Buland, A., Omre, H.: Bayesian linearized AVO inversion. *Geophysics* **68**, 185 (2003)

Hansen, T.M., Cordua, K.S. & Mosegaard, K. Inverse problems with non-trivial priors: efficient solution through sequential Gibbs sampling. *Comput Geosci* **16**, 593–611 (2012). doi.org/10.1007/s10596-011-9271-1

Kjartansson, E. and Rocca, F., 1979, The exploding reflector model and laterally variable media: Stanford Exploration Project Report No. 16, Stanford University.

Stoffa P L and Sen M K 1991 Nonlinear multiparameter optimization using genetic algorithms: inversion of plane wave seismograms *Geophysics* **56** 1794–810

Gallagher K, Sambridge M S and Drijkoningen G G 1991 Genetic algorithms: an evolution on Monte Carlo methods in strongly non-linear geophysical optimization problems *Geophys. Res. Lett.* **18** 2177–80

Wilson W. G. and Vasudevan K 1991 Application of the genetic algorithm to residual statics estimation *Geophys. Res. Lett.* **18** 2181–4

Sambridge M. and Drijkoningen G G 1992 Genetic algorithms in seismic waveform inversion *Geophys. J. Int.* **109** 323–42

Sambridge, M., 1999: Geophysical inversion with a neighborhood algorithm. Searching a parameter space, *Geophysical Journal International*, Volume 138, Issue 2, 479–494,

Scales J. A., Smith M. L. and Fischer T. L. 1992 Global optimization methods for multimodel inverse problems *J. Comput. Phys.* **103** 258–68

Sen M. K. and Stoffa P L 1992 Rapid sampling of model space using genetic algorithms *Geophys. J. Int.* **108** 281–92

Smith M. L., Scales J A and Fischer T L 1992 Global search and genetic algorithms *Geophysics: The Leading Edge of Exploration* pp 22–6

Pedersen J B and Knudsen O 1990 Variability of estimated binding parameters *Biophys. Chem.* **36** 167–76

Gouveia W P and Scales J A 1998 Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis *J. Geophys. Res.* **103** 2759–79

Dahl-Jensen D, Mosegaard K, Gundestrup N, Clow G D, Johnsen S J, Hansen A W and Balling N 1998 Past temperatures directly from the Greenland ice sheet *Science* **9** 268–71

Khan A and Mosegaard K 2001 New information on the deep lunar interior from an inversion of lunar free oscillation periods *Geophys. Res. Lett.* **28** 1791

Khan A, Mosegaard K and Rasmussen K L 2000 A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo Lunar seismic data *Geophys. Res. Lett.* **27** 1591–4

Khoshkholgh, S., Zunino, A., & Mosegaard, K., 2021: Full-waveform Inversion by Informed proposal monte carlo. Earth and Space Science Open Archive, doi:10.1002/essoar.10506565.1

Khoshkholgh, S., Zunino, A., & Mosegaard, K., 2021: Informed Proposal Monte Carlo. *Geophysical Journal International*, Volume 226, Issue 2, 1239–1248, <https://doi.org/10.1093/gji/ggab173>

Lekic, V., Romanowicz, B., 2011: Inferring upper-mantle structure by full waveform tomography with the spectral element method. *Geophysical Journal International*, Volume 185, Issue 2, 799–831,

## *List of Figures*

1.1 The solution of inverse problem and the conjunction of the two states of information.....	6
1.2 The average efficiency of all optimization algorithms .....	11
2.1 Importance sampling, complex probability densities can be represented by a collection of samples in the model .....	15
2.2 Comparing true posterior distribution and approximate posterior ...	20
2.3 Comparing true acoustic impedance and approximated acoustic impedance.....	22
2.4 Comparing true model error envelope and approximate model error envelope.....	23
3.1 Approximate velocity model .....	28
3.2 Realizations from prior distribution.....	31

# Appendices



## **Ricker Wavelet**

Ricker wavelet is a symmetric theoretical wavelet which can be defined as the second derivative of a Gaussian function mathematically (Ricker 1943, 1944; Wang, 2015). The amplitude of Ricker wavelet in time domain is shown as:

$$r(t) = (1 - \frac{1}{2} \omega_p^2 t^2) \exp(-\frac{1}{4} \omega_p^2 t^2)$$

Where  $t$  is time and  $\omega_p$  is the peak frequency. Figure A.1 shows the 2D Ricker wavelet with maximum frequency of  $\omega_p$  and the frequency spectrum of it.

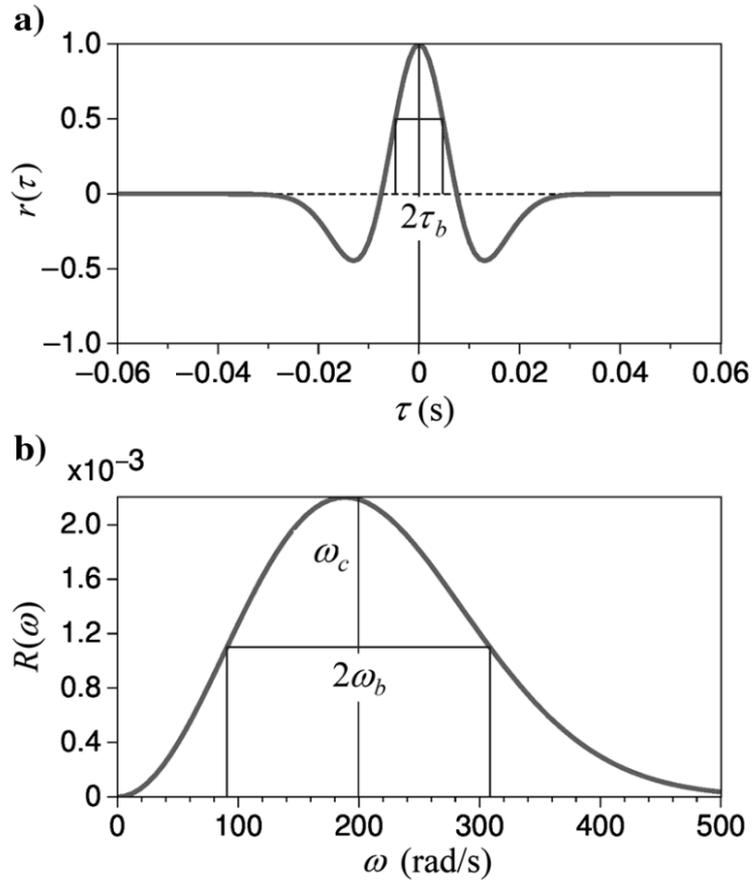


Figure C.1. a) Ricker wavelet  $r(\tau)$ . b) The frequency spectrum. (Wang, 2015)



## The Gaussian Distribution

The multivariate normal (Gaussian) distribution is defined by Johnson and Wichern which is a generalization of univariate normal distribution to several dimensions. The Gaussian distribution in  $p$  dimensions for a random vector  $\mathbf{X}=[X_1, X_2, \dots, X_p]$  is shown as:

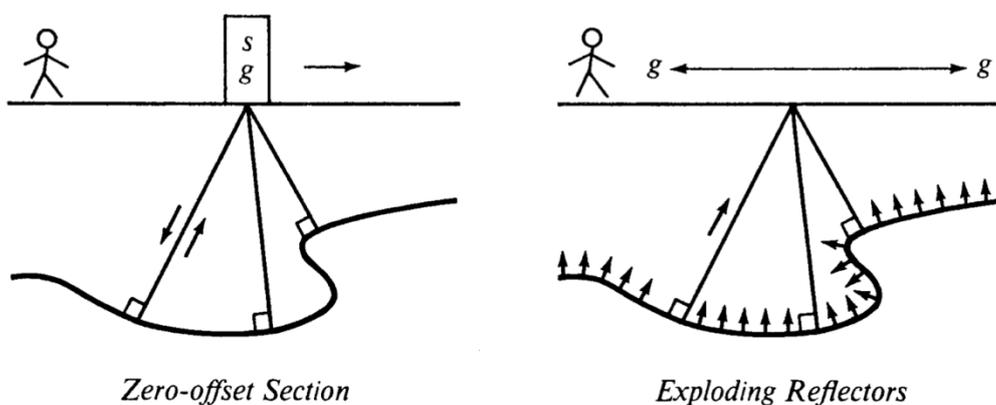
$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |C|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' C^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Where  $C$  is a symmetric  $p \times p$  matrix that is variance -covariance matrix of  $\mathbf{X}$  and  $\boldsymbol{\mu}$  is a  $p \times 1$  vector showing expected value of  $\mathbf{X}$ . The multivariate normal distribution can be denoted by  $N_p(\boldsymbol{\mu}, C)$  (Johnson and Wichern, 2007).



## Exploding reflectors

When the reflectors in the subsurface act as imaginary sources and explode, the waves from explosion travel upward to the surface where we have hypothetical receivers. This conceptual modeling is called exploding reflectors and it is equivalent to zero-offset section. This similarity is not very accurate in existence of strong lateral velocity variations (Kjartansson and Rocca, 1979)



Zero-offset section and exploding reflectors conceptual models



## Full Waveform modelling

### The Elastic wave equation

The process of approximating the seismic wavefield propagation when the subsurface medium properties are known, is called full-waveform modelling. The propagation of seismic wave-field in the subsurface can be modelled by solving the elastic or acoustic wave equations. If elastic wave equation is used for simulation then the propagation can be modelled as:

$$\rho(\mathbf{x}) \ddot{\mathbf{u}}(\mathbf{x}, t) - \nabla \sigma(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t)$$

Where  $\mathbf{u}$  is the displacement field,  $\rho$  is density and  $\sigma$  is the stress tensor and  $\mathbf{f}$  is the external force density (Fichtner, 2011). This equation is a form of Newton's second law that shows equilibrium of momentum of particle, forces caused by internal stress and external forces (Fichtner, 2011). By replacing the stress-strain tensor and using spatial derivatives of  $\mathbf{u}$  we get a set of differential equations that characterize elastic wave equation (Aki and Richards, 2002; Kennett 2001). There are several approaches for deriving a numerical solution for wave equation, and the finite difference method is the most well-known and commonly used procedure in the literature.

## The Acoustic wave equation

Acoustic wave equation is usually used to decrease the computational burden of the problem. In this thesis, we use acoustic wave equation to model the forward problem. Acoustic wave equation considers pressure waves traveling in a fluid media. The linear wave equation for a particle in motion is (Leandro Di Bartolo et al 2012):

$$\rho(\mathbf{r}) \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial t} + \nabla p(\mathbf{r}, t) = \mathbf{f}(\mathbf{r}, t)$$

Where  $\mathbf{v}$  and  $p$  are velocity and acoustic pressure correspondingly.  $t$  and  $\mathbf{r}$  are time and position vector,  $\rho$  is density and  $\mathbf{f}$  is the density external force. This equation can be re-written as following:

$$\frac{1}{k(\mathbf{r})} \frac{\partial p(\mathbf{r}, t)}{\partial t} + \nabla \cdot \mathbf{v}(\mathbf{r}, t) = \frac{\partial i_v(\mathbf{r}, t)}{\partial t}$$

And  $k$  is the medium's compression modules and  $i_v$  is representing the source. Equation 3.3 and 3.4 can be simplified to the following equation

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2(\mathbf{r})} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = -s(\mathbf{r}, t)$$

Finite difference method is one of the most famous schemes for solving partial differential equations in wave propagation problem in full waveform modeling. There are many different examples of using finite difference method for solving

elastic or acoustic wave equation in time or frequency domain (Alford et al 1974, Marfurt 1984, Levander 1988, Kelly et al 1976). Staggered grid or partially staggered grid approach were also introduced and applied in geophysical problems by different researchers (Madariaga 1976, Magnier et al 1994, Saenger et al 2000). In this work we use a Matlab toolkit that uses central finite difference strategy to estimate partial derivatives of acoustic wave equation (Carrie et al 1999).

### Finite difference modelling of acoustic waves

In case of small density variations central difference approach is known to be simple and popular. We have the scalar wave equation in 2D (Carrie et al, 1999):

$$\frac{\partial^2 \phi(x, z, t)}{\partial t^2} = \nu^2(x, z) \nabla^2 \phi(x, z, t)$$

Where the Laplacian is:

$$\nabla^2 \phi = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2}$$

Using central difference operators there will be two types of approximations (second order and fourth order). The second order approximation to the Laplacian operator results in:

$$\nabla^2 \phi_j^n \approx \frac{\phi_j^{n+1} - 2\phi_j^n + \phi_j^{n-1}}{\Delta x^2} + \frac{\phi_{j+1}^n - 2\phi_j^n + \phi_{j-1}^n}{\Delta z^2}$$

n and j are x coordinate and z coordinate grid correspondingly. And fourth order approximation is shown as:

$$\nabla^2 \phi_j^n \approx \frac{1}{\Delta x^2} \left( \frac{-1}{12} \phi_j^{n+2} + \frac{16}{12} \phi_j^{n+1} - \frac{30}{12} \phi_j^n + \frac{16}{12} \phi_j^{n-1} - \frac{1}{12} \phi_j^{n-2} \right) + \frac{1}{\Delta z^2} \left( \frac{-1}{12} \phi_{j+2}^n + \frac{16}{12} \phi_{j+1}^n - \frac{30}{12} \phi_j^n + \frac{16}{12} \phi_{j-1}^n - \frac{1}{12} \phi_{j-2}^n \right)$$

If we assume that the grid spacing is similar in to directions and  $\Delta x = \Delta z$ , then the stability requirement for second order approximation is:

$$\frac{V_{max} \Delta t}{\Delta x} \leq \sqrt{\frac{1}{2}}$$

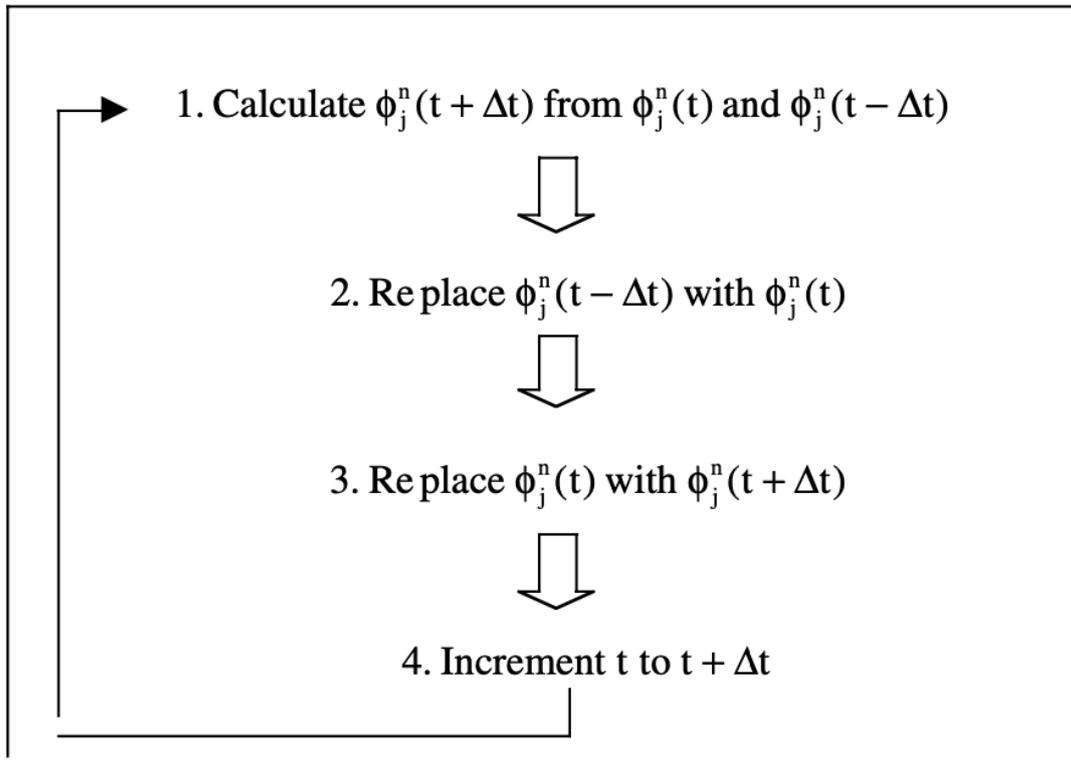
And for fourth order approximation we have :

$$\frac{V_{max} \Delta t}{\Delta x} \leq \sqrt{\frac{3}{8}}$$

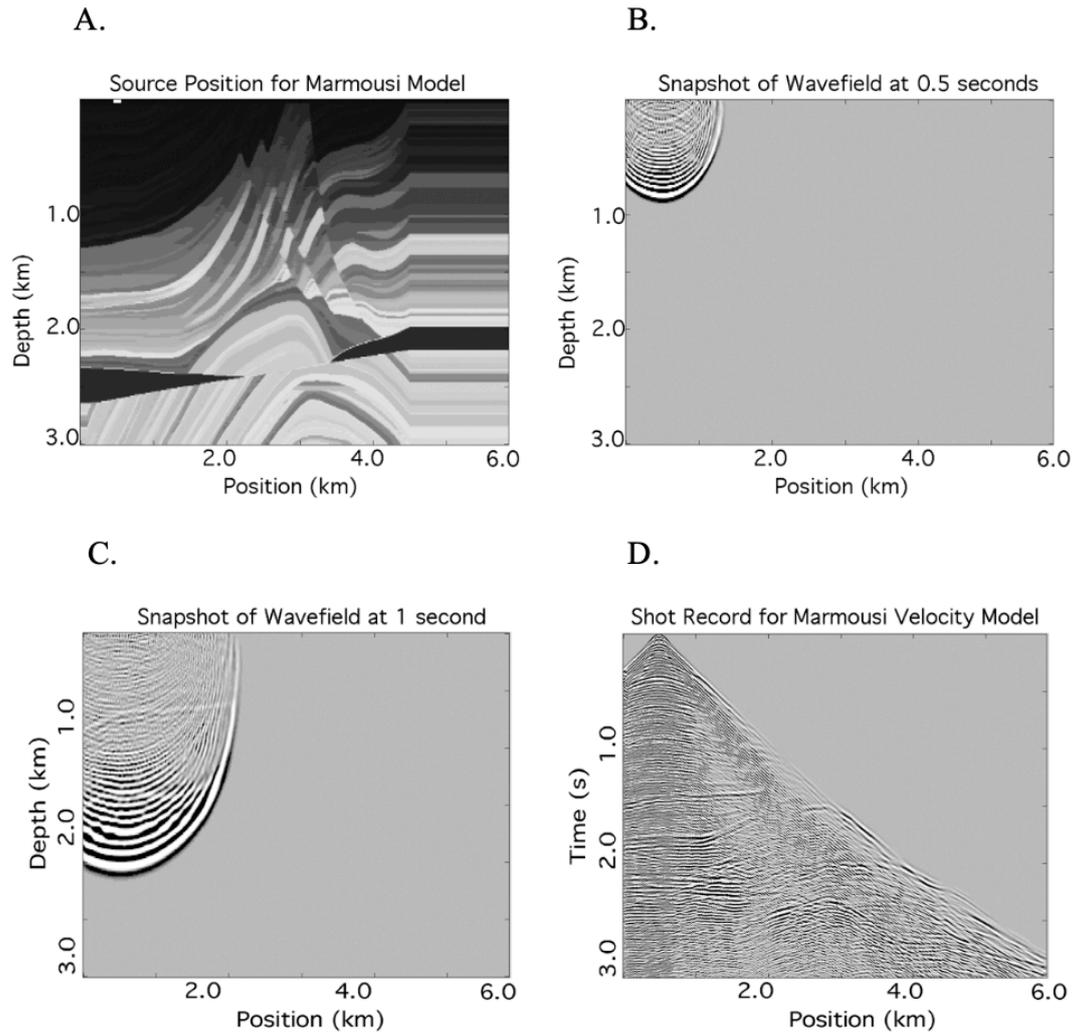
Using second order finite difference approach for wavefield at time  $t + \Delta t$  we will have:

$$\phi_j^n(t + \Delta t) \approx (\Delta t^2 (v_j^n)^2 \nabla^2 + 2) \phi_j^n(t) - \phi_j^n(t - \Delta t)$$

Knowing the seismic waveform at  $t$  and  $t - \Delta t$  we are able to calculate it at  $t + \Delta t$  and generate snapshots showing it at different discrete times. The procedure of the finite difference algorithm is shown in figure 3.1 (Carrie et al, 1999).



Finite difference algorithm procedure (Carrie et al, 1999)



A) The marmousi velocity model. B) A snapshot of the wavefield at 0.5 s C) A snapshot of the wavefield at 1 s D) The shot record of the velocity model (Carrie et al. 1999)

## *Scientific Work*

### *E.1 Informed Proposal Monte Carlo*

*Sarouyeh Khoshkholgh, Andrea Zunino, Klaus Mosegaard*

## Informed proposal Monte Carlo

Sarouyeh Khoshkholgh<sup>1</sup>, Andrea Zunino<sup>\*</sup> and Klaus Mosegaard

*Department of Physics of Ice, Climate and Earth, Tagensvej 16, 2200 Copenhagen N, Denmark. E-mail: sarouyeh@nbi.ku.dk*

Accepted 2021 April 26. Received 2021 April 18; in original form 2020 July 27

### SUMMARY

Any search or sampling algorithm for solution of inverse problems needs guidance to be efficient. Many algorithms collect and apply information about the problem on the fly, and much improvement has been made in this way. However, as a consequence of the No-Free-Lunch Theorem, the only way we can ensure a significantly better performance of search and sampling algorithms is to build in as much external information about the problem as possible. In the special case of Markov Chain Monte Carlo (MCMC) sampling we review how this is done through the choice of proposal distribution, and we show how this way of adding more information about the problem can be made particularly efficient when based on an approximate physics model of the problem. A highly non-linear inverse scattering problem with a high-dimensional model space serves as an illustration of the gain of efficiency through this approach.

**Key words:** Inverse theory; Statistical methods; Waveform inversion; Computational seismology.

### 1 INTRODUCTION

Since its introduction in the late 1900s (Metropolis *et al.* 1953; Hastings 1970; Kirkpatrick *et al.* 1983; Duane *et al.* 1987; Marinari & Parisi 1992), Markov Chain Monte Carlo (MCMC) methods have been established as a main tool for providing solutions and uncertainty estimates for small- to intermediate-scale, highly non-linear inverse problems. This development is closely connected to the dramatic increase in computational speed over the last few decades. However, there has also been an increasing demand for solving inverse problems on a larger scale, with more time-consuming forward calculations (e.g. Dettmer *et al.* 2011; Fichtner *et al.* 2018), and more complex *a priori* information (e.g. Grana 1999; Khan *et al.* 2007; Lange *et al.* 2012; Cordua *et al.* 2015; Zunino *et al.* 2015; Hunziker *et al.* 2017; Laloy *et al.* 2018), some of which do not even rely on smoothness of the problem. In this connection it has become clear that straightforward use of standard MCMC algorithms is unfeasible, and recent years have seen a surge of improved samplers with more and more sophisticated sampling strategies (Tierney 1999; Liu 2002; MacKay 2003; Haario *et al.* 2006; Brooks *et al.* 2011; Vrugt 2016; Ying *et al.* 2020). Most of the methods developed for improved efficiency use gradients of the posterior density (or its logarithm) in their proposal mechanism (Roberts & Tweedie 1996; Dosso 2002; Girolami & Calderhead 2011; Neal 2011; Dosso *et al.* 2014) raising the problem that gradients are generally computationally expensive, except in cases where special, problem-specific

properties of the posterior density can be utilized (Fichtner *et al.* 2018).

Despite all these improvements, there is a growing impression amongst applicants that MCMC strategies are fundamentally slow (Raj *et al.* 2014; Yogatama *et al.* 2014; Andersen *et al.* 2018), and that alternatives should be found. This experience has indeed led to improvements where quite efficient solutions, all tailored to the problem at hand through *a priori* constraints and/or well-chosen simplifying assumptions, have shown promising results. Notable examples are scalable Monte Carlo algorithms which aim at improved computational efficiency through parallelization or subsampling the data (Neiswanger *et al.* 2014; Rabinovich *et al.* 2015; Scott *et al.* 2016; Minsker *et al.* 2017; Nemeth & Sherlock 2018; Srivastava *et al.* 2018). They work for problems where samples can be drawn from a partial posterior that is conditioned on a limited subset of data. For this category of algorithms the challenge is to merge the many partial posterior samples to create a reasonable approximation to the full posterior. For the general non-linear problem, this may be difficult to accomplish with sufficient accuracy (Nemeth & Fearnhead 2020).

Another recent development is an attempt to perform often time-consuming likelihood calculations with neural networks, trained on a very large number of model-data pairs sampled from an a priori probability distribution (Andrieu *et al.* 2003; Hunziker *et al.* 2017; Laloy *et al.* 2017, 2018; Scheidt *et al.* 2018; Nawaz & Curtis 2019; Holm-Jensen & Hansen 2020).

The ‘race of Monte Carlo ideas’ has been accompanied by intense discussions in the research community about the efficiency of algorithms. Not only have intuitive ideas been held up against each other, but arguments for and against methodologies have also been

<sup>\*</sup>Now at: Department of Earth Sciences, ETH Zürich, Sonneggstrasse 5, 8092 Zurich, Switzerland

accompanied by numerical experiments to support the conclusions. This approach rests apparently on a sound basis, but if we take a closer look at the way algorithm comparisons are typically carried out, we discover a common deficiency: In very few cases algorithms are compared by solving *exactly* the same problem. At the surface, test problems look similar, but a closer look reveals that the information available to algorithms in the same test differs significantly. All too often, the total number of numerical operations (forward calculations for an inverse problem) is not used as a measure of computational workload when evaluating the efficiencies. This can be a problem when presenting the results from machine learning, if the computational workload of ‘training’ is not included when the efficiency is evaluated. Another example may be in presentation of gradient-based algorithms where efficiency is evaluated as ‘number of iterations to convergence’, without counting numerical operations needed to compute the gradients. As a result, comparisons often become meaningless, but there is one thing that seems clear from most comparative studies: The more additional information about the posterior probability density of the inverse problem we build into the code of an algorithm, the better is our chance to create an efficient the algorithm.

The purpose of this paper is to explore how additional information in MCMC sampling may significantly reduce the computational workload. We will first discuss the reasons for the often excessive time-consumption of simple MCMC strategies. We will then turn to the problem of finding and applying supplementary information to speed up calculations, not in the form of *a priori* information about possible solutions, but from the approximations to the physical problem (the forward relation). Our aim will be to apply this information in a way that will not bias the sampling asymptotically. We shall explore and support our findings through numerical experiments and by comparing with other methods.

Our test example will be the acoustic inverse scattering problem for a vertical plane wave hitting a horizontally stratified medium with varying acoustic impedance (product of wave speed and mass density). This problem is highly non-linear due to internal multiple scattering (echoes) and attenuation in the medium. Since our aim is to evaluate solutions and their uncertainties, we use MCMC for the analysis. We compare an MCMC sampling approach with a simple proposal distribution with one where the proposal mechanism is designed from an approximation to the forward relation. The result is a significant improvement in the algorithm’s efficiency.

## 2 MCMC AND THE PROPOSAL PROBLEM

### 2.1 Proposal distributions

The basic idea behind any implementation of MCMC is an interplay between *proposals* and *rejections*. In each iteration, sampling from a probability density  $f(\mathbf{x})$  over a space  $\mathcal{X}$  proceeds from a current value  $\mathbf{x}$  by first randomly proposing a new value  $\mathbf{x}'$  according to the so-called *proposal distribution*  $q(\mathbf{x}'|\mathbf{x})$ , followed by a random decision where  $\mathbf{x}'$  is accepted, with probability (Metropolis *et al.* 1953; Hastings 1970):

$$P_{\text{acc}}^{\mathbf{x} \rightarrow \mathbf{x}'} = \min \left( \frac{f(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')}{f(\mathbf{x})q(\mathbf{x}'|\mathbf{x})}, 1 \right). \quad (1)$$

This acceptance probability ensures that, once an equilibrium sampling distribution is established, it will be maintained through *detailed balance*, because the probability  $P_{\text{acc}}^{\mathbf{x} \rightarrow \mathbf{x}'} q(\mathbf{x}'|\mathbf{x}) f(\mathbf{x})$  of a

transition from  $\mathbf{x}$  to  $\mathbf{x}'$  equals the probability of the reverse transition,  $P_{\text{acc}}^{\mathbf{x}' \rightarrow \mathbf{x}} q(\mathbf{x}|\mathbf{x}') f(\mathbf{x}')$  (Mosegaard & Sambridge 2002). At this point it is important to note that the proposal distribution has no influence on the distribution to which the sampling converges, it only influences the speed of convergence.

The two most common types of proposal distributions are:

(i) *Local* proposal distributions  $q$ , where  $q(\mathbf{x}'|\mathbf{x})$  depends on the starting point  $\mathbf{x}$ . A frequent assumption is translation invariance where  $q(\mathbf{x}'|\mathbf{x}) = q(\mathbf{x}' - \mathbf{a}|\mathbf{x} - \mathbf{a})$  for any shift  $\mathbf{a}$  in the parameter space. Another common assumption is symmetry:  $q(\mathbf{x}'|\mathbf{x}) = q(\mathbf{x}|\mathbf{x}')$ , and in this case we get a simpler expression for the acceptance probability (1):

$$P_{\text{acc}}^{\mathbf{x} \rightarrow \mathbf{x}'} = \min \left( \frac{f(\mathbf{x}')}{f(\mathbf{x})}, 1 \right). \quad (2)$$

(ii) *Global* proposal distributions  $q$  that are independent of the starting point  $\mathbf{x}$ . This means that  $q(\mathbf{x}'|\mathbf{x}) = h(\mathbf{x}')$  where  $h(\mathbf{x})$  is fixed during the sampling process. If  $h(\mathbf{x})$  is in some sense close to the target distribution  $f(\mathbf{x})$ ,  $h$  is often called a ‘surrogate’ (for  $f$ ).

An MCMC sampler is only efficient if large enough steps (connecting any two areas of high values of  $f(\mathbf{x})$  in a few steps) are frequently proposed and accepted. This ability critically depends on  $q(\mathbf{x}'|\mathbf{x})$ , and requires that  $q(\mathbf{x}'|\mathbf{x})$  is (at least locally) similar to  $f(\mathbf{x}')$ . This is revealed by a close look at the expression for the transition probability from  $\mathbf{x}$  to  $\mathbf{x}'$ :

$$P(\mathbf{x}'|\mathbf{x}) = q(\mathbf{x}'|\mathbf{x}) \cdot \min \left( \frac{f(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')}{f(\mathbf{x})q(\mathbf{x}'|\mathbf{x})}, 1 \right), \quad (3)$$

showing that the probability of the transition  $\mathbf{x} \rightarrow \mathbf{x}'$  is high if

- (i)  $q(\mathbf{x}'|\mathbf{x})$  is large and
- (ii)  $f(\mathbf{x}') \approx C \cdot q(\mathbf{x}'|\mathbf{x})$  where  $C$  is a constant.

We will now see how implementations of local and global proposals may address these questions.

### 2.2 Local proposals

The use of local proposals is an attempt to satisfy the above two conditions:

- (i) This condition is obviously locally true (close to  $\mathbf{x}$ ) for a local proposal (per definition).
- (ii) This condition is usually met by assuming that  $f$  is smooth and by choosing a smooth, sufficiently narrow  $q(\mathbf{x}'|\mathbf{x})$ . In this way the condition applies in most of  $q$ ’s support.

Local proposals are widely used, but they have at least two serious drawbacks. First, if they are too narrow, the proposed steps will be so small that the algorithm needs many iterations to traverse the parameters space. As a result, many iterations are required to produce sufficiently many independent samples from the space. Secondly, condition (ii) may not be easy/possible to satisfy (or insufficient) in practice. Either because  $f$  is not smooth, or if  $f$  is smooth, but a ‘sufficiently narrow’  $q$  only allows unacceptably small steps.

To exemplify and quantify the possible problems with local proposals in high-dimensional spaces, let us consider the case where the target distribution of  $\mathbf{x}$  is Gaussian with covariance matrix  $\mathbf{C}$  and mean  $\mathbf{x}_0$ :  $f(\mathbf{x}) = \mathcal{N}_{\mathbf{x}}(\mathbf{x}_0, \mathbf{C})$ . Assume for illustration that our proposal distribution is an isotropic Gaussian  $q(\mathbf{x}|\mathbf{x}_q) = \mathcal{N}_{\mathbf{x}}(\mathbf{x}_q, \mathbf{C}_q)$  with mean  $\mathbf{x}_q$  and covariance matrix  $\mathbf{C}_q$ , and that we, in the sampling process, have been fortunate to arrive at point with a high value

of  $f(\mathbf{x})$ , say, for simplicity, at its maximum point  $\mathbf{x}_0$ . We can now calculate the expected acceptance probability  $P^{\mathbf{x}_0 \rightarrow \mathbf{x}}$  proposed in the next step by the algorithm:

$$\begin{aligned} E(P^{\mathbf{x}_0 \rightarrow \mathbf{x}}) &= \int_{\mathcal{X}} \frac{f(\mathbf{x})}{f(\mathbf{x}_0)} q(\mathbf{x}|\mathbf{x}_0) d\mathbf{x} \\ &= \int_{\mathcal{X}} \frac{\mathcal{N}_{\mathbf{x}}(\mathbf{x}_0, \mathbf{C})}{\mathcal{N}_{\mathbf{x}_0}(\mathbf{x}_0, \mathbf{C})} \mathcal{N}_{\mathbf{x}}(\mathbf{x}_0, \mathbf{C}_q) d\mathbf{x} \\ &= \frac{\mathcal{N}_{\mathbf{x}_0}(\mathbf{x}_0, \mathbf{C} + \mathbf{C}_q)}{\mathcal{N}_{\mathbf{x}_0}(\mathbf{x}_0, \mathbf{C})} \int_{\mathcal{X}} \mathcal{N}_{\mathbf{x}}(\mathbf{x}_1, \mathbf{C}_1) d\mathbf{x}, \end{aligned} \tag{4}$$

where

$$\mathbf{x}_1 = (\mathbf{C}^{-1} + \mathbf{C}_q^{-1})^{-1} (\mathbf{C}^{-1} \mathbf{x}_0 + \mathbf{C}_q^{-1} \mathbf{x}_0) = \mathbf{x}_0 \tag{5}$$

and

$$\mathbf{C}_1 = (\mathbf{C}^{-1} + \mathbf{C}_q^{-1})^{-1}. \tag{6}$$

Since the last integral in (4) is 1, we have the following expression for the expected acceptance probability:

$$E(P^{\mathbf{x}_0 \rightarrow \mathbf{x}}) = \frac{\mathcal{N}_{\mathbf{x}_0}(\mathbf{x}_0, \mathbf{C} + \mathbf{C}_q)}{\mathcal{N}_{\mathbf{x}_0}(\mathbf{x}_0, \mathbf{C})} = \left( \frac{\det(2\pi \mathbf{C})}{\det(2\pi(\mathbf{C} + \mathbf{C}_q))} \right)^{1/2}. \tag{7}$$

Both  $\mathbf{C}_q = \sigma_q^2 \mathbf{I}$  (with  $\sigma_q^2 > 0$ ) and  $\mathbf{C}$  are diagonal in the frame spanned by  $\mathbf{C}$ 's eigenvectors, and if we assume that the eigenvalues of  $\mathbf{C}$  are  $\sigma_1^2 \geq \dots \geq \sigma_N^2 > 0$ , where  $N$  is the dimension of  $\mathcal{X}$ , the eigenvalues of  $\mathbf{C} + \mathbf{C}_q$  are  $(\sigma_1^2 + \sigma_q^2), \dots, (\sigma_N^2 + \sigma_q^2)$ . From this we have

$$E(P^{\mathbf{x}_0 \rightarrow \mathbf{x}}) = \prod_{n=1}^N \left( \frac{\sigma_n^2}{\sigma_n^2 + \sigma_q^2} \right)^{1/2}. \tag{8}$$

From (8) we see that for any non-zero values of  $\sigma_n$  and  $\sigma_q$  we have

$$E(P^{\mathbf{x}_0 \rightarrow \mathbf{x}}) \rightarrow 0 \quad \text{for } N \rightarrow \infty. \tag{9}$$

If the proposed steps are kept very short ( $\sigma_q$  is small compared to all  $\sigma_n$ ), the decrease of  $E(P^{\mathbf{x}_0 \rightarrow \mathbf{x}})$  with  $N$  is slow. But this situation is of no practical value, because adequate sampling by the algorithm requires that it can traverse high-probability areas of  $f(\mathbf{x})$  within a reasonable amount of time. For non-negligible step lengths, the situation is radically different. Indeed, if there exists an integer  $K$  and a real constant  $k$  such that  $\sigma_q > k\sigma_n$  for all  $n > K$ , then  $E(P^{\mathbf{x}_0 \rightarrow \mathbf{x}})$  decreases more than exponentially with  $N$ . In other words, if the distribution  $f(\mathbf{x})$  is 'elongated' compared to the proposal  $q$ , that is, if it is broader than  $q$  in only a fixed number  $K < N$  of directions/dimensions, the mean number of accepted moves will decrease at least exponentially with the number of dimensions.

As an example, let us consider the case where  $\sigma_q^2 = 1$ , and  $\sigma_n^2 = 1/n$ . For  $N = 2$  this gives an expected acceptance probability of 0.4082, corresponding to a mean waiting time of about  $0.4082^{-1} \approx 2.5$  iterations between accepted moves. For  $N = 10$  the expectation is  $1.5828 \times 10^{-4}$ , and if we could compute  $N = 100$  it would decrease to  $1.03 \times 10^{-80}$ , giving a waiting time of about  $3.0 \times 10^{62}$  yr for 1 Billion iterations per second.

The above analysis is carried out under the favorable assumption that the maximum of  $f(\mathbf{x})$  has been located by the algorithm, and does not even consider the serious difficulties faced by the sampling algorithm in the initial search for points with high values of  $f(\mathbf{x})$  (the *burn-in* phase). Hence, it is clear that the proposal mechanism, as defined by  $q$ , is the Achilles heel of the standard MCMC approach.

### 2.3 Global proposals

A global proposal  $q(\mathbf{x}'|\mathbf{x})$  is independent of  $\mathbf{x}$  and hence it can be written  $q(\mathbf{x}'|\mathbf{x}) = h(\mathbf{x}')$ . The use of global proposals seeks to meet the requirements of (i) and (ii) by choosing  $h(\mathbf{x}') \approx f(\mathbf{x}')$ . In fact, from (3) it is easily seen that global proposals are ideal if they closely resemble the target distribution. In the ideal case where  $h(\mathbf{x}') = f(\mathbf{x}')$ , the transition probability is equal to  $f(\mathbf{x}')$ , and the sampler has no rejected moves. Arbitrarily large steps in the sample space are allowed, and therefore all sample points are statistically independent.

However, the problem with global proposals is to find them in the first place. There are, in principle, two approaches:

- (i) Using, as proposal, an approximation  $h(\mathbf{x})$  to  $f(\mathbf{x})$ , estimated/interpolated from already visited sample points in the neighbourhood of  $\mathbf{x}$  (Christen 2005; Ying *et al.* 2020). This proposal may be consistent with (similar to)  $f$  in the neighbourhood of existing sample points.
- (ii) Using a global approximation  $h(\mathbf{x})$  derived from external information about  $f(\mathbf{x})$ , that is, *not* derived from already visited sample points. This proposal should be consistent (similar to)  $f$  even far away from existing sample points. One way to do this is to perform a simplified inversion based on simplified physics, and to use the calculated, approximate posterior as  $h(\mathbf{x})$ . We shall see an example of this later. Another approach that is sometimes possible is to use external physical knowledge to perform a coordinate transformation  $(x_1, x_2, \dots, x_M) \rightarrow (y_1, y_2, \dots, y_M)$  in the parameter space such that the transformed distributions  $h_y(\mathbf{y})$  and  $f_y(\mathbf{y})$  are approximately equal (except for a normalization factor) along several coordinate directions in the new system. In this case we call  $h_y(\mathbf{y})$  a *partial* approximation to  $f_y(\mathbf{y})$ . We will see later how the concept of partial approximations is successfully applied in Hamiltonian Monte Carlo (HMC) inversion.

In the following we will take a closer look at methods for finding global proposals for inverse problems. However, before we proceed, we shall first understand the fundamental advantage of (ii) over (i). To this aim, we shall look into an important theorem, proven in the late 1990s, namely the No-Free-Lunch (NFL) Theorem (Wolpert & Macready 1997; Mosegaard 2012).

### 3 NO-FREE-LUNCH THEOREMS AND THE IMPORTANCE OF INFORMATION

We will now make an important distinction between *blind algorithms* and *informed algorithms*. We use the following definitions:

- (i) A *blind algorithm* is an algorithm whose search or sampling is performed only via an *oracle*. An oracle is a function that, when called by the algorithm, is able to evaluate the target distribution  $f$  at a given point  $\mathbf{x}$ . The oracle is used by the algorithm as a black box: No other properties of  $f$  than the corresponding inputs and outputs are used. In computer science, blind algorithms are often called *heuristics*. For inversion, there are many well-known examples of blind algorithms in use: Regular MCMC, Simulated Annealing, Genetic Algorithms, Neural Networks, etc.
- (ii) An *informed algorithm* is an algorithm that, in addition to an oracle, uses known, *external* properties of  $f$  to guide/improve the search or sampling. By external properties we mean any information about  $f$  that is not given by samples from  $f$ . Examples of informed algorithms used in geophysical inversion are HMC (Duane *et al.* 1987; Neal 2011), exploiting that for seismic wave fields adjoint

methods can be used to efficiently compute misfit gradients (Fichtner *et al.* 2018) and Discriminative Variational Bayesian inversion exploiting knowledge about the statistics of the unknown model in case it is a Markov Random Field (Nawaz & Curtis 2019).

Based on the No-Free-Lunch Theorem (Wolpert & Macready 1997), Mosegaard (2012) considered limits for the performance of algorithms designed for solution of inverse problems. The conclusion was that all blind inversion algorithms in finite-dimensional spaces (optimization-based as well as sampling-based) have exactly the same performance when averaged over all conceivable inverse problems. Only an algorithm that take into account more characteristics of the ‘forward model’ than given by the oracle can ensure performance that is superior to blind inversion algorithms.

We can draw the conclusion that efficient inversion algorithms are the ones that operate in accordance with specific properties of the problem it is aiming to solve. If the problem is linear with known Gaussian noise statistics and a given Gaussian prior, it can be solved in ‘one iteration’ (applying a closed-form solution formula). If the problem is mildly non-linear with, for example Gaussian noise and Gaussian prior, our knowledge that the posterior probability distribution is unimodal will render the problem solvable in relatively few iterations. For a highly non-linear problem, the situation is, in principle, the same, except that the term ‘highly non-linear’ usually signals a lack of knowledge of the shape of the posterior. The posterior may be highly multimodal and possess other pathologies, but we may still have some sparse knowledge about it, for instance that it is smooth, so we can compute gradients. Irrespective of what we know about the target posterior distribution, we have the option of building this information into the algorithm. If we have plenty of information, we can create an efficient algorithm. If we have sparse information, our algorithm will need more computation time.

Countless methods use interpolation methods to construct local or global approximations to the posterior and to use them as proposals in the sampling process (Christen 2005; Ginting *et al.* 2011; Jin 2011; Stuart *et al.* 2019; Ying *et al.* 2020). These methods are useful and may improve performance, but they still suffer from the limitations set by the NFL Theorem, because they do not bring in additional, external information.

In the following we will suggest an approach that allows us to design more efficient inversion algorithms through incorporation of additional, external information about the target distribution. The approach is general and can be used in deterministic as well as in sampling approaches. In this exposition we will focus on MCMC sampling, and our approach will be to replace a traditional, blind proposal mechanism with one built from external physical information, providing a reasonable global approximation to the posterior.

#### 4 MCMC WITH PROBLEM-DEPENDENT PROPOSALS

The complete, probabilistic solution to an inverse problem—the posterior probability density—contains a certain, finite amount of information. In addition to the posterior, we may have additional, external information about properties of the posterior (e.g. smoothness, maximum value, principal axes at selected points). If the external information can be used to guide the sampling process, we gain efficiency, measured as number of numerical operations needed to collect a given number of independent samples. The more external information about the ‘structure’ of the posterior we can build into the algorithm, the more efficient our inversion algorithm can

be. Following the NFL Theorem, these statements must be true for any inversion algorithm, whether it is optimization-based (‘deterministic’) or sampling-based (Monte Carlo). Here we will focus on sampling-based algorithms, in particular MCMC.

As stated above, there are basically two ways of obtaining more information about the target (posterior) distribution:

- (i) To collect information while sampling. This can for instance be done via calculation of gradients (e.g. used to derive principal axes proposals). These approaches are ‘blind’, in the sense that, to operate, they do not require the use of external information about the structure of the target distribution. They try to discover properties of the target distribution on the fly, and are therefore limited by the (finite) amount of information expressed therein. This category of algorithms is not the theme of this paper. ‘Blind’ algorithms are subject to the NFL Theorem, stating that the performance of all these algorithms is, when averaged over all conceivable problems, exactly the same. In other words, if an algorithm based on, for example principal axes proposals is good at solving certain problems, it must be poor in other cases (to maintain the general average).
- (ii) The second way to obtain more information about the target distribution is to use external information about the structure of the posterior. Such information will make exploration of the posterior more easy, providing knowledge about the posterior that would otherwise have been difficult (time-consuming) to obtain from samples. This is the subject of our paper. By augmenting an existing blind sampler with a proposal, built from external information, we turn the sampler into a specialized algorithm directed towards sampling our concrete target distribution. In this way we obtain a ‘single purpose’ algorithm that will probably perform well in our case, but most likely will fail in other cases.

#### 4.1 Constructing an MCMC algorithm with an informed proposal

Let us now try to create an MCMC algorithm augmented with external knowledge about the structure (properties) of the posterior to speed up the sampling process. Since our goal is only to improve efficiency of the algorithm (and not to modify the posterior), we want to leave the usual ‘prior’ and ‘likelihood’ untouched. To this aim, we want the external information (e.g. a physical model with simplified mathematics) to enter the Metropolis–Hastings algorithm via the proposal distribution, leaving the posterior asymptotically unbiased. Furthermore, we can allow even oversimplified external information without distorting the result. In the worst case it will only slow down the algorithm.

Let us assume that this information comes from a proposal in the form of an approximation to the posterior probability distribution  $\tilde{\sigma}(\mathbf{m}) \approx \sigma(\mathbf{m})$ . Then, our proposal will not only be close to  $\sigma$  in the neighbourhood of points already visited by the algorithm, it is also expected to work well far away from current samples. We will now see how approximate posteriors can be constructed for linear and non-linear problems.

#### 4.2 Linear, Gaussian problems

Sampling of solutions to a linear Gaussian problem through MCMC sampling is straightforward. Since we have an explicit expression for the Gaussian posterior, the distribution itself can be used as an optimal proposal. Samples  $\mathbf{e}$  from an  $N$ -dimensional standard (isotropic) Gaussian (mean  $\mathbf{0}$  and covariance  $\mathbf{I}$ ) can be generated with, for example the Box–Müller method, and the desired samples

$\mathbf{m}$  from a  $N$ -dimensional multivariate Gaussian with mean  $\mathbf{m}_0$  and covariance  $\mathbf{C}$  can be calculated as  $\mathbf{m} = \mathbf{m}_0 + \mathbf{A}\mathbf{e}$ , where  $\mathbf{A}\mathbf{A}^T = \mathbf{C}$ . The matrix  $\mathbf{A}$  can be found by, for instance, Cholesky decomposition.

### 4.3 Non-linear problems

For non-linear problems there are several ways to proceed, and the method will depend on the external information available about the true posterior. In the following, we will take a look at two important cases, namely (1) when external information about derivatives of the misfit is available and (2) when external information in the form of an approximate relation between model parameters and data is used. The latter case will be investigated in a numerical example in Section (5).

#### 4.3.1 Using information about derivatives of the misfit

Any inverse problem with a smooth posterior can be sampled with fewer rejections if partial derivatives of the misfit function are available (Roberts & Tweedie 1996; Dosso 2002; Girolami & Calderhead 2011; Neal 2011; Dosso *et al.* 2014). However, if derivatives have to be calculated from *internal* information (that is, directly from samples of the posterior itself), the gained efficiency may be lost due to the many extra sample points required to estimate the derivatives. If, however, *external* information about the mathematical structure of the problem allows us to compute derivatives directly at each point in model space, as for instance when using the *method of adjoints* in seismic inversion, it allows us to devise efficient proposals. In general, the higher order derivatives we can get, the more accurate the Taylor approximation to the misfit or the posterior around any point in the model space will be (Girolami & Calderhead 2011). And using this approximation to build a global proposal will lead to high acceptance rates and large step lengths.

To see how gradient information can be used, consider an inverse problem with  $M$  model parameters and an everywhere smooth misfit function  $S(\mathbf{m}) = -\log(\sigma(\mathbf{m}))$ , where  $\sigma(\mathbf{m})$  is the posterior probability density. Assume that the gradient  $\nabla S(\mathbf{m})$  is available for any given  $\mathbf{m}$ , and that we have an approximate solution to the problem  $\tilde{\mathbf{m}}$ . Assume further that in a subset of the parameter space containing  $\tilde{\mathbf{m}}$  we can define an invertible mapping  $\mathbf{x} = \mathbf{f}(\mathbf{m})$  between the original model parameters and new, orthogonal curvilinear coordinates  $\mathbf{x}$ , such that, for any  $\mathbf{m}$ , the first transformed coordinate is  $x_1 = S(\mathbf{m})$ , the first local base vector in the new  $\mathbf{x}$ -coordinate system is  $\mathbf{b}_1 = \nabla S(\mathbf{m})$ , and the remaining base vectors  $\mathbf{b}_2 \dots \mathbf{b}_M$  are orthonormal.  $\mathbf{b}_2 \dots \mathbf{b}_M$  span a subspace in which  $S(\mathbf{m})$  is locally constant around  $\mathbf{m}$ , and this means that local sampling near  $\mathbf{m}$  along all the sampling directions  $\mathbf{b}_1 \dots \mathbf{b}_M$  is simple: The step length in the  $\mathbf{b}_1$ -direction must be tuned according to the size of  $|\nabla S(\mathbf{m})|$ , but local steps in the remaining  $M - 1$  directions can be taken without rejections. This advantage of knowing the gradients is further increased when we observe that we can use the gradients (and indeed higher-order derivatives, if available) to iteratively move in directions of constant  $S(\mathbf{m})$ . This process will, in general, require updated base vectors  $\mathbf{b}_2 \dots \mathbf{b}_M$ , but it will allow large steps without rejections, and this is what we strive for when designing an efficient MCMC algorithm. When using this sampling strategy we should remember that we sample the posterior  $\sigma(\mathbf{x})$  in transformed coordinates  $\mathbf{x}$ , and that the desired posterior  $\sigma(\mathbf{m})$  is computed from  $\sigma(\mathbf{x})$  by multiplication with the absolute value of the Jacobian

determinant of the transformation  $f$ :

$$\sigma(\mathbf{m}) = \sigma(\mathbf{x}) \left| \frac{\partial(x_1, \dots, x_M)}{\partial(m_1, \dots, m_M)} \right|. \tag{10}$$

Since our coordinate transformation  $f$  locally can be viewed (approximately) as a pure rotation with Jacobian determinant 1, followed by a stretching/compression along the gradient with a factor  $|\nabla S(\mathbf{m})|^{-1}$ , the Jacobian determinant for  $f$  becomes:

$$\frac{\partial(x_1, \dots, x_M)}{\partial(m_1, \dots, m_M)} \propto |\nabla S(\mathbf{m})|. \tag{11}$$

This procedure is an important example of how to apply knowledge about gradients everywhere in the model space to define a coordinate transformation  $\mathbf{m} \rightarrow \mathbf{x}$  where a good, partial approximation  $\sigma(\mathbf{x})$  to the posterior can be defined in the new frame. The resulting proposal is approximately proportional to  $\sigma(\mathbf{x})$  (in fact, constant) in all coordinate directions, except one (the gradient direction). The result is a significant gain in sampling efficiency.

An example of using the above method of derivatives is the application of Hamiltonian Monte Carlo (HMC) (Neal 2011) to full-waveform seismic inversion (Fichtner *et al.* 2018; Gebräud *et al.* 2020), in which waveform-misfit gradients are computed using the *method of adjoints*. Using an analogy from analytical mechanics, this method works in an augmented space where half of the parameters are the usual model parameters  $\mathbf{m}$  (viewed as generalized space coordinates), and the other half are ‘generalized momentum parameters’  $\mathbf{p}$ . The posterior distribution in the augmented space  $\hat{\sigma}$  is the product of the posterior distributions over  $\mathbf{m}$  and  $\mathbf{p}$ , respectively:  $\hat{\sigma}(\mathbf{m}, \mathbf{p}) = \sigma(\mathbf{m})\sigma(\mathbf{p})$  where  $\sigma(\mathbf{p})$  is usually set to be Gaussian. Defining the joint misfit as  $\hat{S}(\mathbf{m}, \mathbf{p}) = -\log(\hat{\sigma}(\mathbf{m}, \mathbf{p}))$ , the HMC sampler—knowing the derivatives—is now able to alternate between sampling along contours of constant  $\hat{S}$ , and jumping between contours of different  $\hat{S}$ . Stepping in directions of constant  $\hat{S}$  is accomplished by identifying  $\hat{S}$  with the ‘total energy’, and calculating the orbit of constant energy through integration of Hamilton’s equations. According to Liouville’s theorem in Hamiltonian mechanics (Tolman 2016), this approach will automatically include the Jacobian transformation (10). HMC gives an efficient sampling of the joint posterior, and the marginal posterior  $\sigma(\mathbf{m})$  over the (real) model space is trivially obtained by simply discarding the artificial momentum variables.

The Hamiltonian dynamics used by this method is irrelevant for our discussion, but it is important to understand that the efficiency of this method originates from the knowledge of gradients in the joint space. These gradients carry important information about the (possibly strongly non-Gaussian) posterior, allowing the algorithm to take large steps without risking frequent rejections. The resulting proposal distribution is a partial approximation to the joint posterior in a large subset of the joint space, resulting in efficient sampling of the joint posterior, and hence its marginal in the  $\mathbf{m}$ -subspace.

#### 4.3.2 Using information derived from an approximate forward relation

Let us consider the general expression for the joint posterior probability in the formulation of Tarantola & Valette (1982):

$$\sigma(\mathbf{d}, \mathbf{m}) = \frac{\rho(\mathbf{d}, \mathbf{m})\theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})} \tag{12}$$

where  $\mathbf{d}$  is data,  $\mathbf{m}$  is the model parameters,  $\rho(\mathbf{d}, \mathbf{m})$  is the prior probability density, and  $\mu(\mathbf{d}, \mathbf{m})$  is the homogeneous probability density (assigning equal probabilities to equal volumes) in the joint

( $\mathbf{d}, \mathbf{m}$ )-space. The density  $\theta(\mathbf{d}, \mathbf{m})$  expresses the ‘uncertainty of the forward relation’ between  $\mathbf{m}$  and  $\mathbf{d}$ . For simplicity, let us assume that the homogeneous probability density  $\mu(\mathbf{d}, \mathbf{m})$ , as well as the marginal prior in the model space  $\rho_m(\mathbf{m})$  is constant, which leads us to the following expression for the joint posterior:

$$\sigma(\mathbf{d}, \mathbf{m}) = k \cdot \rho(\mathbf{d})\theta(\mathbf{d}, \mathbf{m}). \quad (13)$$

where  $k$  is a normalization constant. Under the further assumption that the observational data uncertainties are small, compared to the modelization errors (remembering that it is at small data uncertainties that MCMC algorithms show a critical slowing-down), we arrive at the following approximation to the posterior in the model space:

$$\sigma_m(\mathbf{m}) \propto \sigma(\mathbf{d}_{obs}, \mathbf{m}) \approx \theta(\mathbf{d}_{obs}, \mathbf{m}) \quad (14)$$

This is a very rough approximation, but it should be remembered that we will not replace the accurate posterior by this expression. The approximation will only be used as a global proposal distribution to speed up the search/sampling from the correct posterior.

The basic idea will typically be to first solve the inverse problem with simplified physics, obtaining an approximate solution  $\tilde{\mathbf{m}}$ . The deviation of this from the true solution is what we call the (true) modelization error  $\delta\mathbf{m}_{true}$ , and if we can estimate this error, we can use it to build a rough modelization error distribution  $\theta(\mathbf{d}_{obs}, \mathbf{m})$ .

Since we do not know the true solution for a real-data inverse problem, we cannot calculate the error  $\delta\mathbf{m}_{true}$  after having solved the problem.  $\delta\mathbf{m}_{true}$  would have been the difference between the true model vector  $\mathbf{m}_{true}$  and the calculated model vector  $\tilde{\mathbf{m}}$ , but we do not know  $\mathbf{m}_{true}$ . Our solution to this is to create an artificial inverse problem that is ‘close’ to the original problem, the difference being that the true model of the new problem is  $\tilde{\mathbf{m}}$ . The data for this problem can be calculated as  $g(\tilde{\mathbf{m}})$ . We can now solve this problem and compute the modelization error  $\delta\mathbf{m}_{approx}$ . Since the artificial inverse problem is ‘close’ to the original problem, we expect that  $\delta\mathbf{m}_{approx}$  is close to  $\delta\mathbf{m}_{true}$ . In the following section (5) we will adopt the following simple procedure:

- (i) Choose a simplified forward function  $\tilde{g}(\mathbf{m})$  expressing much of the essential physics, and at the same time allowing an efficient (but probably inaccurate) inversion. This step can be skipped if a direct way to the following step (without a formal inversion) is available.
- (ii) Find a solution  $\tilde{\mathbf{m}} = h(\mathbf{d}_{obs})$  to the simplified problem with an acceptable datafit. The ‘pseudo-inverse’  $h$  must give a unique answer, for instance through application of a regularization procedure.
- (iii) Estimate the modelization error introduced by using  $\tilde{g}(\mathbf{m})$  instead of the accurate forward function  $g(\mathbf{m})$ . This error is quantified by the distribution  $\theta(\mathbf{d}_{obs}, \mathbf{m})$ , which is also our rough approximation to the posterior  $\sigma_m(\mathbf{m})$ . The procedure is:

- (a) The ‘true’ modelization error is

$$\delta\mathbf{m}_{true} = \tilde{\mathbf{m}} - \mathbf{m}_{true},$$

but since  $\mathbf{m}_{true}$  is unknown, we compute instead an approximate modelization error

$$\delta\mathbf{m}_{approx} = h(g(\tilde{\mathbf{m}})) - \tilde{\mathbf{m}}.$$

The above formula estimates what the modelization would have been if  $\tilde{\mathbf{m}}$  had been the true model. In case  $\tilde{\mathbf{m}}$  is close to  $\mathbf{m}_{true}$ , we expect that  $\delta\mathbf{m}_{approx}$  will be close to  $\delta\mathbf{m}_{true}$ .

- (b) Use  $\delta\mathbf{m}_{approx}$  to construct a reasonable approximation to the modelization error distribution  $\theta(\mathbf{d}_{obs}, \mathbf{m})$ , centred at  $\tilde{\mathbf{m}}$ . This can be done by assuming a functional form for  $\theta(\mathbf{d}_{obs}, \mathbf{m})$  and by using

the components of  $\delta\mathbf{m}_{approx}$  to obtain the parameters of  $\theta(\mathbf{d}_{obs}, \mathbf{m})$ . An example of this can be found in the following section.

- (iv) Use the approximate modelization error distribution as proposal distribution:

$$q(\mathbf{m}'|\mathbf{m}) = \theta(\mathbf{d}_{obs}, \mathbf{m}') \quad (15)$$

## 5 NUMERICAL EXAMPLE

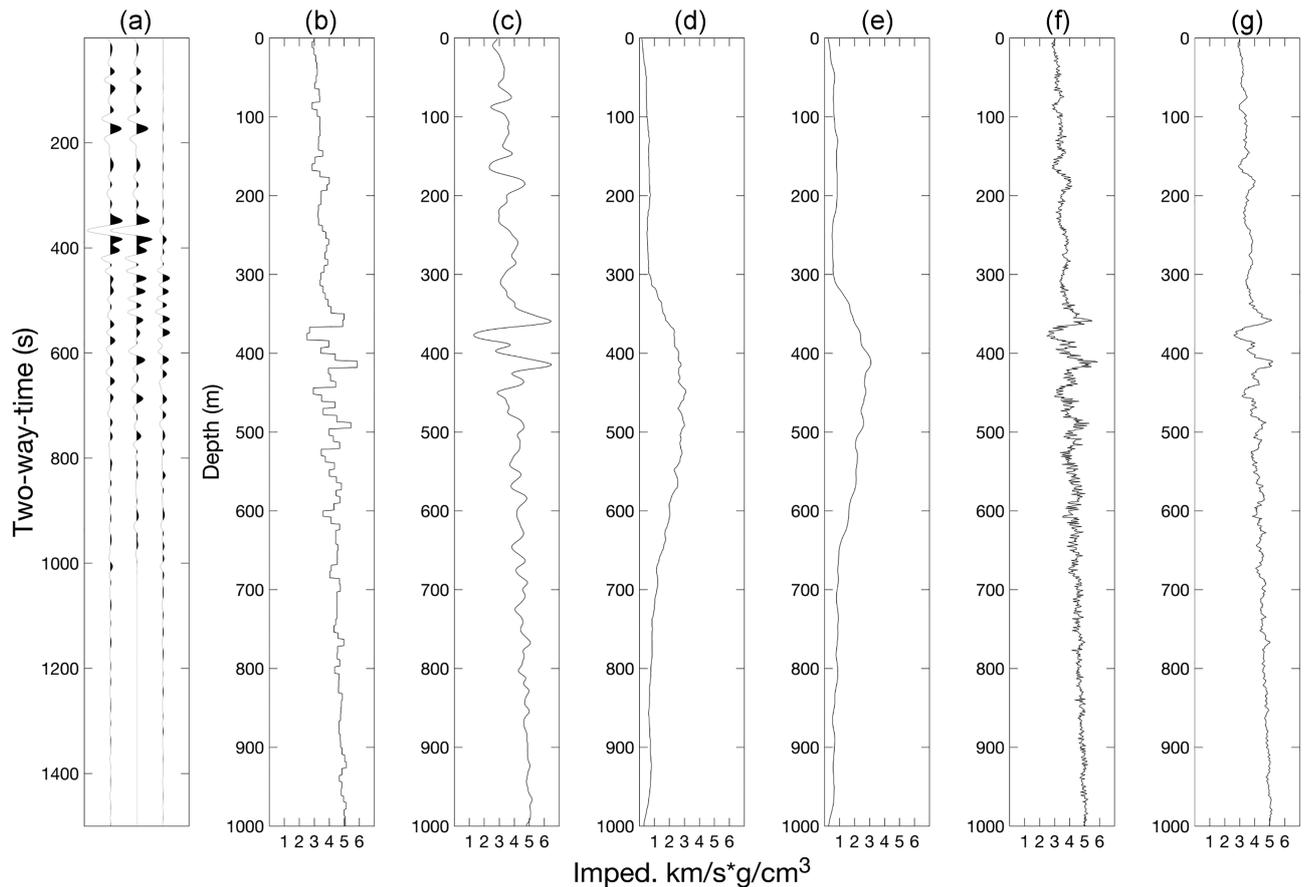
To illustrate the gain of computational efficiency obtained by using even a rough approximation to a high-dimensional target posterior as proposal, we shall look at a 1-D inverse scattering problem (Fig. 1). The unknown model is a horizontally stratified medium with 1000 homogeneous layers. Fig. 1(b) shows the acoustic impedance as a function of distance from the surface. A plane-wave seismic pulse (modelled as a Ricker wavelet) is injected perpendicularly into the medium at the surface, and the data (backscattered waves from the medium) are recorded at the surface (Fig. 1a left-hand side). The data are synthetic 1-D full-waveform seismic signals generated by the propagator matrix method, containing all multiple reflections, transmission losses and damping effects, so the inverse problem of recovering the model from the data is highly non-linear. For comparison, an approximate seismogram, computed by convolution of the reflectivity with the Ricker wavelet, is shown in Fig. 1(a) (middle), together with its error (deviation from the correct seismogram) to the right. Fig. 1(c) shows an approximate solution to the inverse scattering problem in the absence of noise, computed by deconvolution, and converted to impedance through trace integration and addition of the slowly varying trend from Fig. 1(b). The approximate solution requires very little computation time (of the order of one forward calculation), but is clearly inaccurate (compare to the ‘true’ model in Fig. 1b). The purpose of the study is to show how the approximate result can be used to efficiently produce a more accurate solution with uncertainty estimates using Markov Chain Monte Carlo (MCMC).

Our aim is to produce enough samples from the posterior probability distribution in reasonable time, and this raises a well-known problem, namely that the traditional MCMC approach is unfeasible for problems with more than a couple of hundred parameters. Our way of speeding up the sampling is to construct a global proposal distribution for the MCMC sampling using the approximate solution  $\tilde{\mathbf{m}}$ . First, we compute the estimated modelization error vector  $\delta\mathbf{m}_{approx}$  using the method described in the previous section. Fig. 1(e) shows the envelope of the components of this vector, and for comparison, the true modelization error (known in this synthetic data case) is shown in Fig. 1(d). The modelization error distribution is then built as a Gaussian with mean  $\tilde{\mathbf{m}}$  and a diagonal covariance matrix  $\mathbf{C}_\theta$  whose diagonal is the squared components of the envelope function, and used as a proposal distribution.

We now consider the solution of this 1000-parameter problem in four different ways.

### 5.1 Informed proposal Monte Carlo (IPMC)

Fig. 2 (lower curve) shows the convergence of the IPMC algorithm driven by our proposal derived above. A computational overhead of less than 3 forward calculations (deconvolution and calculation of error envelopes) is required to start the sampling process, and we see convergence to equilibrium after about 2000 iterations.



**Figure 1.** (a) Left-hand side: accurate seismogram from (b); Centre: seismogram computed by convolution; Right-hand side: error of the convolution seismogram. (b) True acoustic impedance, (c) acoustic impedance computed by deconvolution (impedance trend from b is added). (d) Envelope of true modelization error (deconvolution impedance minus true impedance). (e) Envelope of estimated modelization error. (f) A sample model from the informed proposal Monte Carlo inversion. (g) Median of 10 000 sample models.

## 5.2 MCMC algorithm with an isotropic, local proposal

We solved the problem without an informed proposal, using a simple MCMC algorithm. We chose an isotropic, local proposal, changing one parameter at a time using a local Gaussian distribution. The step length (standard deviation of the Gaussian) was adjusted to obtain an acceptance rate of approximately 50 per cent. No additional forward calculations were needed to start (initialize) the sampling process. Fig. 2 (upper curve) shows the slow convergence for this algorithm, being far from equilibration after 2000 iterations. A rough estimate showed that the simple MCMC algorithm equilibrated between  $10^3$  and  $10^4$  times slower than the IPMC implementation.

## 5.3 MCMC using a principal axes proposal

An MCMC implementation with a principal axes proposal was considered for solution of the problem. This (blind) method requires an additional computational overhead connected to finding a reasonable initial model, and computation of the Jacobian in this model (and later recompute it after a few hundred iterations). In our test problem we have 4096 data and 1001 unknowns, so a Jacobian requires calculation of more than 4 mill. gradients (at least a few thousand forward calculations). Since the IPMC implementation converges only after about 2000 iterations,

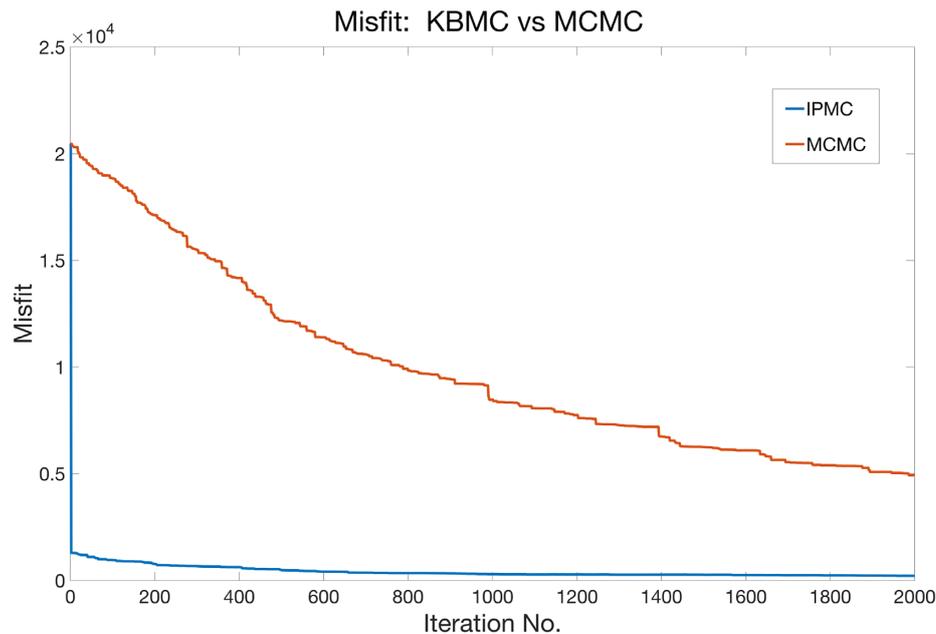
we did not further pursue a solution through the principal axes approach.

## 5.4 Hamiltonian Monte Carlo

We could also have solved our problem with a Hamiltonian Monte Carlo (HMC) approach. As explained above, HMC for seismic inversion is another example of the IPMC strategy using special properties of the posterior density that are not directly available from samples. The method of adjoints can be used to calculate gradients using only 1 forward and 1 adjoint calculation, at the computational cost of approximately 2 forward simulations, and the gradients allow us to define a proposal that is a partial approximation (similar in most directions) to the target distribution in the joint model/momentum space. However, HMC for seismic inversion is well-described elsewhere (Fichtner *et al.* 2018; Gebraad *et al.* 2020), so we have not included this type of solution here.

## 6 DISCUSSION

It is important to realize that the significantly improved efficiency provided by the physical proposal in this study is *not* resulting from prior constraints on the model parameters. Priors generally assign



**Figure 2.** The lower curve shows convergence towards equilibrium of our Informed Proposal Monte Carlo (IPMC) method, which was guided by linearized inversion. The upper curve shows the convergence of a simple MCMC algorithm with a Gaussian proposal, perturbing one parameter at a time, and tuned to an acceptance rate of around 50 per cent. Each iteration of IPMC and simple MCMC required one forward calculation, and involved about 3 and 0 forward calculations for initialization, respectively. The convergence of the simple MCMC was more than  $10^3$  times slower than that of the IPMC algorithm. A gradient-based algorithm was also considered, but would have required calculation of far more than 2000 derivatives (each with 1 forward calculation) to initialize the sampling procedure.

different probabilities to different solutions, but this is not the case with a proposal. A proposal only influences the frequency by which models are presented to the acceptance/rejection algorithm. The bias of the proposal will, asymptotically, be neutralized because it is compensated for in the acceptance probability. In this way it will only influence the efficiency of the sampler, not the asymptotic result. It should, however, be remembered that the most serious problem in non-linear inversion is that the number of models we can practically test is limited. And considering that highly non-linear problems are often so complex that they can only be safely solved with a high number of approximately independent samples from the posterior, it is clear that using an efficient proposal will not only be an improvement in speed, but also a potential improvement in quality of solutions. Simply speaking, we can expect to discover more significantly different solutions (peaks of the target distribution) within the allowed computer resources than with an arbitrary local proposal.

We have illustrated how important it is for the proposal to mimic the posterior in MCMC sampling of solutions to inverse problems. However, the idea of using the physics of the problem to build a posterior-like proposal is not restricted to Monte Carlo sampling. Any method depending on a search for sample solutions or good data fits can potentially benefit from this strategy. In an interesting recent paper on variational full-waveform inversion (Zhang & Curtis 2020), it is shown how variational methods may be used to modify samples from the prior into samples of the posterior in the solution of large-scale inverse problems. It is likely that this class of methods may, in the future, be further improved through application of informed proposal mechanisms.

In this paper, we have chosen to explain the basic ideas of the IPMC strategy and to support them with a strongly non-linear, synthetic test example in a high-dimensional model space. The test

example was synthetic to ensure that we know the true solution and in this way have full control over the behavior of the algorithm. However, to give the reader an idea of how the algorithm could be applied to real data cases and real problems, we give two possible data inversion scenarios.

### 6.1 Elastic full-waveform inversion of pre-stack seismic reflection data

From seismic reflection data we wish to compute realizations of elastic properties of the upper few kilometres of the subsurface. The computationally demanding elastic wave simulations, combined with the strong non-linearity of the problem, is a serious challenge to MCMC methods. For this problem, an approximate solution can be found using classical data-processing and interpretation techniques. An interpreted, depth-migrated profile (volume), combined with optimized migration velocities, can be used to build a rough, approximate model of the subsurface under study. The approximate model and its (accurately computed) synthetic data will now be used in an artificial inverse problem to estimate modelization errors, and the proposal can be built. In a simple implementation, the approximate model will be the ‘centr’ of the global proposal, and the modelization errors will provide the ‘shape’ (dispersion) of the proposal.

### 6.2 Inversion of flow data from a geothermal reservoir

In this problem we wish to compute realizations of the permeability field of a geothermal reservoir from water injection and water production in a large number of wells. Accurate reservoir simulations are computationally expensive, and the problem is highly non-linear.

Assume that we know that five discrete geological facies with approximately known permeabilities are present in the subsurface, and that we wish to build the model by assigning a facies, and a permeability value at each grid point. Each facies has its own prior probability, and its own prior probability distribution for the permeability. Due to the discrete facies parameters, the problem is not smooth. For this problem, an approximate flow simulation can be carried out by a so-called streamline simulator. This simulator estimates the pressure field in the reservoir, followed by flow-line tracing. The algorithm is extremely effective compared to accurate simulators, but its results are imprecise. We first fit a (smooth) least-squares permeability field to the flow data, using the streamline simulator in an iterative linearized inversion. Using our geological knowledge, we then assign facies to each grid point, using the computed permeability field. We now have a (very) rough approximate model. The approximate model and its (accurately computed) synthetic flow data together form an artificial inverse problem from which we can estimate modelization errors. The proposal can now be built, and in a simple implementation we will use the approximate model as the ‘centr of the global proposal, and the modelization errors as ‘dispersion’.

## 7 CONCLUSION

We have analysed the impact of proposal distributions on the performance of MCMC sampling methods when applied to the solution of inverse problems. We concluded that the ‘small step’ strategies used in traditional implementations are relatively efficient because they impose a local consistency between the proposal distribution and the target (posterior) distribution: the target probabilities tend to be large where the proposal probabilities are large. Nevertheless, we showed by a simple analytical example that even local consistency may be difficult to obtain when local ‘small-step’ proposals are arbitrary. Furthermore, a main problem with local proposals is the limited step length, which is strongly hampering the exploration of vast, high-dimensional spaces. The volumes of high-probability areas are negligible in such spaces, so burn-in times, and the times needed to pass from one maximum to another can be prohibitive for small-step algorithms.

Our solution to these problems is to use global proposals built from external information about the target distribution. We propose ways of using physical knowledge of the problem to ensure global consistency between the proposal and the target distribution. The efficiency of this ‘informed proposal’ approach is highly problem-dependent and strongly conditioned on the choice of the external proposal, but we successfully carried out a test on a 1000-parameter, highly non-linear inverse scattering problem. The performance of an MCMC algorithm, augmented with an informed proposal, compared favorably with the three other algorithms considered in this study.

## ACKNOWLEDGEMENTS

This work was supported by Innovation Fund Denmark through the OPTION Project—Optimizing Oil Production by Novel Technology Integration. Joint industry project with Lloyd’s Register, Welltec, Technical University of Denmark (DTU), and University of Copenhagen (5184-00025B). Klaus Mosegaard would like to thank Dr Amir Khan and colleagues at the Department of Earth Sciences, ETH, for their hospitality and inspiring discussions during the fall 2017 where this work was initiated.

## DATA AVAILABILITY

No new data were generated or analysed in support of this research.

## REFERENCES

- Andersen, M., Winther, O., Hansen, L.K., Poldrack, R. & Koyejo, O., 2018. Bayesian structure learning for dynamic brain connectivity, in *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, PMLR **84**, 1436–1446.
- Andrieu, C., De Freitas, N., Doucet, A. & Jordan, M. I., 2003. An introduction to MCMC for machine learning, *Mach. Learn.*, **50**, 5–43.
- Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., 2011. *Handbook of Markov chain Monte Carlo*, CRC Press.
- Christen, J.A. & Fox, C., 2005. Markov Chain Monte Carlo using an approximation, *J. Comput. Graph. Stat.*, **14**, 795–810.
- Cordua, K.S., Hansen, T.M. & Mosegaard, K., 2015. Improving the pattern reproducibility of multiple-point-based prior models, *Math. Geosci.*, **47**, 317–343.
- Dettmer, J., Dosso, S.E. & Holland, C.W., 2011. Sequential trans-dimensional Monte Carlo for range-dependent geoacoustic inversion, *J. acoust. Soc. Am.*, **129**, 1794–1806.
- Dosso, S.E., 2002. Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach, *J. acoust. Soc. Am.*, **111**, 129–142.
- Dosso, S.E., Dettmer, J., Steininger, G. & Holland, C., 2014. Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation, *Inverse Problems*, **30**, 114018.
- Duane, S., Kennedy, A.D., Pendleton, B.J. & Roweth, D., 1987. Hybrid Monte Carlo, *Phys. Lett. B*, **195**, 216–222.
- Fichtner, A., Zunino, A. & Gebraad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.*, **216**, 1344–1363.
- Gebraad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.*, **125**.
- Ginting, V., Pereira, F., Presho, M. & Wo, S., 2011. Application of the two-stage Markov chain Monte Carlo method for characterization of fractured reservoirs using a surrogate flow model, *Comput. Geosci.*, **15**, 691–707.
- Girolami, M. & Calderhead, B., 2011. Riemann manifold Langevin and Hamiltonian Monte Carlo methods (with discussion), *J. R. Stat. Soc. B*, **73**(2), 123–214.
- Grana, D. & Della Rossa, E., 2010. Probabilistic petrophysical-properties estimation integrating statistical rock physics with seismic inversion, *Geophysics*, **75**(3), 1M1–1M7.
- Haario, H., Laine, M., Mira, A., et al., 2006. DRAM: efficient adaptive MCMC, *Stat. Comput.*, **16**, 339–354.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–109.
- Holm-Jensen, T. & Hansen, T.M., 2020. Linear Waveform tomography inversion using machine learning algorithms, *Math. Geosci.*, **52**, 31–51.
- Hunziker, J., Laloy, E. & Linde, N., 2017. Inference of multi-Gaussian relative permittivity fields by probabilistic inversion of crosshole ground-penetrating radar data, *Geophysics*, **82**(5), H25–H40.
- Jin, Y., 2011. Surrogate-assisted evolutionary computation: recent advances and future challenges, *Swarm Evolut. Comput.*, **1**, 61–70.
- Khan, A., Connolly, J.A.D., Maclennan, J. & Mosegaard, K., 2007. Joint inversion of seismic and gravity data for lunar composition and thermal state, *Geophys. J. Int.*, **168**(1), 243–258.
- Kirkpatrick, S., Gelatt, C.D.Jr & Vecchi, M.P., 1983. Optimization by simulated annealing, *Science*, **220**, 671–680.
- Laloy, E., Hraut, R., Lee, J., Jacques, D. & Linde, N., 2017. Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, *Adv. Water Resour.*, **110**, 387–405.
- Laloy, E., Hraut, R., Jacques, D. & Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network, *Water Resour. Res.*, **54**, 381–406.
- 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, *Math. Geosci.*, **44**(7), 783–803.
- Liu, J.S., 2002. *Monte Carlo Strategies in Scientific Computing*, Springer, ISBN-13: 978-0387952307.
- MacKay, D., 2003. *Information Theory, Inference and Learning Algorithms*, Cambridge Univ. Press, ISBN-13: 978-0521642989.
- Marinari, E. & Parisi, G., 1992. Simulated tempering: a new Monte Carlo scheme, *Europhys. Lett.*, **19**, 451–458.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **1**, 1087–92.
- Minsker, S., Srivastava, S., Lin, L. & Dunson, D.B., 2017. Robust and scalable Bayes via a median of subset posterior measures, *J. Mach. Learn. Res.*, **18**(1), 4488–4527.
- Mosegaard, K., 2012. Limits to nonlinear inversion, in *Applied Parallel and Scientific Computing. Proceedings of the PARA 2010 Meeting: Revised Selected Papers, Part I*, Lecture Notes in Computer Science, no. Part 1, Vol. **7133**, Springer, pp. 11–21.
- Mosegaard, K. & Sambridge, M., 2002. Monte Carlo analysis of inverse problems, *Inverse Problems*, **18**, R29–R54.
- Nawaz, M.A. & Curtis, A., 2019. Rapid discriminative variational Bayesian inversion of geophysical data for the spatial distribution of geological properties, *J. geophys. Res.*, **124**, 5867–5887.
- Neal, R.M., 2011. MCMC using Hamiltonian dynamics, in *Handbook of Markov chain Monte Carlo*, pp. 113–162, eds Brooks, S., Gelman, A., Jones, G. L. & Meng, X.-L., CRC Press.
- Neiswanger, W., Wang, C. & Xing, E.P., 2014. Asymptotically exact, embarrassingly parallel MCMC, in *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, pp. 623–632.
- Nemeth, C. & Fearnhead, P., 2020. Stochastic gradient Markov chain Monte Carlo, *J. Am. Stat. Assoc.*, **116**(533), 433–450.
- Nemeth, C. & Sherlock, C., 2018. Merging MCMC subposteriors through Gaussian-process approximations, *Bayesian Anal.*, **13**(2), 507–530.
- Rabinovich, M., Angelino, E. & Jordan, M.I., 2015. Variational consensus Monte Carlo, in *Advances in Neural Information Processing Systems*, pp. 1207–1215.
- Raj, A., Stephens, M. & Pritchard, J.K., 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets, *Genetics*, **197**(2), 573–589.
- Roberts, G.O. & Tweedie, R.L., 1996. Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli*, **2**(4), 341–363.
- Scheidt, C., Li, L. & Caers, J., 2018. *Quantifying Uncertainty in Subsurface Systems*, Wiley.
- Scott, S.L., Blocker, A.W., Bonassi, F.V., Chipman, H.A., George, E.I. & McCulloch, R.E., 2016. Bayes and big data: the consensus Monte Carlo algorithm, *Int. J. Manag. Sci. Eng. Manag.*, **11**(2), 78–88.
- Srivastava, S., Li, C. & Dunson, D.B., 2018. Scalable Bayes via barycenter in Wasserstein space, *J. Mach. Learn. Res.*, **19**(1), 312–346.
- Stuart, G.K., Minkoff, S.E. & Pereira, F., 2019. A two-stage Markov chain Monte Carlo method for seismic inversion and uncertainty quantification, *Geophysics*, **84**, R1015–R1032.
- Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *J. Geophys.*, **50**, 159–170.
- Tierney, L. & Mira, A., 1999. Some adaptive Monte Carlo methods for Bayesian inference, *Stat. Med.*, **18**, 2507–2515.
- Tolman, R.C., 1979. *The Principles of Statistical Mechanics*, pp. 48–51, Dover, ISBN 9780486638966.
- Vrugt, J.A., 2016. Environmental modelling and software, **75**, 273–316.
- Wolpert, D.H. & Macready, W.G., 1997. No free lunch theorems for optimization, *IEEE Trans. Evolut. Comput.*, **1**, 67–82.
- Ying, H., Mao, K. & Mosegaard, K., 2020. Moving target Monte Carlo, preprint arXiv:2003.04873.
- Yogatama, D., Wang, C., Routledge, B.R., Smith, N.A. & Xing, E.P., 2014. Dynamic language models for streaming text, *Trans. Assoc. Comput. Linguist.*, **2**, 181–192.
- Zhang, X. & A Curtis, A., 2020. Variational full-waveform inversion, *Geophys. J. Int.*, **222**(1), 406–411.
- Zunino, A., Mosegaard, K., Lange, K., Melnikova, Y. & Hansen, T., 2015. Monte Carlo reservoir analysis combining seismic reflection data and informed priors, *Geophysics*, **80**, R31–R41.

## ***E.2 Full-waveform Inversion by Informed-Proposal***

### ***Monte Carlo***

*Sarouyeh Khoshkholgh, Andrea Zunino, Klaus Mosegaard*

# Full-waveform Inversion by Informed-Proposal Monte Carlo

Sarouyeh Khoshkholgh<sup>1</sup>, Andrea Zunino<sup>1,2</sup> and Klaus Mosegaard<sup>1</sup>

<sup>1</sup> Department of Physics of Ice, Climate and Earth, Tagensvej 16, 2200 Copenhagen N, Denmark

<sup>2</sup> Now at: Department of Earth Sciences, ETH Zürich, Sonneggstrasse 5, 8092 Zrich, Switzerland

## SUMMARY

Markov Chain Monte Carlo (MCMC) sampling of solutions to large-scale inverse problems is, by many, regarded as being unfeasible due to the large number of model parameters. This statement, however, is only true if arbitrary, local proposal distributions are used. If we instead use a global proposal, informed by the physics of the problem, we can dramatically improve the performance of MCMC and even solve highly nonlinear inverse problems with vast model spaces. This is illustrated by a seismic full-waveform inverse problem in the acoustic approximation, involving close to  $10^6$  parameters.

## INTRODUCTION

Full-waveform inversion (FWI) is emerging as a promising method for computing subsurface properties and high-resolution images from seismic data. However, since its introduction in the late 1980s through the theoretical work of Lailly (1983) and Tarantola (1984, 1986, 1988) and subsequent numerical tests (*Gauthier et al.*, 1986; *Igel et al.*, 1996). it has been known that inference about the low-wavenumber components of the Earth model (the background velocity field) is a highly nonlinear problem. Early attempts to solve this problem using Markov-Chain Monte Carlo (MCMC) techniques (see, e.g., *Koren et al.*, 1991) ran into serious problems, not only because of the shortcomings of computational resources at the time, but also because large-scale inverse problems have vast parameter spaces.

Later improvements (*Bunks et al.*, 1995; *Pratt*, 1990; *Pratt*, 1999; *Virieux and Operto*, 2009), of which the frequency-domain approaches (*Pratt et al.*, 1998; *Pratt*, 1999; *Pratt et al.*, 1999) were amongst the most important, gave further momentum to the development of more efficient algorithms, but there were still some way to go before a probabilistic full-waveform inversion could be attempted. Characterizing the full posterior probability distribution, either through a sample of solutions, or in a parameterized form, is a formidable problem, far exceeding the calculation of best-fitting solutions.

Only recently, improved methods for many-parameter highly nonlinear seismic inversion have come to light, of which inversion based on Hamiltonian Monte Carlo (HMC) (*Fichtner et al.*, 2018; *Gebraad et al.*, 2020) and Variational Full-Waveform Inversion (VFWI) (*Zhang & Curtis*, 2020) are notable examples. HMC is an MCMC with improved sampling efficiency using gradient information of the misfit function - information that can be computed relatively fast for seismic inverse problems through the method of adjoints. VFWI is a non-sampling method seeking a continuous mapping over the parameter space that transforms samples from the prior probability density into approximate samples from the posterior.

A close look at all methods (deterministic and stochastic) used for solution of inverse problems (linear, weakly nonlinear and strongly nonlinear) reveals that efficiency always requires specific properties/assumptions about the problem built into the algorithm: In regular MCMC, good performance depends heavily on smoothness information obtained through initial experimentation with step lengths of the algorithm. In HMC, efficient computation of misfit gradients through adjoints, which is a specific property of the wave equation, is built into the algorithm. VFWI assumes a predefined family of posterior distributions among which it seeks a solution. The role of information in defining the efficiency of an inversion algorithm is further explored in (*Khoshkholgh et al.*, 2020) where it is shown that goal-directed use of specific information, characteristic of the problem, can dramatically improve performance.

In this study we propose a MCMC methodology for large-scale full-waveform inversion, based on the Informed Proposal Monte Carlo (IPMC) technique described by *Khoshkholgh et al.* (2020). Our approach in this paper is pragmatic, in that we use approximate information acquired through classical seismic processing and subsequent interpretation to build a proposal strategy for the MCMC sampler. IPMC ensures that errors and inaccuracies in the proposal information does not pollute the final sampling result. Only the speed by which statistically independent samples are collected is influenced: The closer the proposal is to the posterior, the faster is the sampling.

First we provide a brief overview of the Informed Proposal Monte Carlo method, and then we apply the method to a synthetic 2D test example and compare the results with those obtained by a regular MCMC method.

## METHODS

### Probabilistic Inversion

Probabilistic inverse theory is based on the assumption that any state of information about a parameterized system can be described by a probability density function. In the two most widely used formulations, the Bayesian approach (Bayes, 1764; Jackson & Matsu'ura, 1985) and the approach of Tarantola and Valette (1982), the outcome of the inversion is a so-called posterior probability distribution  $\sigma(\mathbf{m})$  over the model parameter space. In both formulations,  $\sigma(\mathbf{m})$  is expressed as a product of two distributions

$$\sigma(\mathbf{m}) = \rho(\mathbf{m})L(\mathbf{m}) \quad (1)$$

where  $\rho(\mathbf{m})$  quantifies our (uncertain) prior information about  $\mathbf{m}$ , and  $L$  contains information about  $\mathbf{m}$  from uncertain, observed data and from physical law. Since a probabilistic inversion produces, not only one particular solution, but a probability distribution over the entire model space, we are left with the problem of characterizing this distribution. One choice is to look for a *parametric* description of the posterior (as is done in variational methods (Zhang & Curtis, 2020)), but another approach is to create a *non-parametric* description of the posterior. This method, which is essentially equivalent to forming a multi-dimensional histogram of the posterior, is the one used by MCMC- and other sampling methods.

### Markov-Chain Monte Carlo

The goal of an MCMC sampler is to produce a collection of models with a sampling density proportional to a given target distribution. Each step of the sampling from a probability density  $\sigma(\mathbf{m})$  proceeds from a current  $\mathbf{m}$  by first randomly proposing a new model  $\mathbf{m}'$  according to a so-called *proposal distribution*  $q(\mathbf{m}'|\mathbf{m})$ , followed by accepting  $\mathbf{m}'$  only with probability

$$P_{\text{acc}}^{\mathbf{m} \rightarrow \mathbf{m}'} = \min \left( \frac{\sigma(\mathbf{m}')q(\mathbf{m}|\mathbf{m}')}{\sigma(\mathbf{m})q(\mathbf{m}'|\mathbf{m})}, 1 \right). \quad (2)$$

This acceptance probability guarantees that, in the limit where the number of models  $N \rightarrow \infty$ , the distribution  $\sigma(\mathbf{m})$  will be correctly sampled. There is great freedom in the choice of the proposal distribution  $q(\mathbf{m}'|\mathbf{m})$ , as long as equation (2) is well defined for all  $\mathbf{m}'$ . It is, however, clear from (2) that if we could choose  $q(\mathbf{m}'|\mathbf{m}) = \sigma(\mathbf{m}')$  we would have an algorithm with maximum efficiency: It would be allowed to move freely between models of non-zero values of the target probability density, and all moves would be accepted. It is, however, also clear that this choice of  $q$  would require that we already had full knowledge of the structure of  $\sigma$  and hence that the solution would be known from the beginning!

In practice, the situation is quite the opposite. We usually have very little information about  $\sigma$  or, to simplify implementation of our algorithm, we ignore this information. A commonly used, minimum information about  $\sigma$  used in an MCMC implementation is about the smoothness of  $\sigma$ , and it is normally found by experimentation. The smoothness tells us how far away from the current point  $\mathbf{m}$  in the model space we can go without changing the value of  $\sigma$  significantly, and this allows us to build the proposal  $q$  so narrow that any new proposed model have a good chance of being accepted. The advantage of a narrow  $q$  is that it is quite similar (near-proportional) to  $\sigma$  in a small neighborhood around  $\mathbf{m}$ . For this reason it gives a high acceptance probability, but the disadvantage is that the resulting moves away from  $\mathbf{m}$  are so small that successive, accepted models become highly correlated. Consequently, it takes many moves to produce new, uncorrelated models.

It is the above-mentioned experience with information-deficient, narrow proposals that leads many to conclude that MCMC is inherently inefficient. We view the situation quite differently: MCMC is an algorithm that allows (possibly slow) sampling of a distribution  $\sigma$ , even in cases where the approximate smoothness is the only thing known about  $\sigma$ . If, on the other hand, we have more comprehensive information about  $\sigma$ , this can be built into the proposal  $q$ , and the algorithm can be made much more efficient. The purpose of this study is to demonstrate that the latter approach is very efficient. We use a semi-realistic, synthetic, seismic inversion example where inference about  $\approx 10^6$  model parameters is sought.

### Using a Proposal, Informed by Approximate Physics

Following the formulation of Tarantola and Valette (1982), a general expression for the joint posterior probability can be written

$$\sigma(\mathbf{d}, \mathbf{m}) = \frac{\rho(\mathbf{d}, \mathbf{m})\theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})} \quad (3)$$

where  $\mathbf{d}$  is data,  $\mathbf{m}$  is the model parameters,  $\rho(\mathbf{d}, \mathbf{m})$  is the prior and  $\mu(\mathbf{d}, \mathbf{m})$  is the homogeneous probability density in the joint  $(\mathbf{d}, \mathbf{m})$ -space. The density  $\theta(\mathbf{d}, \mathbf{m})$  is the distribution of errors/uncertainties of the relation between  $\mathbf{m}$  and  $\mathbf{d}$ , including data uncertainties and possible, physical modelization errors. In this study we will, without loss of generality, assume that  $\mu$  is constant over the model space, and that  $\rho(\mathbf{d}, \mathbf{m}) = \rho(\mathbf{d})\rho(\mathbf{m})$  (prior information about  $\mathbf{d}$  and  $\mathbf{m}$  are independent). This leads to (ignoring the normalization constant):

$$\sigma(\mathbf{d}, \mathbf{m}) = \rho(\mathbf{m})\rho(\mathbf{d})\theta(\mathbf{d}, \mathbf{m}) \quad (4)$$

Following (Khoshkholgh *et al.*, 2020), we now build our proposal distribution as a rough approximation  $\tilde{\sigma}_m(\mathbf{m})$  to the marginal posterior in the model space.  $\tilde{\sigma}_m$  will be based on simplified physics where the modelization errors overwhelm the observational data uncertainties, hence (ignoring nor-

malizations):

$$\tilde{\sigma}_m(\mathbf{m}) \approx \sigma(\mathbf{d}_{obs}, \mathbf{m}) \approx \rho(\mathbf{m})\theta(\mathbf{d}_{obs}, \mathbf{m}) . \quad (5)$$

expressing the approximate posterior as a product of the prior and the approximate likelihood  $\theta(\mathbf{d}_{obs}, \mathbf{m})$ .

In the following we will use  $q(\mathbf{m}'|\mathbf{m}) = \tilde{\sigma}_m(\mathbf{m}')$  as a global proposal distribution to speed up the MCMC algorithm. Even if  $q$  is only a poor approximation to the likelihood  $L$ , it will not bias the final sampling result. Only the efficiency of the algorithm will be influenced.

In this study we obtain  $q(\mathbf{m}'|\mathbf{m})$  through traditional processing and interpretation of the seismic data  $\mathbf{d}$ . This procedure is in the following formally denoted  $h$ , a pseudo-inverse operator mapping from data space into the model space:  $\tilde{\mathbf{m}} = h(\mathbf{d})$ . For real data this may include muting, velocity analysis, normal moveout correction, stacking, deconvolution, migration, and subsequent interpretation. In our illustrative synthetic example we simplify this procedure.

### Constructing the Approximate Likelihood

In our study, we take a simple approach and construct  $\theta(\mathbf{d}_{obs}, \mathbf{m})$  as follows:

- (i) Using a simplified approach, we construct an approximate solution  $\tilde{\mathbf{m}}$  to the inverse problem.
- (ii) The "true" modelization error is  $\delta\mathbf{m}_{true} = \tilde{\mathbf{m}} - \mathbf{m}_{true}$  but since  $\mathbf{m}_{true}$  is unknown, we instead compute an approximation to the modelization error:

$$\delta\mathbf{m}_{approx} = h(g(\tilde{\mathbf{m}})) - \tilde{\mathbf{m}},$$

where  $g$  is the forward function used to compute synthetic data. Since  $h(g(\tilde{\mathbf{m}}))$  is an approximate solution to the inverse problem with  $g(\tilde{\mathbf{m}})$  as data, the above formula estimates what the modelization error would have been if  $\tilde{\mathbf{m}}$  had been the true model. In case  $\tilde{\mathbf{m}}$  is close to  $\mathbf{m}_{true}$ , we expect that  $\delta\mathbf{m}_{approx}$  will be close to  $\delta\mathbf{m}_{true}$ .

- (iii) We now define now  $\theta(\mathbf{d}_{obs}, \mathbf{m})$  as a Gaussian with mean  $\tilde{\mathbf{m}}$  and a diagonal covariance matrix where the  $n$ 'th standard deviation is equal to the  $n$ 'th component of  $\delta\mathbf{m}_{approx}$ .

### Informed Proposal Sampling

Given that  $S(\mathbf{m})$  is the (accurately computed) misfit function for the problem, we can now use an MCMC algorithm to sample the posterior, using  $\rho(\mathbf{m})$  as the prior,  $L(\mathbf{m}) = \exp(-S(\mathbf{m}))$  as the accurately computed likelihood, and  $q(\mathbf{m}'|\mathbf{m}) = \rho(\mathbf{m}')\theta(\mathbf{d}_{obs}, \mathbf{m}')$  as the proposal. We assume here that values of the prior can be computed explicitly, so our proposal is now fully defined.

Each step of the algorithm now runs as follows:

- (i) Perturb the current model  $\mathbf{m} \rightarrow \mathbf{m}'$  using the proposal  $q(\mathbf{m}'|\mathbf{m})$ .
- (ii) Accept/reject the perturbed model  $\mathbf{m}'$  with probability

$$P_{\text{acc}}^{\rho} = \min \left( \frac{\rho(\mathbf{m}')L(\mathbf{m}')q(\mathbf{m}|\mathbf{m}')}{\rho(\mathbf{m})L(\mathbf{m})q(\mathbf{m}'|\mathbf{m})}, 1 \right) = \min \left( \frac{L(\mathbf{m}')\theta(\mathbf{d}_{\text{obs}}, \mathbf{m})}{L(\mathbf{m})\theta(\mathbf{d}_{\text{obs}}, \mathbf{m}')}, 1 \right). \quad (6)$$

If  $\mathbf{m}'$  is rejected, the current model  $\mathbf{m}$  will be repeated.

In practice, the rate of accepted models found by the above algorithm can be improved if we introduce a burn-in period (say, within the first  $N_B$  iterations) in the sampling procedure where we replace the approximate model  $\theta(\mathbf{d}_{\text{obs}}, \mathbf{m})$  with better-fitting models discovered in the process. We will see an example of this in the numerical example below.

## RESULTS

### Model Parameters and Data

The above-mentioned method was applied to acoustic full waveform inversion of synthetic reflection data from a subset of the Marmousi velocity model (*Versteeg, 1994*) giving P-wave velocities in a 755 x 1255 Cartesian grid (= 947525 parameters) with a grid size of 1.5 m in vertical and horizontal directions (see Figure (1)). Two shot records were generated by a finite difference algorithm to solve the constant-density, variable-velocity acoustic wave equation in two dimensions (*CREWES Library: Youzwishen, 1999, and Margrave, 2000*). The algorithm uses second-order finite-difference operators for the time derivative and the Laplacian operator and applies simple, absorbing boundary conditions. Figure (2) shows one of the two shot records used as observed data in our study. The dominant frequency of the Ricker wavelet related to the source term is 40 Hz. The seismic sources were located at  $x = 375$  m and  $x = 1500$  m, and receivers were located at the surface, equally distributed with distances of 7.5 meters.

### The Likelihood Function

Assuming Gaussian white noise on the data, we use the likelihood function

$$L(\mathbf{m}) = \exp(-\|\mathbf{d} - g(\mathbf{m})\|/\sigma^2)$$

where  $g$  is the forward function calculating synthetic data from a model  $\mathbf{m}$ , and  $\sigma$  is the standard deviation of the noise.  $\sigma$  was chosen to give a signal-to-noise ratio  $S/N \approx 2.0$ . Synthetic noise was not added to the data.

### The Prior Distribution

In our study, we use a prior probability density assigning non-zero probability only to piecewise constant velocity models that can be derived from:

(i) Smooth, continuous 1-1 deformations ('warpings') of the approximate model. A warping is defined by a random, smooth displacement field  $\mathbf{u}$  with  $|\partial u_i / \partial x_i| < 1$  where  $x_i$  ( $i = 1, 2$ ) are the image location coordinates. Warping takes place in a randomly centered square window with a predefined dimension (here  $250 \times 250$ ), a cosine variation of displacement coordinates in both vertical and horizontal direction, and with zero displacement at the boundary. Any such warping will result in a new, piecewise constant velocity model.

(ii) Choosing new velocities in each layer within given fixed intervals.

All models that can be generated according to the above rules are assumed a priori equally likely. Figure 3 shows four sample models from the prior. It is seen that the prior allows considerable variations in the model, but retains its basic 'topology'.

### Building the Proposal Distribution

Our proposal distribution is a rough approximation to the posterior, which in turn is a product of the prior and a Gaussian, approximate likelihood. The approximate likelihood is centered at the approximate solution to the problem and with standard deviations proportional to the modelization errors of each model parameter. Hence, our proposal sampler is based on four components:

(i) **A rough approximation to the posterior.** To find this, we assume that we have a 2D seismic reflection profile across the area, from which we will derive a rough subsurface model through classical processing and interpretation. To simulate this situation, we generate zero offset seismic data using the exploding reflector model (*Loewenthal et al., 1976*). The first step in our data processing is a depth migration using a rough background velocity model derived from the true velocity model through Gaussian smoothing (Figure 4), followed by computing an approximate reflectivity profile through frequency domain spiking deconvolution (Figure 5). Then we simulate seismic interpretation by identifying clearly visible reflectors, and after combining this interpreted image with the long-wavelength migration velocity field, we arrive at an approximate, homogeneous-layer velocity model for the area (Figure 6). This velocity model is a rough solution  $\tilde{\mathbf{m}}$  to the full-waveform inverse problem.

(ii) **Standard deviations of the approximate likelihood.** The difference between the exact likelihood and the approximate likelihood is obtained by generating synthetic reflection data from the approximate model, and then performing a rough inversion of this data through processing and interpretation to obtain a second (and even less accurate) approximate model. Subtracting this from the

(first) approximation allows estimation of the modeling error, providing standard deviations of our Gaussian, approximate likelihood. The spatial envelope of the velocity differences at each pixel in the model, computed through a Hilbert transform, is shown in Figure 7 and reveals regions of large and small modelization errors.

(iii) **A warping scheme that allows deformation of a model into a new model with nonzero prior probability.** Our warping consists of a series of deformations, each in a quadratic window centered at the point chosen for maximum perturbation. The size of the maximum displacement is random and is adjusted such that, in that point, the change will be no larger than the estimated modelization error at that point. The displacement field in each window is smooth and has random directions, and at the boundary of each window the displacement is zero. We chose 100 quadratic windows of size 1/16 of the model size for random warping.

(iv) **A simple velocity perturbation scheme that allows a random change of velocity within one layer,** within given bounds (here  $\pm 5\%$ ). The size of the velocity perturbation is adjusted such that, in any point, the change will be limited by the estimated modelization error.

This composition of the proposal algorithm ensures that the sampling is guided by (1) information about the approximate solution and its errors, and (2) information about the structure of the solution (piecewise constant).

### Producing Samples from the Posterior Distribution:

We use the approximate model derived from processing and interpretation as the starting model from our sampling. In 30 percent of the perturbations the velocity of a random layer is perturbed and the layer boundaries are kept constant. In the rest of the perturbations we perturb the layer boundaries. During the first 300 iterations we replaced the approximate model  $\tilde{\mathbf{m}}$  (which is the center of the proposal distribution) with the best-fitting model found so far by the algorithm, and adjusted the modeling error accordingly. In this way we have deliberately introduced a burn-in sequence of 300 iterations in order to improve the acceptance rate of the sampler.

Figure 8 shows four realizations from the posterior probability density produced after 200, 400, 600 and 800 iterations of the IPMC algorithm. Figure 9 compares the convergence of IPMC with a regular Extended Metropolis Algorithm (EMA) (Mosegaard and Tarantola, 1995). The EMA starts in a random realization from the prior, and its proposal distribution uses the same prior as the IPMC. However, as is customary for the EMA, the proposal establishes a random walk with a limited steplength:

$$q(\mathbf{m}'|\mathbf{m}) = \rho(\mathbf{m}')u(\mathbf{m}'|\mathbf{m}) \quad (7)$$

where  $u(\mathbf{m}'|\mathbf{m})$  is a uniformly sampling random walk. Note that, in contrast to the IPMC, the EMA proposal does not use the approximate likelihood  $\theta(\mathbf{d}_{obs}, \mathbf{m}')$ .

Looking at the log-likelihood curves of the two methods (Figure 9) we see a significant difference between their convergence properties. The IPMC algorithm starts advantageously at a reasonable approximation to the solution with a relatively high value of the log-likelihood. From there it proceeds towards equilibrium, which is attained after around 100 iterations. The log-likelihood curve for the MCMC is, however, showing long-term correlations throughout the 3000 iterations. Even the models sampled after 3000 iterations (not shown) show no similarity to the true model, but rather resembles white noise. The proposed IPMC method shows a remarkable improvement in efficiency as compared to the regular MCMC algorithm.

## DISCUSSION

It is often claimed that Markov-Chain Monte Carlo methods are highly inefficient for solution of large-scale inverse problems. This statement is only true if simple, local proposal mechanisms are used. In this study we show that a global proposal, incorporating substantial external information about the problem (from approximate physics, and from knowledge about the character of the solution), dramatically changes the situation and allows MCMC to equilibrate much faster. In our case with  $\sim 10^6$  model parameters, the Informed Proposal Monte Carlo converged to equilibrium within  $\sim 100$  iterations whereas the classical local-proposal MCMC was unable to approach any close-to-equilibrium sampling within this time frame.

As always, when comparing algorithms, or differently tuned versions of the same algorithm, one has to be critical about the conditions under which the comparison was carried out. Here, it is easy to see what the difference between IPMC and MCMC is. The IPMC algorithm suggested in this paper is a specialized MCMC algorithm using much more external information about feasible solutions than the regular MCMC implementation. Knowing in this case that the solution is near-piecewise constant is one important piece of information, and knowing that classical processing and interpretation will produce a reasonable solution is another. Since our approximate likelihood enters only via our proposal and assigns nonzero probability to all models, it does not asymptotically bias our solution and is therefore risk-free to use (it does not exclude any models). It can only, in the best case, speed up the sampling. On the other hand, our prior has the special property that it excludes (assigns zero probability) to all non-piecewise-constant models. For this reason it actually biases the solution. However, this is desirable (and intentional), because we want the prior information to have an imprint on the solutions.

When interpreting our numerical results, one should remember the following:

(i) The convergence speeds of both the IPMC and the regular MCMC depend on the noise variance of the data. The lower the variance, the slower the convergence.

(ii) Our data consist only of two shot records. For this reason the solution is not strongly constrained by the data. This is apparent on the four realizations from the posterior probability density shown in Figure 8. More data would make the solution more well-determined and would probably slow down the sampling process.

(iii) The proposed IPMC uses a global proposal, and for this reason can generate independent samples quite fast. In our numerical example we could produce uncorrelated samples separated by  $\sim 200$  iterations. This, however, does not mean that a complete description of the posterior can be obtained within short time. More than  $10^6$  samples may be required to completely describe a posterior defined in a  $10^6$ -dimensional model space, and this is independent of the sampling algorithm used.

## CONCLUSIONS

In this study we have investigated how external physical information can be used to establish a proposal distribution for efficient MCMC sampling of solutions to an inverse problem. Our test problem was an acoustic full-waveform inverse problem where earth models consistent with data from two seismic shots we produced. Our aim have been to simulate a practical situation where a preliminary, approximate subsurface model created from processing and interpretation of reflection data was available. The preliminary model was used in two different, independent ways: (1) To define a prior distribution assigning nonzero probability only to models consisting of a stack of homogeneous layers intersected by a major fault, and (2) to quantify the modelization error of combined processing and interpretation and to use this information to build a global proposal distribution centered at the preliminary model and with a dispersion proportional to the modelization error. Our study showed how informed proposal distributions can have significant impact on the computational speed of Monte Carlo sampling of solutions to inverse problems.

## ACKNOWLEDGEMENTS

This work was supported by Innovation Fund Denmark through the OPTION Project (5184-00025B). Klaus Mosegaard would like to thank Dr. Amir Khan and colleagues at the Department of Earth Sciences, ETH, for their hospitality and inspiring discussions during the fall 2017 where this work was initiated.

**Data availability**

No new data were generated or analysed in support of this research.

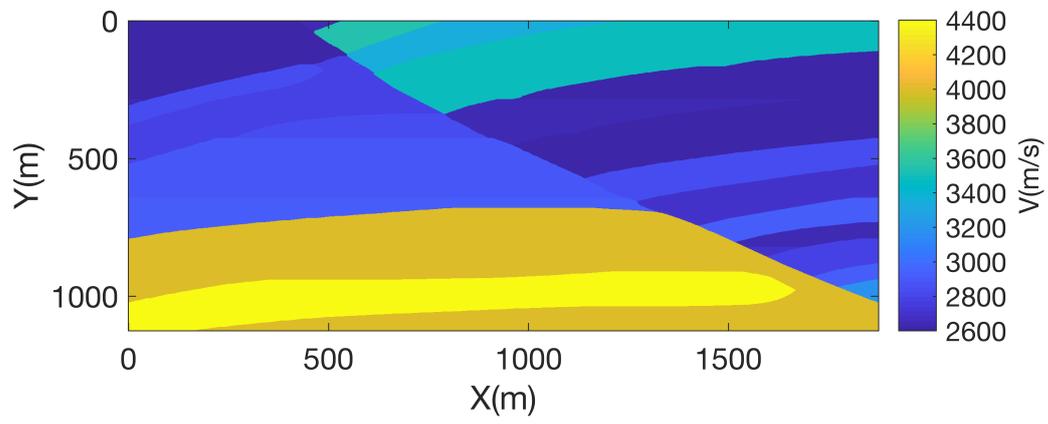


Figure 1. True velocity model

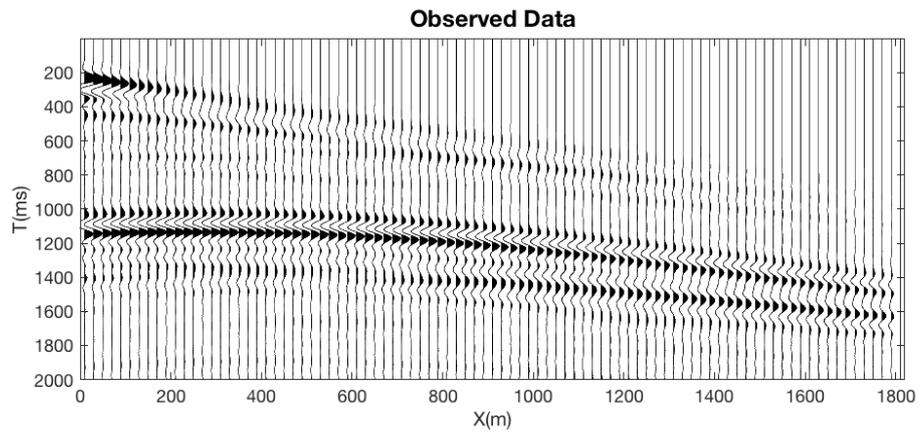


Figure 2. Synthetic shot record generated from true velocity model

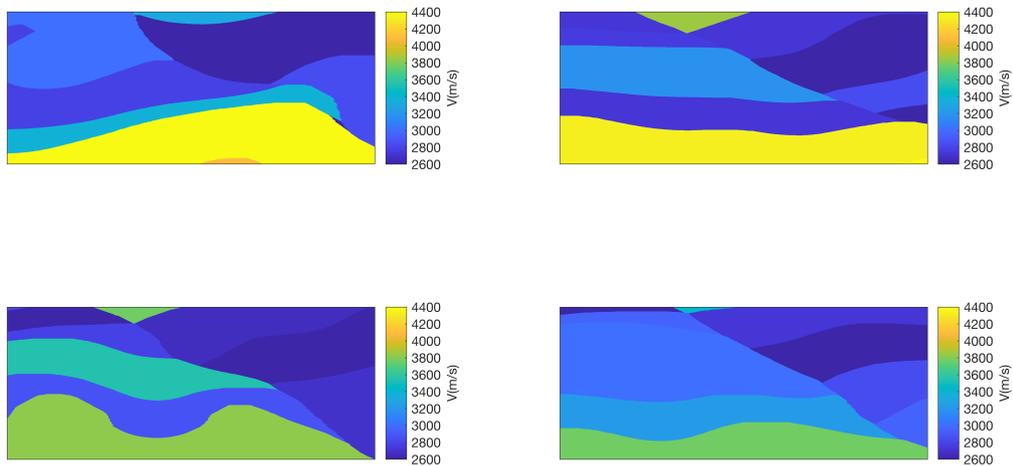


Figure 3. Four realizations from the prior

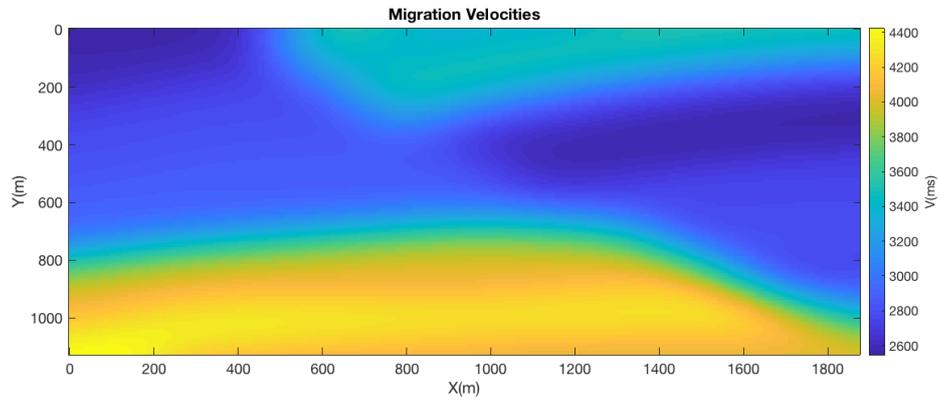


Figure 4. Migration Velocities

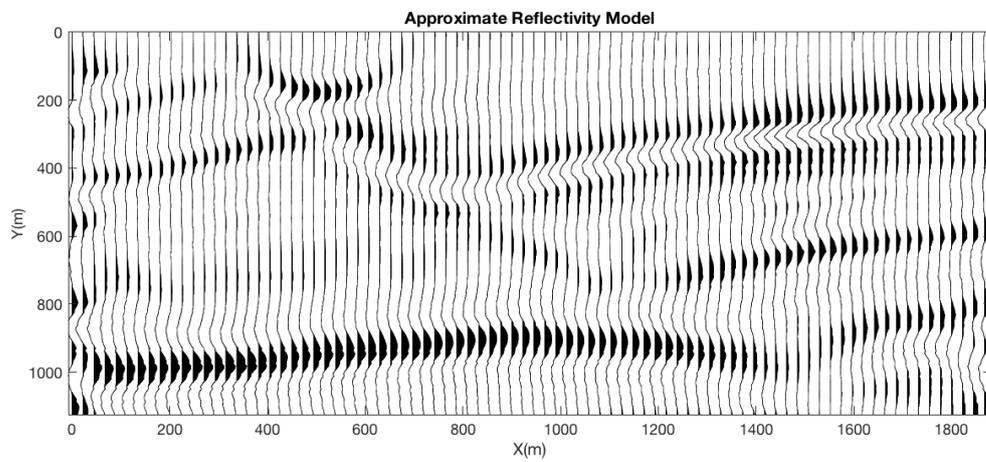
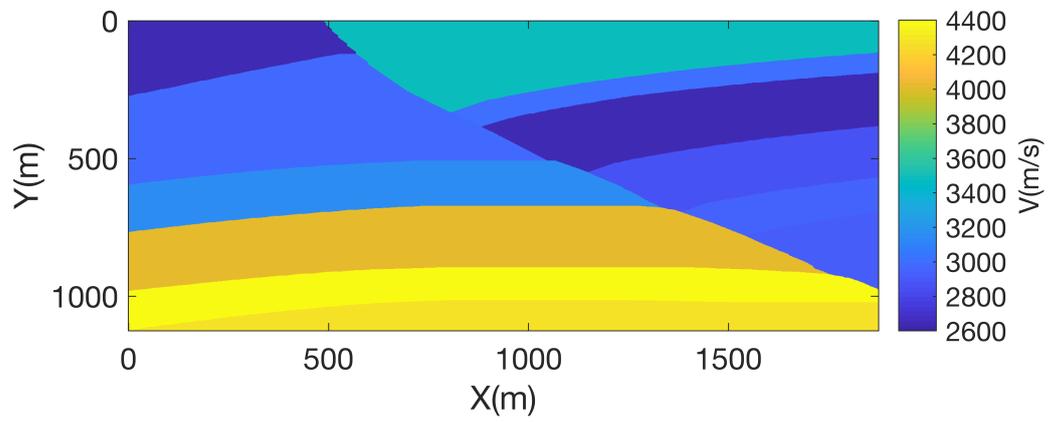
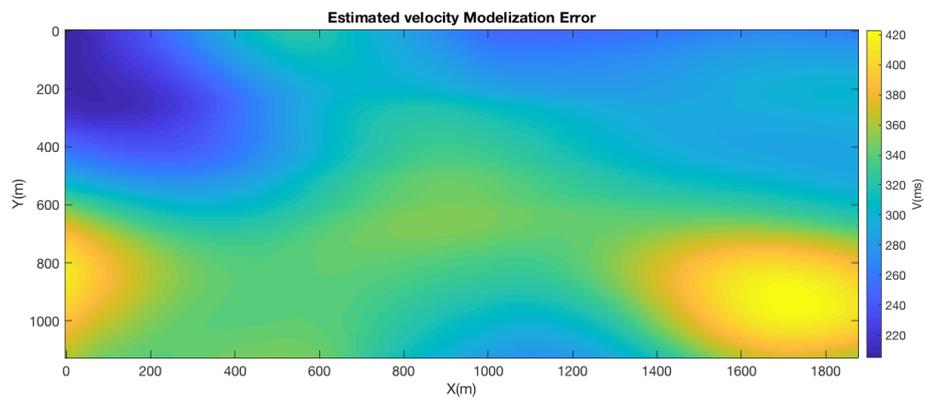


Figure 5. Approximate Reflectivity



**Figure 6.** Approximate velocity from interpretation



**Figure 7.** Error envelope obtained from approximation error

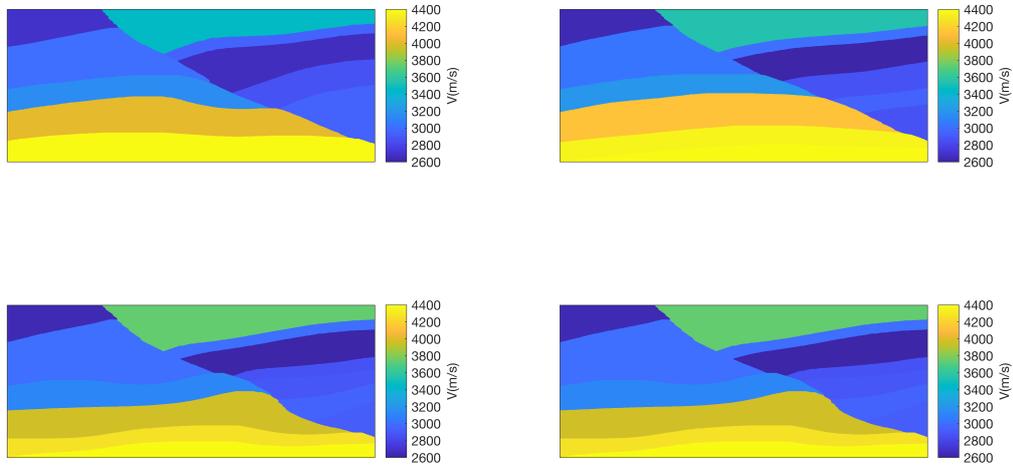


Figure 8. Four realizations from the posterior

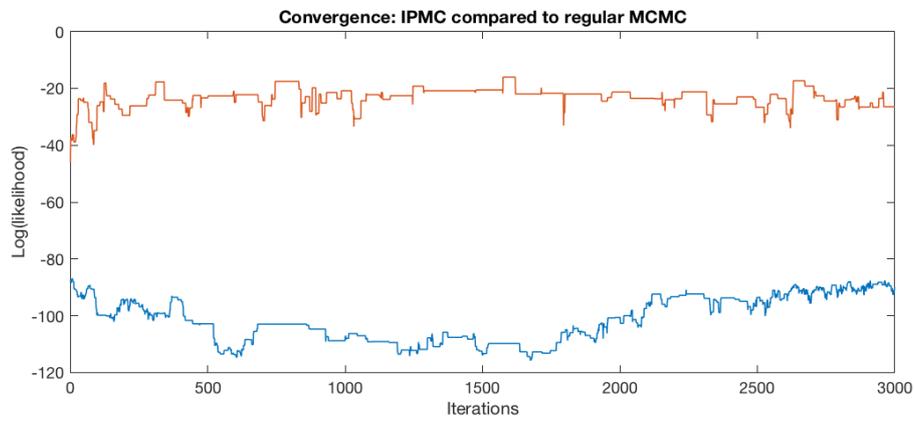


Figure 9. Likelihood vs number of iterations for IPMC compared to a regular MCMC

## REFERENCES

- Bayes, T., 1764. An essay towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc.*, **53**, 370–418. Reprinted, with biographical note by G. A. Barnard, 1958, in *Biometrika*, **45**, 293–315.
- Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, 60, 14571473.
- Fichtner, A., Zunino, A., and Gebraad, L., 2018. Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophysical Journal International*, **216**, 1344-1363.
- Gauthier, O., Virieux, J., and Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results., *Geophysics*, 51, 13871403.
- Gebraad, L., Boehm, C., and Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. Geophys. Res.*, 125.
- Igel, H., Djikpesse, H., and Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poissons ratio., *Geophys. J. Int.*, 124, 363371.
- Khoshkholgh, S., Zunino, A. and Mosegaard, K., 2020. Informed Proposal Monte Carlo. arXiv:2005.14398
- Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo Estimation and Resolution Analysis of Seismic Background Velocities: *Journal of Geophysical Research*, **96**, B12.
- Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in Conference on Inverse Scattering: Theory and Application, Soc. Industr. appl. Math., Philadelphia, PA.
- Loewenthal, D., Lu, L., Roberson, R., and Sherwood, J.W.C., 1976, The wave equation applied to migration: *Geophys. Prosp.*, 24, 380399.
- Jackson, D. D., and Matsu'ura, M., 1985. A Bayesian approach to nonlinear inversion, *J. geophys. Res.*, **90**, 581–591.
- Pratt, R., 1990, Frequency-domain elastic modeling by finite differences: a tool for crosshole seismic imaging: *Geophysics*, 55, 626632. doi:10.1190/1.1442874
- Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model: *Geophysics*, 64, 888901. doi:10.1190/1.1444597
- Pratt, R., Shin, C., and Hicks, G. J., 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, 133, 341362. doi:10.1046/j.1365-246X.1998.00498.x
- Pratt, R., and Shipp, R., 1999, Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data: *Geophysics*, 64, 902914. doi:10.1190/1.1444598
- Tarantola, A., and B. Valette, 1982, Inverse Problems 1?4 Quest for Information: *Journal of Geophysics*, 50, 159?170.?
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, 1893–1903.

- Tarantola, A., 1988, Theoretical Background for the Inversion of Seismic Waveforms, Including Elasticity and Attenuation: *Pure and Applied geophysics* **128**, 1/2, 365–399.
- Versteeg, R., 1994. The Marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge* 13 (9): 927936. doi:10.1190/1.1437051
- Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, 74, WCC1WCC26. doi:10.1190/1.3238367
- Zhang X., A Curtis, A., 2020. Variational full-waveform inversion. *Geophysical Journal International*, **222**,1, 406-411.

*E.3 Evolution of the Stress and Strain field in the Tyra field during the Post-Chalk Deposition and Seismic Inversion of fault zone using Informed-Proposal Monte Carlo*

*Sarouyeh Khoshkholgh, Ivanka Orozoa-Bekkevold, Klaus Mosegaard*

---

*Evolution of the Stress and Strain field  
in the Tyra field during the Post-Chalk  
Deposition and Seismic Inversion of  
fault zone using Informed-Proposal  
Monte Carlo*

---

Sarouyeh Khoshkholgh, Ivanka Orozova-Bekkevold and Klaus Mosegaard  
Niels Bohr Institute  
University of Copenhagen

September, 2021

## Table of Contents

---

<b>Introduction</b> .....	<b>2</b>
<b>Well Data</b> .....	<b>3</b>
<b>Modelling of the Post-Chalk overburden in the Tyra field</b> .....	<b>5</b>
Numerical representation of the Post-Chalk Deposition in the Tyra field.....	6
Finite elements modelling of the Post-Chalk Deposition in the Tyra field.....	9
<b>Results: Evolution of the Stress and Strain state in the Tyra Field during the Cenozoic Deposition</b> .....	<b>12</b>
<b>Seismic Study of the Fault Detection Limit</b> .....	<b>13</b>
Seismic Inversion.....	14
<b>Results: Seismic Inversion</b> .....	<b>17</b>
<b>Conclusions and Discussions</b> .....	<b>22</b>
<b>Acknowledgements</b> .....	<b>23</b>
<b>References</b> .....	<b>24</b>

## Abstract

---

When hydrocarbon reservoirs are used as a CO<sub>2</sub> storage facility, an accurate uncertainty analysis and risk assessment is essential. An integration of information from geological knowledge, geological modelling, well log data, and geophysical data provides the basis for this analysis. Modelling the time development of stress/strain changes in the overburden provides prior knowledge about fault and fracture probability in the reservoir, which in turn is used in seismic inversion to constrain models of faulting and fracturing. One main problem in solving large scale seismic inverse problems is high computational cost and inefficiency. We use a newly introduced methodology - Informed-proposal Monte Carlo (IPMC) - to deal with this problem, and to carry out a conceptual study based on real data from the Danish North Sea. The result outlines a methodology for evaluating the risk of having sub-seismic faulting in the overburden that potentially compromises the CO<sub>2</sub> storage of the reservoir.

## Introduction

---

Sequestration of CO<sub>2</sub> in former oil and gas reservoirs can contribute to amelioration of the global increase in CO<sub>2</sub> emissions, and it is already in use in a limited number of sites worldwide (Bachu, 2008; Michael et al., 2010; Ringrose et al., 2017). Injection of CO<sub>2</sub> for enhanced oil recovery has already been utilised by the oil industry for decades, particularly in onshore North America (Gozalpour et al., 2005). CO<sub>2</sub> is currently stored offshore Norway, in Sleipner and Snøhvit, with  $\sim 1.5 \cdot 10^6$  tonnes annually (Eiken et al., 2011). Pilot-projects have been carried out in Germany (Kempka & Kühn, 2013; Bergmann et al., 2016), Spain (Vilamajó et al., 2013; Ogaya et al., 2013) and Texas (Daley et al., 2008; Doughty et al., 2008), and this has greatly increased our understanding of CO<sub>2</sub> migration, monitoring and injection strategies in geological reservoirs.

Injection of CO<sub>2</sub>, or any other fluid, into subsurface reservoirs might increase the risk of caprock failure and migration of fluid along faults, fracture corridors, and other pre-existing weak zones (Ogata et al., 2014a). Understanding the detection thresholds of such structures calls for careful mechanical modelling of the reservoir stress- and strain field, careful inversion of available seismic data, combined with geological knowledge from well data and outcrop data. In this way it may be possible to quantify

the probability of significant CO<sub>2</sub> migration through the caprock, laying the ground for a meaningful risk evaluation.

During the last couple of decades considerable progress has been made in geophysical and geostatistical data analysis methods to correctly estimate model uncertainties and thereby to evaluate fault detection thresholds (see, e.g., Zunino et al., 2015). The goal of this pilot project is to propose a way of exploiting these methods for risk assessment in connection with CO<sub>2</sub> storage. Conceptual models are developed to model the time evolution of the subsurface, giving information about current and future stress fields resulting from geological processes, such as sediment deposition for example. This analysis provides the prior information to a subsequent probabilistic inversion of seismic data. Monte Carlo methods are used to simulate the noise in the data, and the noise is back-propagated through the geophysical (e.g., seismic) equations into the geophysical model, generating a model variability, and reflecting the uncertainty of the reservoir structure. Combining this approach with prior information about the mechanical properties of the reservoir, we evaluate the probability of fault migration scenarios. In this pilot project, we carry out a concrete, highly simplified numerical study of the sub-problem of estimating the density of sub-seismic faults in the overburden of an existing North Sea hydrocarbon reservoir, and established a simple probability model for releases through existing faults. The study is a starting point for developing a full-scale risk analysis system based on the principles outlined above.

### Well data

---

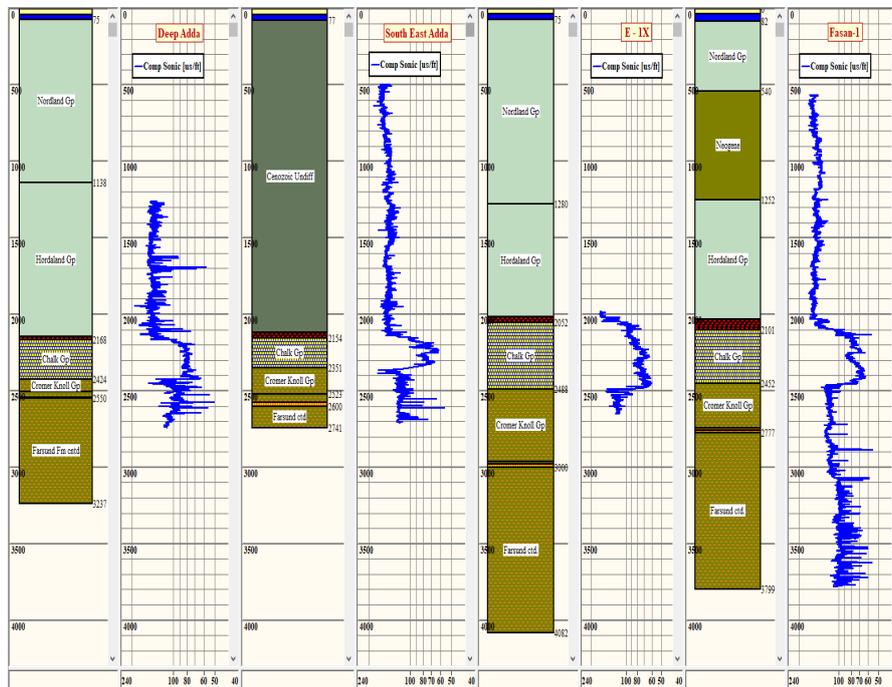
Since the overburden provides the reservoir seal and hosts significant part of the infrastructure, it is important to have sufficient data of good quality in order to analyse its integrity and/or strength. One very important parameter is the fracture pressure of the seal – if the pressures at the top of the reservoir exceeds the fracture pressure of the seal a breach will occur and the reservoir fluid (CO<sub>2</sub> or hydrocarbons) will escape.

Data (petrophysical logs and well reports) for the following wells: Fasan-1, Deep Adda, South East Adda and E1-X (Figure 1), were provided by DHRTC (Danish Hydrocarbon Research and Technology Centre). The wells are located in the Tyra field, in the Danish sector of the Central Graben, approximately 200 km west from the city of Esbjerg.

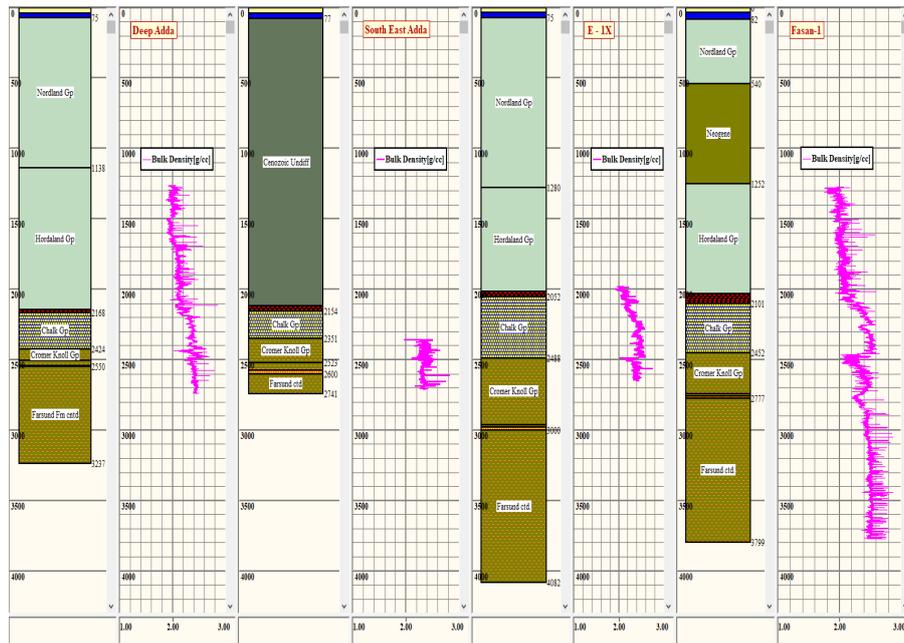


**Figure 1.** Location of the wells used in the analysis.

Figure 2 and 3 show the lithology columns and, respectively, the depth coverage of the sonic and density logs acquired in the four above mentioned wells. The figures illustrate very well the challenges related to the availability of log data in the overburden. In all four wells, no sonic (or density) data were acquired in the shallow section from sea bed to approximately 500 m depth. In the well E-1X, both sonic and density logs were acquired only in the reservoir (chalk) section, while in South East Adda, density data are available only below the Chalk group, in the deepest section of the well. This probably is due to the fact that the purpose of the South East Adda well was to investigate the hydrocarbon potential of the Cromer Knoll Group (Lower Cretaceous).



**Figure 2.** Overview of the sonic compressional logs (us/ft), acquired in the four wells. From left to right: Deep Adda, South East Adda, E-1X and Fasan-1. The depth reference is Rotary Table (RT, i.e. the rig floor).



**Figure 3.** Overview of the density logs (g/cc), acquired in the four wells. From left to right: Deep Adda, South East Adda, E1X and Fasan-1. The depth reference is Rotary Table (RT, i.e. the rig floor).

### Modelling of the the Post-Chalk overburden in the Tyra Field

The hydrocarbon field Tyra is situated in the Danish sector of the Central Graben in the North Sea. The hydrocarbon accumulation in this field is mainly concentrated in the Chalk Formations Ekofisk (Danian), Tor and Hod (both Upper Cretaceous).

The present-day stress-strain state of the subsurface is the result of complex geological and planetary processes, taking place over millions of years, such as plate tectonic movements, sedimentary deposition, climate changes, erosion, uplifting etc.

The most dominating geological process in the Post-Chalk period, i.e. the last 61 Million Years (MY) from Mid Paleocene (Selandian) to Present, was the filling in of the North Sea Basin with sediments originating from Fennoscandia and Scotland (Konradi, 2005, Schiøler et al., 2007, Gibbard and Lewin, 2016). The most intensive sedimentary flux into the North Sea occurred during the Late Pliocene-Pleistocene (Gibbard and Lewin, 2016). These sediments, deposited in offshore environment (Schiøler et al., 2007), form the present day overburden above the chalk reservoirs in the North Sea.

Several phases of inversion and compression in the North Sea during the Paleocene (Nielsen *et al.*, 2005; Japsen et al., 2014) and in the Early Miocene (Rasmussen, 2004) have been suggested. However, it is believed that the magnitude of these events, especially the Miocene one, did not affect significantly the depositional process in the Central Graben.

Thus, the present-day stress-strain state of the Post-Chalk overburden in the North Sea is mainly the result of the sedimentary deposition. In addition, the weight of the thick overburden affects also the stress-strain field of the underlying strata.

As mentioned above, the overburden provides the reservoir seal, and its integrity is a crucial parameter to consider when evaluating the risk of leakage of reservoir fluid (CO<sub>2</sub> or hydrocarbons).

Reservoir depletion, caused by hydrocarbon production, leads to stress and strain changes both in the reservoir and the overburden as well, among the examples are the subsidence experienced in the Tyra (Plischke, 1994; Schutjens et al., 2019) and the Ekofisk (Sulak and Danielsen, 1989) fields.

In the Tyra field, the Post-Chalk overburden is approximately 2000 m thick and is represented by three main groups: the thick Nordland and Hordaland Groups, composed mainly of smectite-rich shales (Nielsen and Rasmussen, 2015), and the much tinner Rogaland Group (Figure 2 and 3).

The well reports of the four wells (Jensen, 2004; Rong et al., 1985; Kleist et al., 1977), used in this study, hold additional useful information about the lithological composition of the overburden. The upper (most shallow 0-500 m depth) part of the Nordland Group consist of predominantly Quaternary sand sand/clay mixtures, while the lower part consist predominantly of claystone with occasional thin limestone layers (Jensen, 2004). The Hordaland Group consists predominantly of clay-rich (shale) formations, in some interval interbedded with thin limestone layers (Jensen, 2004). The Rogaland Group is situated at the top of the Chalk Group and thus represents the seal for the uppermost Chalk reservoirs (Danian age, Lower Paleocene). The thickness of Rogaland varies across the field, as illustrated by the well lithology columns in Figure 2 and 3. The upper part of the Rogaland Group (the Balder formation) is characterized by tuffaceous claystone, while the lower part (Sele, Lista, Vaale) is predominantly claystone with stringers of marl (Jensen, 2004).

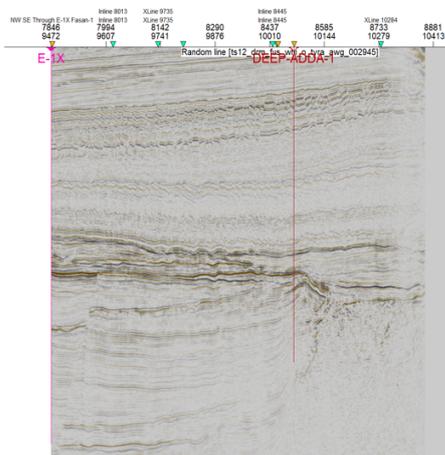
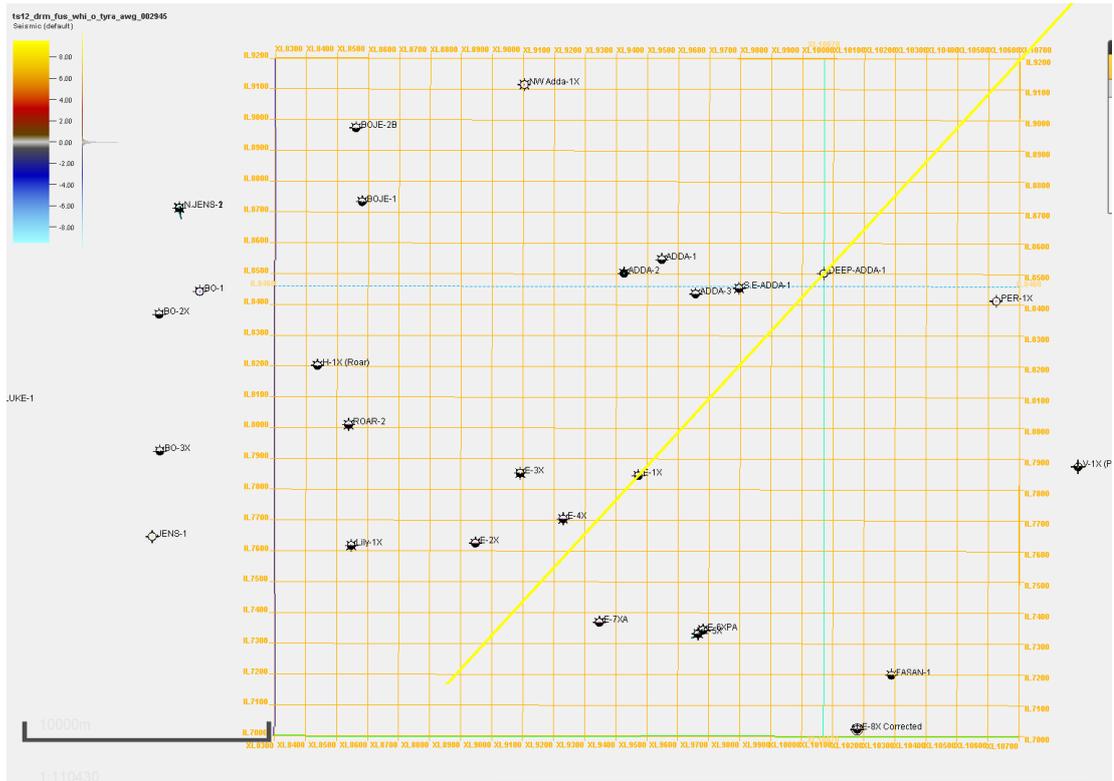
The Rogaland Group, which is overlaying the Danian Chalk sequences, spans the period from Selandian (Mid Paleocene) to Ypresian (Early Eocene), approximately from 61 to 53 million years ago (Ma); the Hordaland Group spans the period from Lutetian (Mid Eocene) to Tortonian (Mid/Late Miocene), from 53 to around 7 Ma; and the Nordland Group spans the period from Messinian (Late Miocene) to Present Day, roughly the last 7 million years. (Lithostratigraphic Chart of the Central North Sea, 2014, <https://www.npd.no>).

### Numerical representation of the Post-Chalk Deposition in the Tyra Field

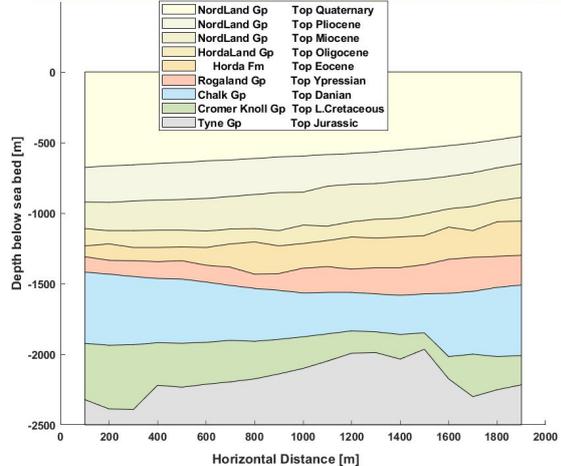
---

A 2D subsurface geometry in the Tyra Field was derived from seismic interpretation of the main reflective horizons down to (approximately) 2500 m depth below sea bed.

These main reflective horizons and the corresponding stratigraphic units are shown in Figure 4. The bottom of the domain was set to 2500 m below sea. The lateral (horizontal) extension is approximately 2000 m. It was assumed that the horizons correspond roughly to the stratigraphic groups/formations and the geological age given in the Table 1. The overburden section is composed by the strata above the Chalk Group (given in blue on Fig.4, bottom right). The steep sections of the horizons representing the tops of the Jurassic and the Lower Cretaceous sections probably indicate the presence of faults.



Reflective horizons and associated stratigraphic units and geological age



**Figure 4.** Sketch of the location of the considered seismic profile, the seismic data, and the main reflective horizons, derived from seismic interpretation and the corresponding stratigraphic units and geological age. The step sections of the horizons representing the tops of the Jurassic and the Lower Cretaceous sections probably indicate the presence of faults.

In the present model, the overburden was deposited upon a pre-existing domain (called also underburden), composed by the strata from Top Jurassic to Top Danian, Fig. 4. Thus the pre-existing domain is composed by the units (Fig. 4) Tyne (grey), Cromer Knoll (green) and Chalk (blue).

The horizons from Top Ypresian to Top Quaternary (Fig.4, Table 1) , were used to define the depositional stages in the time evolution of the Post-Chalk period.

Horizons, from Fig. 4	Stratigraphic Group / Formation	Numerical Age (Ma)
Sea Bed / Top Quaternary	Top Nordlang Gp	0.00
Top Pliocene	Mid Nordland Gp	2.60
Top Miocene	Lower Nordland Gp	5.30
Top Oligocene	Upper Hordaland Gp /Mid Lark Fm	23.00
Top Eocene	Lower Hordaland Gp/ Top Horda Fm	33.90
Top Ypresian (Mid Eocene)	Top Rogaland Gp	47.80
Top Danian (Early Eocene)	Top Chalk / Top Ekofisk Fm	61.60
Top Upper Cretaceous	Top Shetland Gp / Top Tor Fm	66.00
Top Lower Cretaceous	Top Cromer Knoll Gp	100.0
Top Jurassic	Top Tyne Gp	145.0

**Table 1.** Tentative assignment of the reflective horizons to stratigraphic groups an geological age.

The post-Chalk deposition was modelled using the approach, recently presented by Orozova-Bekkevold et al. (2021). Each stage is defined by specific duration in millions of years and a prescribed number of discrete depositional layers, composing the respective section. The duration in time and the number of discrete layers used for each depositional stage are summarized in Table 2.

Since the focus of this study was the Post-Chalk overburden, the geological evolution of the Jurassic and the Cretaceous sections pre-existing base (sedimentary deposition, tectonic events, erosion, uplifting, etc.) was not modelled in details at this stage. In addition we lacked data related to material properties and timing of past tectonic events for these sections.

Stage	Geological Period	Duration [MY]	Nr of layers
1: Pre-existing base	Upper Jurassic	0.1 - Settling under gravity	1
2: No detailed modelling	L. Cretaceous	1.0 - Settling under gravity	1
3: No detailed modelling	U. Cretaceous – Danian	1.0 - Settling under gravity	1
4:	Ypresian (L. Eocene)	8.0	4
5:	Mid to Late Eocene	19.3	5
6:	Oligocene	10.7	5
7:	Miocene	17.7	5
8:	Pliocene	2.7	2
9:	Quaternary	2.6	2
10:	Post-depo settling	Projected 0.5 MY after present	Not used

**Table 2.** Main deposition stages with duration (in Million Years, MY) and present day thickness.

The layers were deposited one-by-one on the top of the underlying ones until the height of the respective horizon (Fig. 4) was reached. At the end, after the deposition of the last layer, a ‘settling’ period of the duration of 0.5 MY was introduced to simulate conditions of relaxation with no deposition.

The Jurassic and Cretaceous sections were modelled as “deposited” in one chunk and were “allowed” to settle under gravity, before the onset of the Post-Chalk deposition. The material, composing the Upper Jurassic (stage 1 in Tab. 2) is assumed to be consolidated shale, while the Low Cretaceous to Top Danian (stage 2) is assumed to be composed by consolidated sandstones.

The stratum between Top Lower Cretaceous and Top Danian can be seen as a proxy for a reservoir.

It is assumed that the deposited material composing the Post-Chalk overburden is unconsolidated clay and all materials are considered to be

- isotropic and homogeneous;
- fully saturated with water;
- and at any given time the sediments are at their maximum burial depth.

Clay diagenesis and other chemical and temperature effects are not modelled at the current stage.

### Finite elements modelling of the Post-Chalk deposition in the Tyra Field

The evolution in geological time is modelled in the terms of a finite element method, using the software Elfen (ELFEN, Rockfield Software Ltd.). The framework and the theory behind the software are given in details in Crook et al. (2003), Peric and Crook (2004), (Crook et al. (2006a, 2006b), Thornton and Crook (2014).

The non-consolidated clay material composing the overburden is represented as a poro-elasto-plastic material, fully saturated with water. The mechanical field (the solid part) is solved explicitly, while the seepage field is solved by implicit time integration schemes and the two fields are coupled at given time intervals (Thornton and Crook, 2014).

The mechanical properties of the water-saturated medium, were expressed as:

$$\text{div}(\boldsymbol{\sigma}') + [(1 - \varphi)\rho_s + \varphi\rho_f](\mathbf{g} - \mathbf{a}_s) = 0 \quad (1)$$

The fluid (water only) flow over geological time is represented with a transient equilibrium equation:

$$\text{div}\left[\frac{k(\varphi)}{\mu} \nabla P_f - \rho_f(\mathbf{g} - \mathbf{a}_s)\right] = [\varphi/K_f + (\alpha - \varphi)/K_s] \frac{\partial P_f}{\partial t} + \alpha \frac{\partial \epsilon_v}{\partial t} \quad (2)$$

where:

$\boldsymbol{\sigma}' = \boldsymbol{\sigma} - \alpha \mathbf{P}$  is the effective stress;  $\rho_s$  and  $\rho_f$  are the solid and the fluid density, respectively;  $\mathbf{g}$  is the Earth’s gravitational acceleration,  $\mathbf{a}_s$  is the acceleration of the solid phase,  $P_f$  is the fluid pressure,  $\varphi$  is the porosity,  $k(\varphi)$  is the porosity-dependent permeability,  $K_f$  is the fluid bulk modulus,  $K_s$  is the frame bulk modulus,  $\alpha$  is the Biot’s coefficient and  $\epsilon_v$  is the volumetric strain.

The bulk modulus of the non-consolidated clay is expressed as a function of the mean effective stress  $P'$  (Thornton and Crook, 2014):

$$K = K_0 + \frac{(1-A)P_{co}}{\kappa} \exp \left[ \frac{\varphi_0 - \varphi}{\lambda(1-\varphi_0)(1-\varphi)} \right] + \frac{A*\sigma'}{\kappa(1-\varphi)} \quad (3)$$

where

$\varphi_0$  is the initial porosity,  $K_0$  is the initial bulk modulus and  $P_{co}$  is the initial pre-consolidation pressure,  $A$  is a weighting factor,  $\kappa$  and  $\lambda$  are material constants.

The vertical stress,  $S_v$ , resulting by the weight of the overlaying sediments is assumed to be the maximum stress. The horizontal stresses is assumed to be isotropic, and in the absence of tectonic forces, it is derived from the vertical stress:

$$S_h = K_{eff} * S_v \quad (4)$$

where  $K_{eff}$  is the so-called effective stress ratio (Matthews and Kelly, 1967). Finally, the mean effective stress is obtained as

$$\sigma' = \left[ \frac{S_v + 2*S_h}{3} \right] - \alpha P_p = \sigma - \alpha P_p \quad (5)$$

Where is  $P_p$  the pore pressure and  $\alpha$  is the Biot's coefficient. Density and porosity are derived by well data (sonic and density logs), following the procedure presented in Orozova-Bekkevold et al. (2021).

The evolution of these properties in time is driven by the rate of deposition and subsequent burial of fully saturated material, as described in Orozova-Bekkevold et al. (2021), Crook et al. (2003), Peric and Crook (2004), Crook et al. (2006a, 2006b), Thornton and Crook (2014).

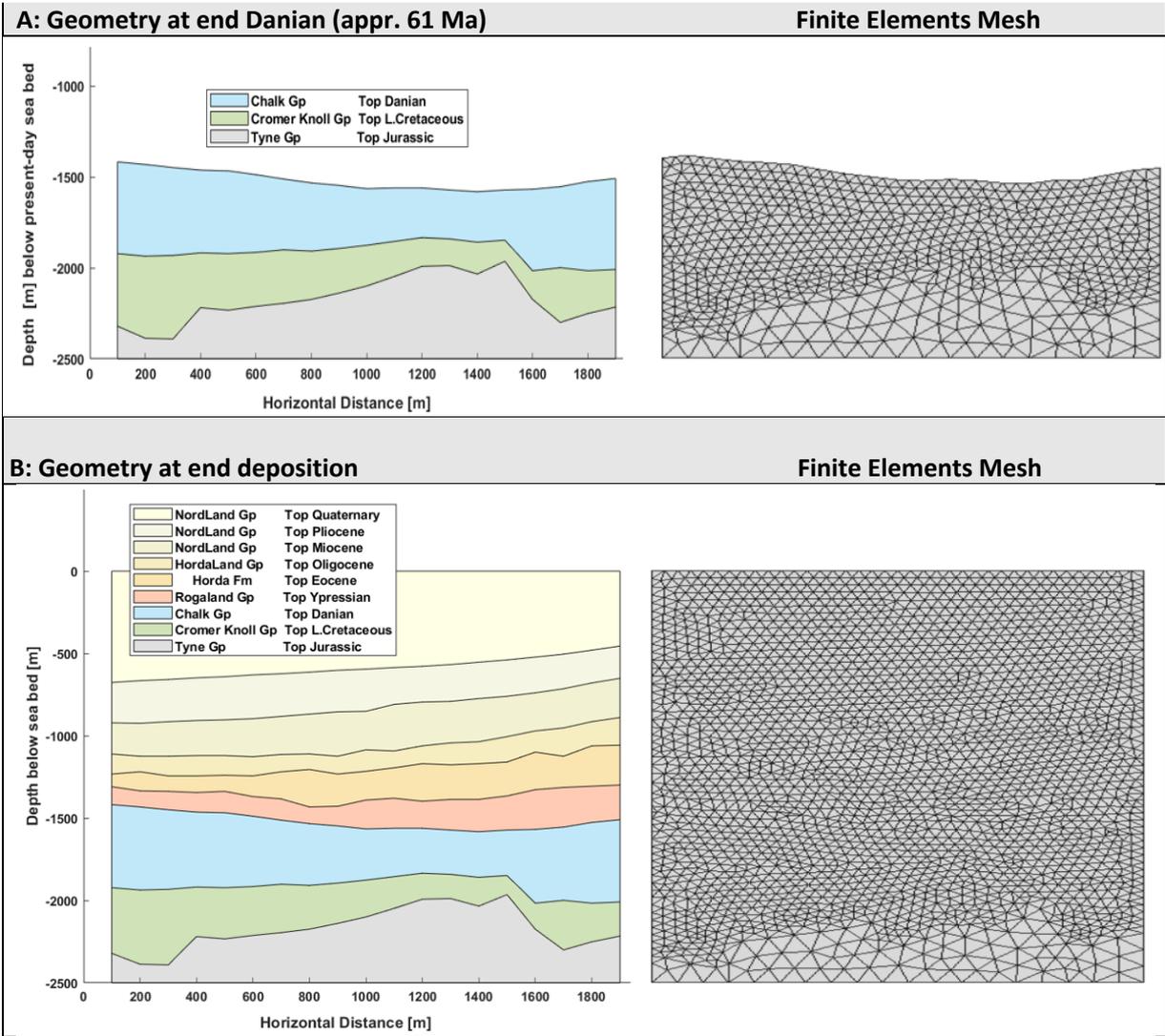
**Initial and boundary conditions:** The main force acting upon the domain is the gravity. The gravitational load originates at the top of the sediments and acts downwards. This setup is considered representative for the Cenozoic Period in the North Sea basin, where no major tectonic events (uplift, erosion, collision, subduction etc.) occurred, and thus the maximum stress is caused only by the weight of the deposited material and acts in the vertical direction. Uniaxial compaction (i.e. plain strain conditions) is assumed.

The domain below the overburden (Jurassic to Danian, stage 1-3 in Tab. 2) is not allowed to deform at the bottom, along the top and across the sides.

As mentioned above, the deposited material is modelled as a fully water saturated porous medium. The formation water can flow both vertically and horizontally within the domain, but there is no fluid flow from outside sources. The water does not flow through the bottom and no capillary and temperature effect are taken into account at this stage.

**Meshing:** The finite element mesh is generated by an advancing front algorithm, adding new elements as the geometry expands, following the deposition of new layers of material (Crook et al. (2003), Peric and Crook (2004), Crook et al. (2006a, 2006b)). The elements are triangular, with initial size of 100m. The size of the elements is rescaled during the simulation, depending on the estimated plastic strain at a given step: plastic strain exceeding 2 results in diminished element size.

At the beginning of the simulation, before start of the deposition, the finite elements mesh consist of around 120 triangular elements, all with size 100. At the end of the simulation, the mesh consists of around 6750 elements with size ranging from 50 to 100. The model geometry and the finite element mesh at the beginning and the end of the deposition is summarized in Figure 5.

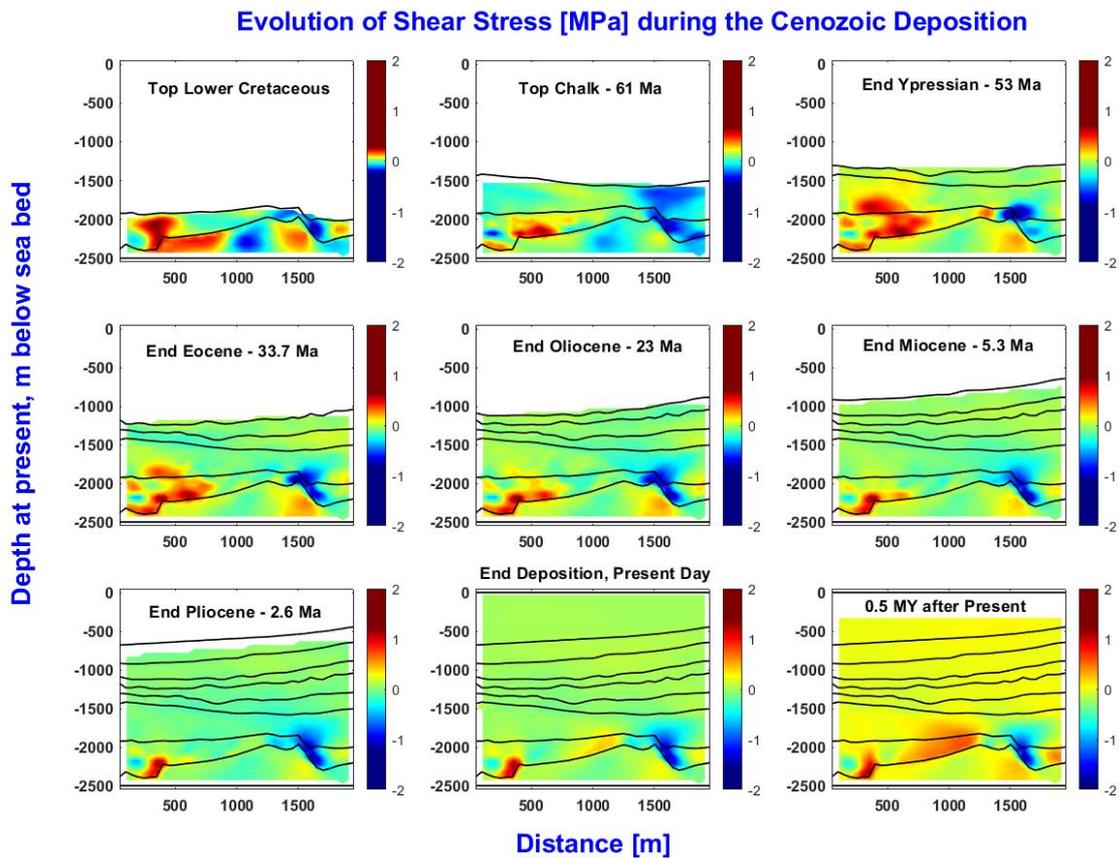


**Figure 5.** Model Geometry and the finite elements mesh in time. The depth is given meters [m] below sea bed at present. From top to bottom: a) Pre-existing domain (underburden) at time appr. 61 Ma (end Danian) before the onset of the overburden deposition; b) The final stage at present time after appr. 61 MY of deposition.

## Results: Evolution of the stress and strain state in the Tyra Field during the Cenozoic Deposition

The estimated changes in time in the subsurface stress-strain fields as result of the overburden deposition in the last 61 MY are reported in this section.

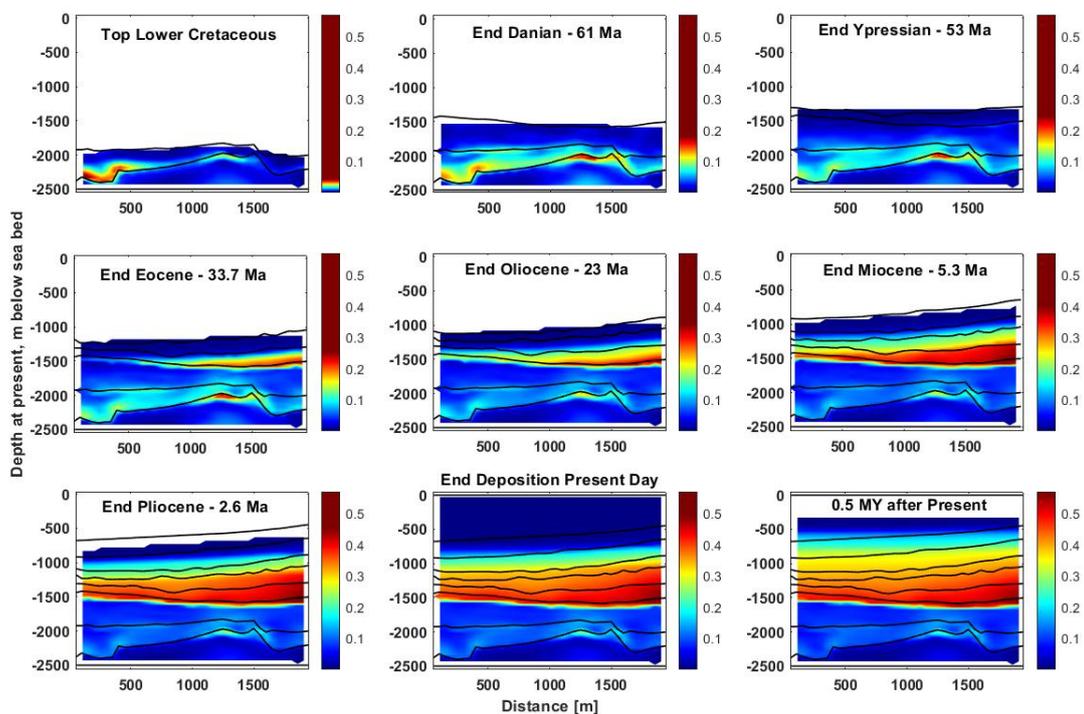
If the shear (differential) stress exceeds the strength of the material, it might result in fracturing of the formation. Figure 6 shows the evolution of the shear stress during the deposition of the overburden. The plot in the centre of the top row (labelled “Top Chalk – 61 Ma”) corresponds to the stage just before the onset of the Cenozoic deposition, approximately 61 Million years Ago (Ma). As the figure shows, the largest shear stress was estimated along the very steep shoulders of the deepest horizon, which corresponds to the boundary between the Upper Jurassic and the Lower Cretaceous. Relatively high stress magnitudes were also estimated in the lowest strata, roughly corresponding to the Chalk reservoirs, while the shear stress estimated in the overburden was low. It is interesting to observe that in the 0.5 MY after the end of deposition, the projection suggests that the subsurface continues to react and a re-distribution of the shear stress might occur, especially in the deepest layers (right bottom corner of the plot).



**Figure 6.** Evolution of the Shear Stress [MPa] during the deposition of the overburden. Depth is in meters, below sea bed. The horizons are at their present day-depth. Shear Stress is given in MPa

The zones with large shear stress might be considered being at higher risk of fracturing. The largest shear stress magnitude are estimated at present time where the overburden build-up is completed. If no further deposition is going to occur in the next 0.5 MY, the medium relaxes and the magnitude of the shear stress could decrease.

Figure 7 shows the evolution in the effective strain during the burial process. This could also be interpreted as a kind of deformation accumulated during the Post-Chalk sedimentation. The maximum values were estimated in the overburden, along the Danian-Ypressian boundary, in the strata which forms the top seal of the Upper Chalk reservoirs. The strain continues to increase also in the projected post-deposition period (the right bottom plot in Fig. 7).



**Figure 7.** Evolution of the effective strain as result of the deposition of the Post-Chalk sediments.

### Seismic Study of the fault Detection Limit

Seismic reflection data is used extensively in the oil and gas sector for imaging and characterizing the subsurface. Reflected energy and arrival time of reflected waves provides information on elastic properties in specified locations and allows us to map geological formations on large and smaller scales, only limited by the resolution of the data.

Seismic resolution implies the capability of distinguishing between different geological micro-structures. Some geological features such as small faults and fractures are below the seismic resolution threshold. Recognition of small-scale structures helps us to study thin layers and minor fractures and faults in

reservoirs for future exploration or risk analysis (Ashraf, 2020). When conducting risk assessment in the area, the fault detection threshold in the model, or the smallest fault size that can be detected by seismic data, as well as prior knowledge of the mechanical properties of the rock, such as the stress and strain condition of the overburden, must be taken into account.

Using probabilistic models for evaluating the fault density in a certain area, we can assess the likelihood of CO<sub>2</sub> migration in the reservoir. One of the first investigations of possible leakage risks and safety of CO<sub>2</sub> underground injection was conducted by Holloway (1997), and probabilistic methods for risk assessment of such problems has become more and more popular during years (Kopp et al., 2010; Smith et al., 2011; Zunino et al., 2015 ).

In this study we use a simplified probabilistic approach to locate areas of increased risk of fracturing in the cap rock. We use a Monte Carlo method to generate small subseismic faults escaping the resolution limit of the data, but at the same time having a fault density proportional to a prior probability density derived from the differential stresses presented in the previous section.

Markov Chain Monte Carlo algorithms are commonly used for solving nonlinear inverse problems and sampling posterior probability density where the aim is integrating independent sources of information such as geological information and prior knowledge, geophysical data and data from nearby well logs. In stochastic approaches such as the Monte Carlo method, the size of the model space and high calculation cost associated with forward modelling could be challenging. Khoshkholgh et al., 2021 proposed a new methodology known as Informed Proposal Monte Carlo (IPMC), for MCMC sampling method that uses external information obtained from simplified physics to form a global proposal distribution that guides the sampling procedure. An example of solving a probabilistic Full-waveform inversion by IPMC for a near surface velocity model in large scale can be found in Khoshkholgh et al., 2021.

In this study we use Informed Proposal Monte Carlo in order to integrate (1) prior knowledge about strain and stress fields and fault and fracture probability models in the reservoir, and (2) information from observed seismic data, to provide uncertainty analysis and to assess the probability of CO<sub>2</sub> migration. The aim is to contribute to probabilistic models that could appraise possible CO<sub>2</sub> leakage through faults in the overburden.

## Seismic Inversion

---

In our inversion approach, our aim is to produce subsurface models presenting relevant physical characteristics of the caprock in the chosen area. If  $\mathbf{m}$  denotes the model parameters (in our case acoustic impedance),  $\mathbf{d}$  the data and  $g$  stands for the forward operator (represented by the wave propagation algorithm), the forward problem can be expressed as:

$$\mathbf{d} = g(\mathbf{m}).$$

The inverse problem, in its widest sense, can be described as a problem of integrating measurable data information, prior information and theoretical connections between models and data (Tarantola et al. 1982). As mentioned by Tarantola and Valette (1982) the solution of the inverse problem is then defined as a posterior probability density which is obtained as a product of the prior probability density  $\rho$ , providing the probability of models based on non-seismic information, and the likelihood function  $L$ , measuring the degree of fit between the observed data, and synthetic data calculated from the elastic reservoir model:

$$\sigma(\mathbf{m}) = k \rho(\mathbf{m}) L(\mathbf{m}).$$

$k$  is the normalization constant and  $\sigma(\mathbf{m})$  denotes the posterior probability density that is considered as the solution of inverse problem. In our case,  $\rho(\mathbf{m})$  is derived from a simple hypothesis about proportionality of fracture risk and shear (differential) stress, and  $L(\mathbf{m})$  is given by

$$L(\mathbf{m}) = C \exp\left(-\frac{1}{2}(\mathbf{d}_{obs} - g(\mathbf{m}))^T \mathbf{C}_D^{-1}(\mathbf{d}_{obs} - g(\mathbf{m}))\right)$$

where  $\mathbf{d}_{obs}$  are the (noisy) observed data, and  $\mathbf{C}_D$  is the covariance matrix of the noise. In this study we assume that  $\mathbf{C}_D = \sigma_D^2 \mathbf{I}$ , where  $\sigma_D$  is the standard deviation of the seismic noise.

When the posterior probability density cannot be calculated analytically, Monte Carlo sampling is required (Mosegaard, 2006). The posterior distribution can be described by an ensemble of realizations implemented by Markov-Chain Monte Carlo (MCMC) algorithms, and their variability expresses the uncertainty of subsurface structures (Hastings 1970; Metropolis et al. 1963; Mosegaard and Tarantola 1995; Tarantola, 2005). This requires the ability of calculating likelihood  $L$  at each point, as well as an algorithm that can sample the prior  $\rho$ . When the prior model is complex, as in this study, sampling through numerical operations is often the only way to incorporate this information (Mosegaard 1998; Zunino et al. 2015). In this study our geological prior information comes from two main sources: (1) the shear stress analysis presented in the previous section, providing a fault risk map, and (2) the simplifying assumption that the reservoir zone consists of a stack of (non-horizontal) homogeneous layers whose properties are calibrated to well information in the area (the wells E-1X and Deep Adda).

There are several studies looking at the properties of fault and fractures zones in relation to hydrocarbon investigation or CO2 storage (Aydin, 2000 ; Rotevatn et al., 2011; Shipton et al., 2004; Dockrill et al., 2010). A fault displacement causes two damage zones on both sides, which could be categorized into three different groups: along fault, around tip and cross fault (Choi et al., 2015). The thickness of the damage zone is influenced by the size of the fault displacement. In our prior model generator, fault models in a specific angle range have been proposed in overburden and the fault density is proportional to the shear stress.

Khoshkholgh et al., 2021 proposed a warping strategy for generating prior models that works by warping and deforming the subsurface image and perturbing the velocity value at each layer. This is done by finding a random window inside the model and allowing deformations such that they are maximum in the center and gradually fade away and become zero at the boundaries. The velocity perturbation within one layer is adjusted by an estimated modelization error (Khoshkholgh et al., 2021). In this study, the same technique is applied in such a way that fault planes are randomly (according to the prior) introduced in overburden during perturbation and the displacement vector directions on either side of the fault line are opposite.

We use the seismic convolutional model as the forward model, which produces the seismic data by convolving the reflectivity series with a wavelet. The reflectivity is computed from the seismic impedance, which is a product of density and velocity. We assume here that the density is roughly proportional to the velocity (Liner, 1999). In order to simulate the residual, horizontal smearing left in the data after (an unavoidable imperfect) migration, we use a 2D wavelet (see Figure 8):

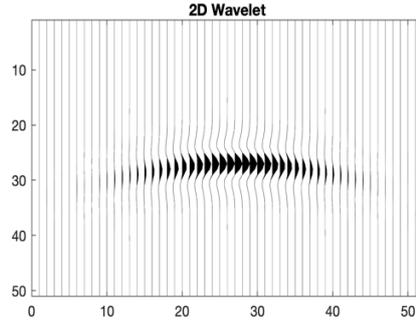


Figure 8. 2D wavelet used in this study

where the variation in time is a zero phase Ricker wavelet. The following is the formula for convolutional modeling of the seismic profile:

$$R(x, t) * W(x, t) + n(x, t) = D(x, t)$$

Where  $R(x, t)$  is the 2D reflectivity,  $W(t)$  is the 2D seismic wavelet and  $n(x, t)$  is the added noise (Partyka, 1999). For the sake of simplicity, we will assume in this study that the wavelet is constant everywhere in the depth domain, and hence the modeling can be carried out with  $t$  replaced by depth  $z$ .

To carry out the inversion we use a Markov Chain Monte Carlo algorithm which proceeds as follows:

- From current starting model  $\mathbf{m}$ , propose  $\mathbf{m}'$  using the proposal distribution  $q(\mathbf{m}' | \mathbf{m})$
- Accept  $\mathbf{m}'$  with probability:

$$p_{acc}^{(1)} = \min\left(\frac{\rho(\mathbf{m}')q(\mathbf{m} | \mathbf{m}')}{\rho(\mathbf{m})q(\mathbf{m}' | \mathbf{m})}, 1\right)$$

where  $\rho(\mathbf{m})$  is the prior probability density.

- If  $\mathbf{m}'$  is accepted, perform another test, where it is accepted with probability

$$p_{acc}^{(2)} = \min\left(\frac{L(\mathbf{m}')}{L(\mathbf{m})}, 1\right)$$

where  $L(\mathbf{m})$  is the likelihood function.

- If  $\mathbf{m}'$  is rejected by both of the above tests, repeat  $\mathbf{m}$ .

In order to construct an MCMC algorithm with an informed proposal, we first assume that data uncertainties are very small in comparison to the modelization error (Khoshkholgh et al. 2021). This results in the following formula:

$$\sigma_m(\mathbf{m}) \propto \sigma(\mathbf{d}_{obs}, \mathbf{m}) \approx \theta(\mathbf{d}_{obs}, \mathbf{m})$$

where  $\sigma(\mathbf{d}, \mathbf{m})$  is the posterior distribution in the joint data-model parameter space  $D \times M$ , and  $\theta(\mathbf{d}, \mathbf{m})$  is the distribution in  $D \times M$  that describes the correlation between model parameters and data (allowing an uncertain forward relation). Since the true modelization error  $\delta\mathbf{m}_{true}$  is unknown, we build an approximate modelization error distribution  $\theta(\mathbf{d}_{obs}, \mathbf{m})$  by considering a simplified inverse problem that is similar to the original problem: First we solve the problem using a simplified pseudo-inverse  $h$  which gives us the rough estimate  $\tilde{\mathbf{m}} = h(\mathbf{d}_{obs})$ . We now compute synthetic data from the estimate, using the correct forward  $g$ :  $\mathbf{d}_{synt} = g(\tilde{\mathbf{m}})$ , and invert this result again using  $h$ :  $\tilde{\mathbf{m}}_2 = h(\mathbf{d}_{synt})$ . If  $\mathbf{m}_{true}$  and  $\tilde{\mathbf{m}}$  are close in the model space, this back-and-forth process approximately reveals the modelization error, which can now be estimated as  $\delta\mathbf{m}_{approx} = \tilde{\mathbf{m}}_2 - \tilde{\mathbf{m}}$  (Khoshkholgh et al. 2021). A simple modelization error distribution  $\theta(\mathbf{d}_{obs}, \mathbf{m})$  can now be constructed as an isotropic Gaussian with mean  $\tilde{\mathbf{m}}$  and with the components of  $\delta\mathbf{m}_{approx}$  as standard deviations. This approximate modelization error distribution is used as a global proposal  $q(\mathbf{m}'|\mathbf{m})$ . In this way, our external knowledge and information about the target distribution can be injected into the problem through the proposal. Using the global proposal will not bias the problem or distort the result, but just accelerate the sampling process and leads to a more efficient algorithm (Khoshkholgh et al., 2021).

## Results: Seismic Inversion

---

A 2D seismic profile is chosen from the Tyra field in the Danish part of North see. There are two well logs located in the chosen area. By simple interpretation and from the well information, a velocity function was generated. This velocity model was considered as the center of the informed proposal distribution, modelization error distribution was calculated from it. The following steps were taken to create the informed proposal distribution:

- a) Synthetics were computed from the reflectivity obtained from the center velocity model.
- b) Synthetics were inverted with a simple linear inversion (deconvolution) and a second reflectivity model is obtained.
- c) Modelization error for the layer boundaries were calculated by finding the difference between the first and the second reflectivity model. The modelization error was then turned into a modelization error envelope.
- d) The center velocity model and the error envelope were used for the global proposal strategy in the IPMC method.

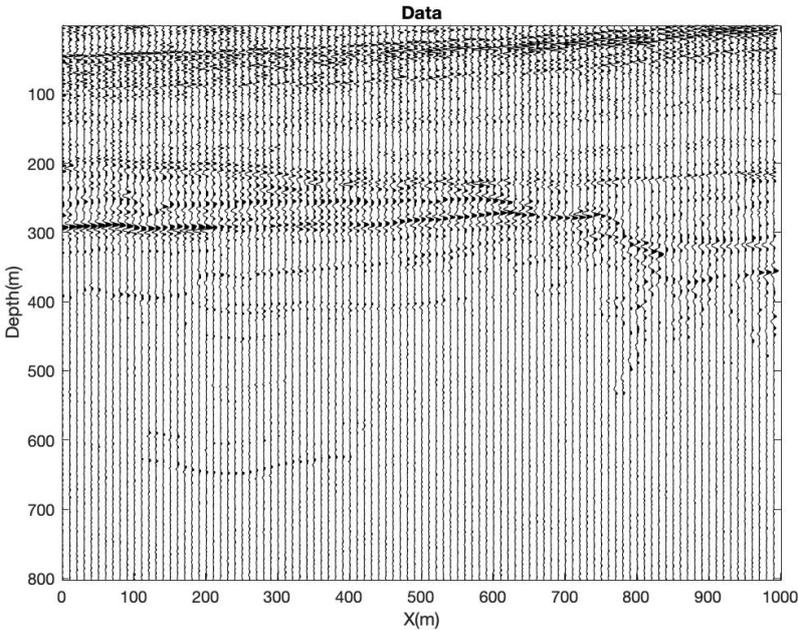
As previously explained, our strategy for generating prior models is inspired from Khoshkholgh et al. 2021, except for the fact that in this study fault lines are introduced into the model during perturbation. The probability of having fault structures in each point is obtained from the differential stress field and used in the sampling procedure. The prior probability is non-zero for piece-wise constant velocity models (Khoshkholgh et al., 2021). Overall, perturbing the prior model is a combination of warping or deforming the layer boundary shape, perturbing and changing the velocity value in each layer and introducing fault lines with regard to the fault and fracture probability.

Figure 9 shows a 2D seismic profile from the Tyra field that was used in this study. Figure 10 shows the center velocity model obtained by simple interpretation. The reflectivity modelization error estimation is shown in figure 11 and the modelization envelope is shown in figure 12. In figure 13 the prior fracture probability model, obtained from the stress field, is shown.

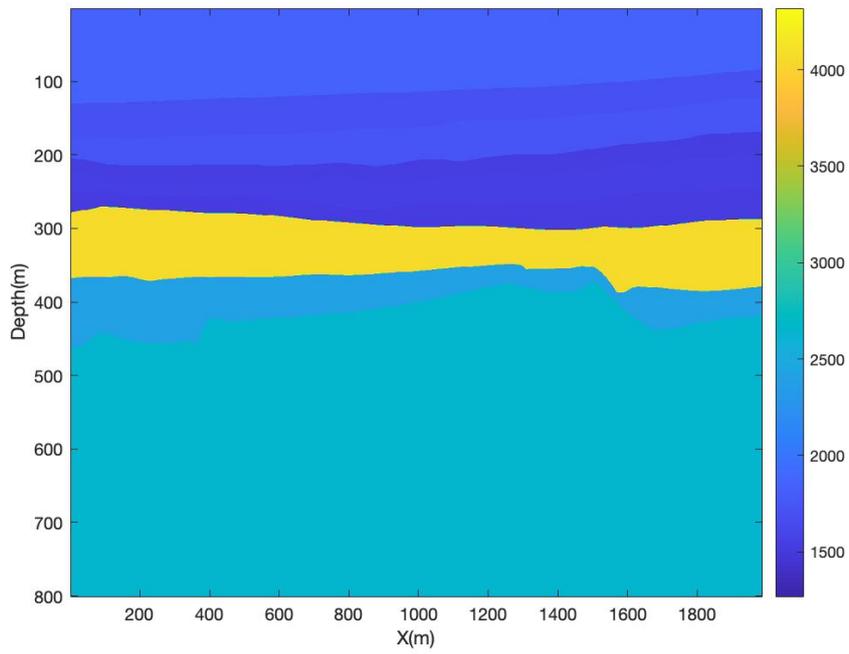
To start sampling the posterior probability density the center velocity model is chosen as the starting model. The maximum velocity perturbation is chosen as 1 percent of the maximum velocity and the maximum displacement in warping is chosen as 5 pixels. Warping happens in a square window centered at a randomly selected point. The size of the window is 200 by 200. The maximum and minimum fault angles are 12 degrees and 36 degrees respectively.

Figure 14 shows two different realizations from posterior. Using a global, informed proposal containing external information about the posterior sped up the sampling process, and made the sampling possible for the full model with more than  $10^6$  parameters.

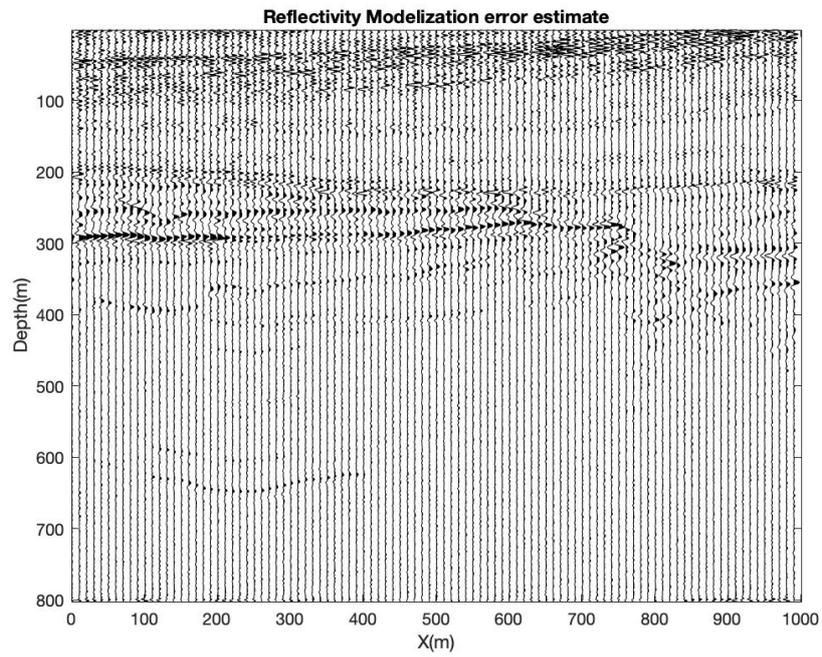
The posterior fault/fracture density is shown in figure 15 and suggests that the bottom of our 2D selected area, which is located under the reservoir, has a very low probability of having a fault or a fracture zone.



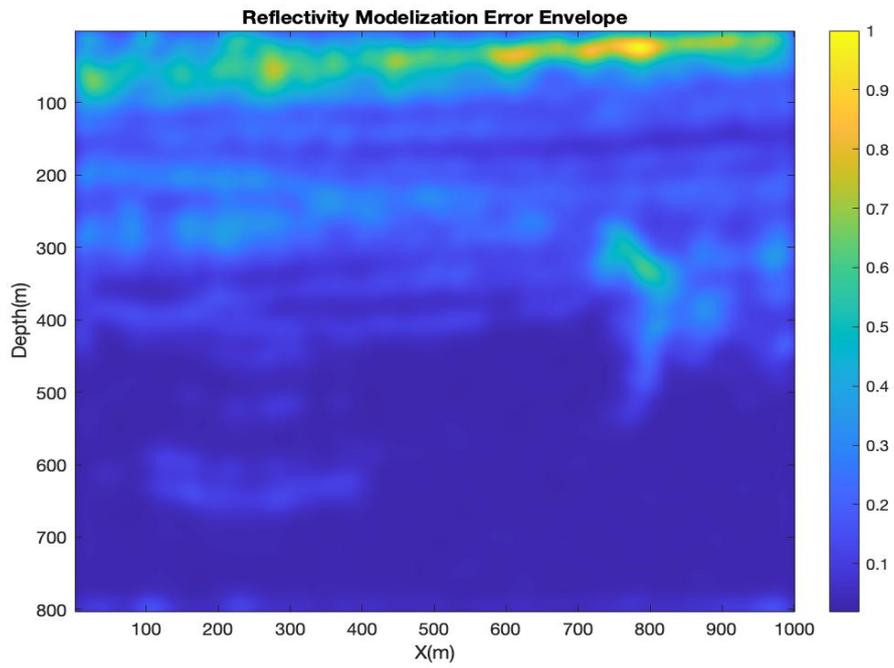
**Figure 9.** 2D observed seismic profile from the Tyra field.



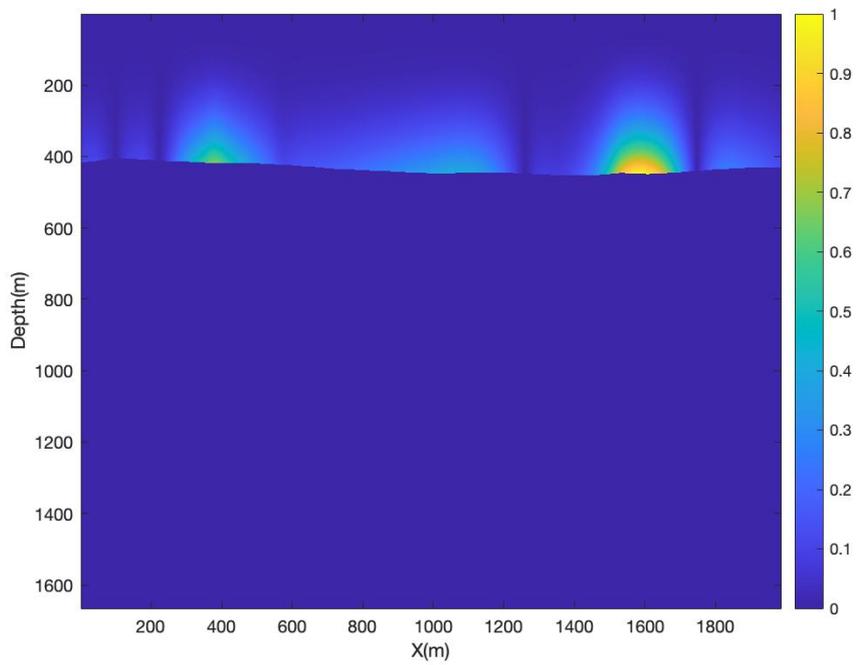
**Figure 10.** The center velocity model obtained by simple interpretation.



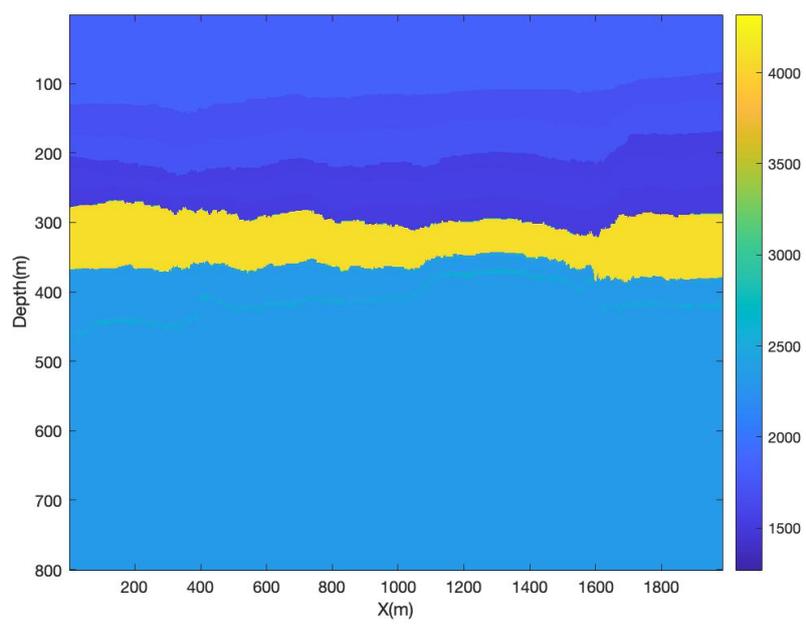
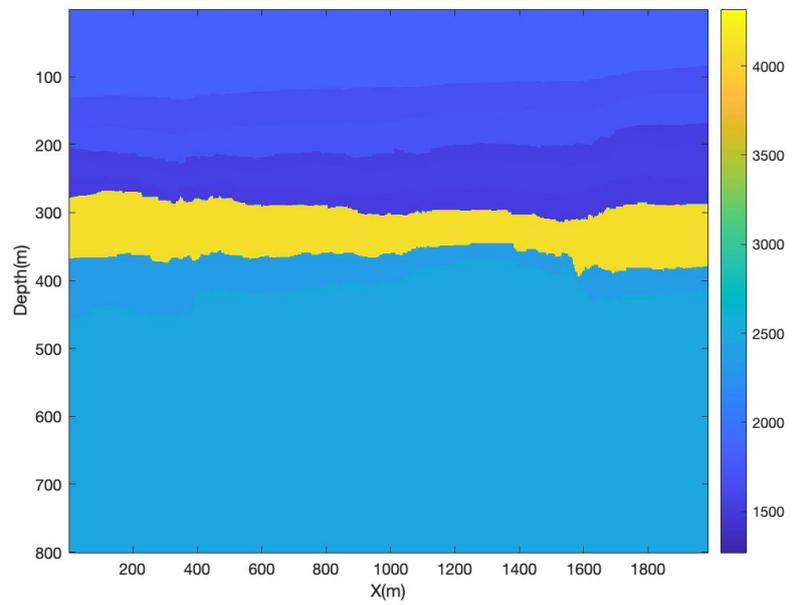
**Figure 11.** The reflectivity modelization error estimation.



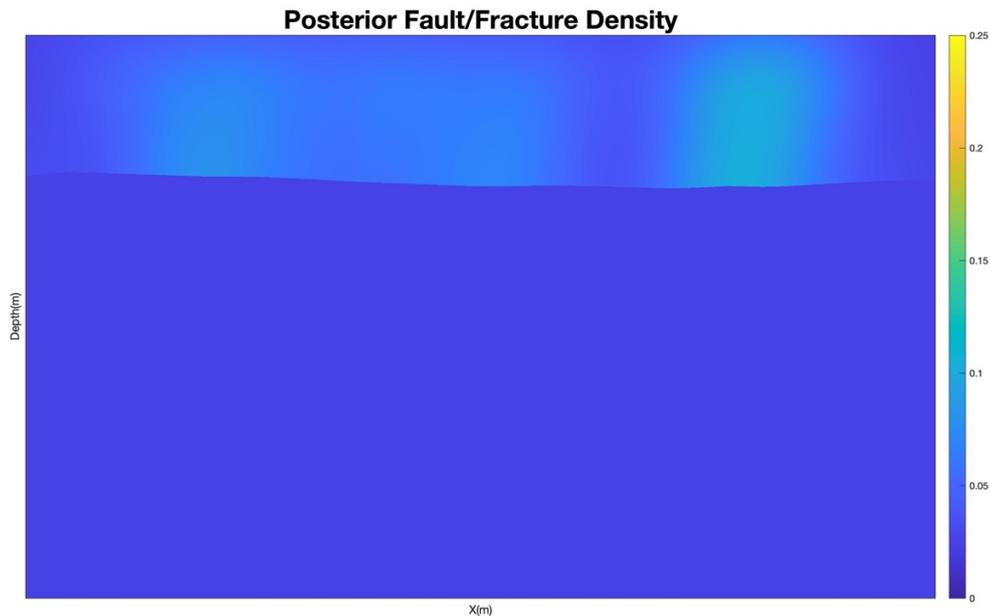
**Figure 12.** Reflectivity modelization error envelope.



**Figure 13.** Prior fracture probability model obtained from the stress field.



**Figure 14.** Two realizations from the posterior probability density.



**Figure 15.** Posterior Fault and fracture density.

## Conclusions and Discussions

---

A forward conceptual model was used to simulate the evolution in time of the stress and strain fields in the Tyra field as result of the deposition of the approximately 2000m thick clay-rich Post-Chalk overburden in the last 61 Million years.

The main purpose was to study possible risks, in terms of probability of fracturing or fault re-activation, related to seal integrity.

Since the resolution of even the best quality seismic data does not allow to detect small fractures and faults, the forward conceptual model allowed us to identify areas of stress-strain concentration, which possibly can be at increased risk of fracturing and probable fault re-activation, especially in the cap rock (immediate overburden).

The simulation showed that the deposition of thick clay-rich Cenozoic Overburden in the last 61 MY on the top of a pre-existing structure could have produced significant shear stress in the strata below the Overburden, i.e. the Upper Jurassic, the Lower Cretaceous and the Chalk Group. The largest values of the shear stress were estimated along the very steep sections at Top Jurassic, which might corresponds to fault zones. The areas of high shear stress could be at higher risk for fracturing.

The largest effective strain (deformation) was estimated in the section above Top Chalk, which forms the seal of the Chalk reservoir.

The very intensive sedimentation in last 5.3 MY seems to lead to significant increase and re-distribution in both the shear stress of the deepest strata and the strain accumulated in the overburden.

The projection in time after end deposition showed that the strain-stress field continues to evolve even if no new sedimentation is occurring.

The stress-strain field, estimated with the present model is believed to correspond to natural state of the subsurface, i.e. the state before any kind of human intervention.

It should be kept in mind that the forward model of the evolution of the subsurface presented here is conceptual and was used mainly to illustrate the importance of the Cenozoic deposition for the present-day subsurface stress-strain field. The model can be improved by including further details.

A simplified probabilistic analysis of a seismic 2D profile across the Tyra reservoir, intersecting the E1-X and Deep Adda well sites was carried out. The focus was on the detection limit of the data, and the result was a map of faulting and fracturing probabilities across the area. The result is a combination of seismic information with information about stress fields, derived from the subsurface evolution model.

Based on this study, we currently cannot make very detailed conclusions about the actual conditions at specific locations in the Tyra field. Both the subsurface evolution model and the seismic analysis were oversimplified and important processes, among others, anisotropy in stress and rock properties, the pre-Cenozoic geological and tectonic history, the effect of temperature and capillary phenomena were not included in the current study. A notable weakness in the seismic analysis is the lack of accurate data modelling (the wavelet estimation and the wave simulation), and we have also not included an obvious source of information, namely anisotropy of rock properties, which may be derived from full waveform data or AVO/AVA data. An investigation of anisotropy may contain valuable information about rock fracturing, and should be taken up in future studies.

The outcome of the study is, however, encouraging. We have proposed a new way of analysing caprock integrity of reservoirs for CO<sub>2</sub> storages. Our subsurface evolution model potentially allows us to predict current and future changes in a reservoir, and our probabilistic approach to seismic data analysis allows our numerical model of the stress field to be integrated with seismic information, producing fault/fracture probability maps that are ready to be included in a quantitative risk analysis.

## Acknowledgements

---

The authors would like to thank the DHRTC for providing the data the financial support for this study.

Ivanka Orozova-Bekkevold acknowledges the support of this work by Halliburton Software and Services, a Halliburton Company by allowing the authorized use of the "Drillworks/Predict" software used to produce Figure 1 and 2.

## References

---

- Ashraf, U., Zhang, H., Anees, A., Nasir Mangi, H., Ali, M., Ullah, Z., and Zhang, X. 2020. "Application of Unconventional Seismic Attributes and Unsupervised Machine Learning for the Identification of Fault and Fracture Network" *Applied Sciences* 10, no. 11: 3864. <https://doi.org/10.3390/app10113864>
- Aydin, A., 2000. Fractures, faults and hydrocarbon entrapment, migration and flow. *Mar. Pet. Geol.* 17, 797814.
- Bachu, S. 2008: CO2 storage in geological media: Role, means, status and barriers to deployment. *Progress in Energy and Combustion Science* 34, 254–273. <https://doi.org/10.1016/j.pecs.2007.10.001>.
- Bergmann, P., Diersch, M., Götz, J., Ivandic, M., Ivanova, A., Juhlin, C., Kummerow, J., Liebscher, A., Lüth, S. & Meekes, S. 2016: Review on geophysical monitoring of CO2 injection at Ketzin, Germany. *Journal of Petroleum Science and Engineering* 139, 112–136. <https://doi.org/10.1016/j.petrol.2015.12.007>.
- Choi, J., Edwards, P., Ko, K., Kim, Y., 2015. Definition and classification of fault damage zones: A review and a new methodological approach: *Earth-Science Reviews* 152 (2016) 7087.
- Crook, A.J.L., S.M. Willson, J.G. Yu and D.R.J. Owen (2006a) Predictive modelling of structure evolution in sandbox experiments. *Journal of Structural Geology*, v. 28, 729–744.
- Crook, A.J.L., Owen, D.R.J., Willson, S., and J. Yu (2006b) Benchmarks for the evolution of shear localization with large relative sliding in frictional materials. *Comput. Methods Appl. Mech. Eng.*, v. 195, 4991-5010.
- Crook, T., Willson, S., Yu, J.G. and R. Owen (2003) Computational modeling of the localized deformation associated with borehole breakout in quasi-brittle materials. *J. Petrol. Sci. Eng.*, v. 38, 177-186.
- Daley, T., Myer, L., Peterson, J., Majer, E. & Hoversten, G. 2008: Time-lapse crosswell seismic and VSP monitoring of injected CO2 in a brine aquifer. *Environmental Geology* 54, 1657–1665. <https://doi.org/10.1007/s00254-007-0943-z>.
- Dockrill, B., Shipton, Z.K., 2010. Structural controls on leak-age from a natural CO2 geologic storage site: Central Utah, U.S.A. *J. Struct. Geol.* 32, 17681782.
- Doughty, C., Freifeld, B. & Trautz, R. 2008: Site characterization for CO2 geologic storage and vice versa: the Frio brine pilot, Texas, USA as a case study. *Environmental Geology* 54, 1635–1656. <https://doi.org/10.1007/s00254-007-0942-0>.
- Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T.A. & Høier, L. 2011: Lessons learned from 14 years of CCS operations: Sleipner, In Salah and Snøhvit. *Energy Procedia* 4, 5541–5548. <https://doi.org/10.1016/j.egypro.2011.02.541>.
- Elfen, Finite Element and Discrete Element software tool written at, and by Rockfield Software Ltd., Rockfield Software, Swansea, United Kingdom.

Gardner, G.H., Gardner L.W. and A.R. Gregory (1974) Formation velocity and density – the diagnostic basics for stratigraphic traps. *Geophysics*, v. 39.6, 770-780.

Gibbard, P.L. and J. Lewin (2016) Filling the North Sea Basin: Cenozoic sediment sources and river styles. *Geologica Belgica*, v. 19, n. 3-4, 201-217. <http://dx.doi.org/10.20341/gb.2015.017>

Gozalpour, F., Ren, S. & Tohidi, B. 2005: CO2 EOR and storage in oil reservoir. *Oil & gas science and technology* 60, 537–546. <https://doi.org/10.2516/ogst:2005036>.

Hastings, W.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**(1), 97 (1970)

Japsen, P., Green, P.F., Bonow, J.M., Nielsen, T.F.D. and J.A. Chalmers (2014) From volcanic plains to glaciated peaks: Burial and exhumation history of southern East Greenland after opening of the NE Atlantic. *Global Planetary Change*, v. 116, 91-114.

Jensen, J.R., 2004. FASAN-1, Final Well Report, DONG E&P, 2004

Kempka, T. & Kühn, M. 2013: Numerical simulations of CO2 arrival times and reservoir pressure coincide with observations from the Ketzin pilot site, Germany. *Environmental Earth Sciences* 70, 3675– 3685. <https://doi.org/10.1007/s12665-013-2614-6>.

Khoshkholgh, S., Zunino, A., & Mosegaard, K., 2021: Informed Proposal Monte Carlo. *Geophysical Journal International*, Volume 226, Issue 2, 1239–1248, <https://doi.org/10.1093/gji/ggab173>

Khoshkholgh, S., Zunino, A., & Mosegaard, K., 2021: Full-waveform Inversion by Informed proposal monte carlo. *Earth and Space Science Open Archive*, doi:10.1002/essoar.10506565.1

Kleist, J., Newhouser J., Lee, L., Booth, S., Ducote C. and Williams, W, 1977. ADDA-1, Well Completion Report. Chevron Petroleum Denmark, 1977.

Konradi, P.B. (2005) Cenozoic stratigraphy in the Danish North Sea basin. *Netherlands Journal of Geosciences*, vol 84, no. 2, 109-111. Doi: <https://doi.org/10.1017/S001677460002299X>

Kopp, A., P.J. Binning, R. Helmig, and H. Class. A contribution to risk analysis for leakage through abandoned wells in geological CO2 storage. *Advances in Water Resources*, 33:867879, 2010.

Lithostratigraphic Chart of the Central North Sea, 2014, <https://www.npd.no>

Maersk Olie & Gas, 1992, South East ADDA-1, Final Well Report. Maersk Olie & Gas, 1992.

Matthews, W.R. and J. Kelly (1967) How to Predict Formation Pressure and Fracture Gradient. *Oil and Gas Journal*, v. 65, 92-1066.

Metropolis, N., Rosenbluth, M., Rosenbluth, A., Teller, A., Teller, E.: Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1963)

Michael, K., Golab, A., Shulakova, V., Ennis-King, J., Allinson, G., Sharma, S. & Aiken, T. 2010: Geological storage of CO<sub>2</sub> in saline aquifers—A review of the experience from existing storage operations. *International Journal of Greenhouse Gas Control* 4, 659– 667.

<https://doi.org/10.1016/j.ijggc.2009.12.011>.

Mosegaard, K., Tarantola, A.: Monte Carlo sampling of solutions to inverse problems. *J. Geophys. Res.* **100**(B7), 12431– 12447 (1995)

Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: *Inverse Problems*, 14, 405–426, [https://doi: 10.1088/0266-5611/14/3/004](https://doi.org/10.1088/0266-5611/14/3/004).

Mosegaard, K.: *Monte Carlo Analysis of Inverse Problems*. University of Copenhagen, Copenhagen (2006)

Nielsen, S.B., Thomsen, E., Hansen, D.L. and O.R. Clausen (2005) Plate-wide stress relaxation explains European Paleocene basin inversions. *Nature* v. 435, 195–198. <https://doi.org/10.1038/nature03599>

Nielsen, O. B., Rasmussen, E. S. and B. I. Thyberg (2015) Distribution of Clay Minerals In the Northern North Sea Basin During the Paleogene and Neogene: A Result of Source-Area Geology and Sorting Processes, *Journal of Sedimentary Research*, v. 85.6, 562–581. <https://doi.org/10.2110/jsr.2015.40>.

Ogata, K., Senger, K., Braathen, A. & Tveranger, J. 2014: Fracture corridors as seal-bypass systems in siliciclastic reservoir-cap rock successions: Field-based insights from the Jurassic Entrada Formation (SE Utah, USA). *Journal of Structural Geology* 66, 162– 187. <https://doi.org/10.1016/j.jsg.2014.05.005>.

Ogaya, X., Ledo, J., Queralt, P., Marcuello, Á. & Quintà, A. 2013: First geoelectrical image of the subsurface of the Hontomín site (Spain) for CO<sub>2</sub> geological storage: a magnetotelluric 2D characterization. *International Journal of Greenhouse Gas Control*, 13, 168–179. <https://doi.org/10.1016/j.ijggc.2012.12.023>.

Orozova-Bekkevold, I. and T.G. Petersen, 2021. Numerical Forward Modeling of the Overpressure Build-up in the Cenozoic Section of the Central Graben in the North Sea *Journal of Petroleum Exploration and Production*, 11(4), 1621-1642.  
DOI: <http://link.springer.com/article/10.1007/s13202-021-01135-z>  
<https://doi.org/10.1007/s13202-021-01135-z>

Partyka, G., Gridley, J. and Lopez, J., 1999. Interpretational applications of spectral decomposition in reservoir characterization. *The Leading Edge*, 18, 353-360.

Peric, D. and A.J.L. Crook (2004) Computational strategies for predictive geology with reference to salt tectonics. *Comput. Methods Appl. Mech. Eng.*, v. 193, 5195-5222. <https://doi.org/10.1016/j.cma.2004.01037>

Plischke, B. (1994) Finite element analysis of compaction and subsidence - Experience gained from several chalk fields. Paper presented at the Rock Mechanics in Petroleum Engineering, Delft, Netherlands, August 1994. doi: <https://doi.org/10.2118/28129-MS>

Rasmussen, E. S. (2004) The interplay between true eustatic sea-level changes, tectonics, and climatic changes: What is the dominating factor in sequence formation of the Upper Oligocene-Miocene succession in the eastern North Sea Basin, Denmark?, *Global and Planetary Change*, v. 41.1, 15–30. <https://doi.org/10.1016/j.gloplacha.2003.08.004>

Ringrose, P., Thorsen, R., Zweigel, P., Nazarian, B., Furre, A.-K., Paasch, B., Thompson, N. & Karstad, P. 2017: Ranking and Risking Alternative CO2 Storage Sites Offshore Norway. Fourth Sustainable Earth Sciences Conference 3–7 September, Malmö, Sweden.

Rong O., Baker, C., Cowie I., and Wagner, E.R. 1985. DEEP ADDA-1, Well Completion Report. Chevron Petroleum Denmark, 1985.

Rotevatn, A., Fossen, H., 2011. Simulating the effect of subseismic fault tails and process zones in a siliciclastic reservoir analogue: implications for aquifer support and trap definition. *Mar. Pet. Geol.* 28, 16481662.

Schiøler, P., Andsbjerg, J., Clausen, O.R., Dam, G., Dybkjaer, K., Hamberg, L., Heilmann-Clausen, C., Johannessen, E.P., Kristensen, L.E., Prince, I., and J. Rasmussen (2007) Lithostratigraphy of the Palaeogene – Lower Neogene succession of the Danish North Sea, *Geological Survey of Denmark and Greenland Bulletin*.

Schutjens, P.M.T.M., Fokker, P., Rai, U.B.B. et al. (2019) Compaction- and Shear-Induced Well Deformation in Tyra: Geomechanics for Impact on Production. *Rock Mech Rock Eng.*, v52, 5205–5224, doi: <https://doi.org/10.1007/s00603-019-01892-8>

Sheriff, R., Geldart, L., 1995. *Exploration Seismology* (2nd). Cambridge University Press.

Shipton, Z., Evans, J., Kirchner, D., Kolesar, P., Williams, A., Heath, J., 2004. Analysis of CO2 leakage through low-permeability faults from natural reservoirs in the Colorado Plateau, southern Utah. In: Baines, S., Worden, R. (Eds.), *Geological Storage of Carbon Dioxide*. Geological Society of London, Special Publications 233, pp. 4358.

Smith, J., S. Durucan, A. Korre, and J. Shi. Carbon dioxide storage risk assessment: Analysis of caprock fracture network connectivity. *International Journal Of Green- house Gas Contro*, 5:226240, 2011.

Sulak, R. M. & Danielsen, J. (1989), Reservoir aspects of Ekofisk subsidence *Journal of Petroleum Technology*, Society of Petroleum Engineers, v41, 709-716

Thornton, D.A., and A.J.L. Crook (2014) Predictive modeling of the evolution of fault structure: 3-D modeling and coupled geomechanical/ flow simulation. *Rock Mech. Eng.*, v. 47, 1533-1549.

Tarantola A and Valette B 1982 Inverse problems = Quest for information J. Geophys. 50 159–70

Tarantola, A. (2005). *Inverse Problem Theory*. Siam.

Vilamajó, E., Queralt, P., Ledo, J. & Marcuello, A. 2013: Feasibility of Monitoring the Hontomín (Burgos, Spain) CO<sub>2</sub> Storage Site Using a Deep EM Source. *Surveys in Geophysics* 34, 441–461.  
<https://doi.org/10.1007/s10712-013-9238-y>.

Yilmaz, O. *Seismic Data Analysis, volume 2*. Society of Exploration Geophysicists, 1 2001. ISBN 978-1-56080-094-1. doi: 10.1190/1.9781560801580.

Zunino, A., Mosegaard, K., Lange, K. Melnikova, Y., Hansen, T.M., 2015. Monte Carlo reservoir analysis combining seismic reflection data and informed priors. *Geophysics* 80 (1), R31-R41.

## ***E.4 Probabilistic Approach for Risk assessment of CO<sub>2</sub>***

### ***Storage***

*Ivanka Orozoa-Bekkevold*

*Sarouyeh Khoshkholgh, Klaus Mosegaard*

---

# *Probabilistic Approach for Risk Assessment of CO<sub>2</sub> Storage*

---

A project under DHRTC's Sprint programme 2020

Ivanka Orozova-Bekkevold, Technical University of Denmark

Sarouyeh Khoshkholgh and Klaus Mosegaard  
Niels Bohr Institute  
University of Copenhagen

February 17, 2021

## Table of Contents

---

<b>Introduction</b> .....	<b>2</b>
<b>Well data</b> .....	<b>3</b>
<b>Conceptual modeling - Tyra: Future evolution of the overburden</b> .....	<b>5</b>
Conceptual modeling - Tyra: Forward finite elements method.....	6
Conceptual Model 1 - Tyra: Past evolution of the deposition of the Cenozoic overburden .....	7
Results: Conceptual Model 1: stress state at the pre-production stage.....	10
Conceptual Model 2 - Tyra: Future subsurface evolution after abandonment.....	11
Results: Conceptual Model 2: Reservoir compaction – slip on fault .....	13
<b>Seismic Study of the Fault Detection Limit</b> .....	<b>15</b>
Seismic inversion.....	17
<b>Conclusions and Discussions</b> .....	<b>17</b>
<b>Conclusions and Discussions</b> .....	<b>19</b>
<b>References</b> .....	<b>20</b>

## Introduction

---

Sequestration of CO<sub>2</sub> in former oil and gas reservoirs can contribute to amelioration of the global increase in CO<sub>2</sub> emissions, and it is already in use in a limited number of sites worldwide (Bachu, 2008; Michael et al., 2010; Ringrose et al., 2017). Injection of CO<sub>2</sub> for enhanced oil recovery has already been utilised by the oil industry for decades, particularly in onshore North America (Gozalpour et al., 2005). CO<sub>2</sub> is currently stored offshore Norway, in Sleipner and Snøhvit, with  $\sim 1.5 \cdot 10^6$  tonnes annually (Eiken et al., 2011). Pilot-projects have been carried out in Germany (Kempka & Kühn, 2013; Bergmann et al., 2016), Spain (Vilamajó et al., 2013; Ogaya et al., 2013) and Texas (Daley et al., 2008; Doughty et al., 2008), and this has greatly increased our understanding of CO<sub>2</sub> migration, monitoring and injection strategies in geological reservoirs.

Injection of CO<sub>2</sub>, or any other fluid, into subsurface reservoirs increases the risk of caprock failure and migration of fluid along faults, fracture corridors, and other pre-existing weak zones (Ogata et al., 2014a). Understanding the detection thresholds of such structures calls for careful mechanical modeling of the reservoir stress- and strain field, careful inversion of available seismic data, combined with geological knowledge from well data and outcrop data. In this way it may be possible to quantify the probability of significant CO<sub>2</sub> migration through the caprock, laying the ground for a meaningful risk evaluation.

During the last couple of decades considerable progress has been made in geophysical and geostatistical data analysis methods to correctly estimate model uncertainties and thereby to evaluate fault detection thresholds (Zee, e.g., Zunino et al., 2015). The goal of this pilot project is to propose a way of exploiting these methods for risk assessment in connection with CO<sub>2</sub> storage. Conceptual models

are developed to model the time evolution of the subsurface, giving information about current and future stress fields. This analysis provides the prior information to a subsequent probabilistic inversion of seismic data. Monte Carlo methods are used to simulate the noise in the data, and the noise is back-propagated through the geophysical (e.g., seismic) equations into the geophysical model, generating a model variability, and reflecting the uncertainty of the reservoir structure. Combining this approach with prior information about the mechanical properties of the reservoir, we evaluate the probability of fault migration scenarios. In this pilot project, we carry out a concrete, highly simplified numerical study of the sub-problem of estimating the density of sub-seismic faults in the overburden of an existing North Sea hydrocarbon reservoir, and established a simple probability model for releases through existing faults. The study is a starting point for developing a full-scale risk analysis system based on the principles outlined above.

## Well data

---

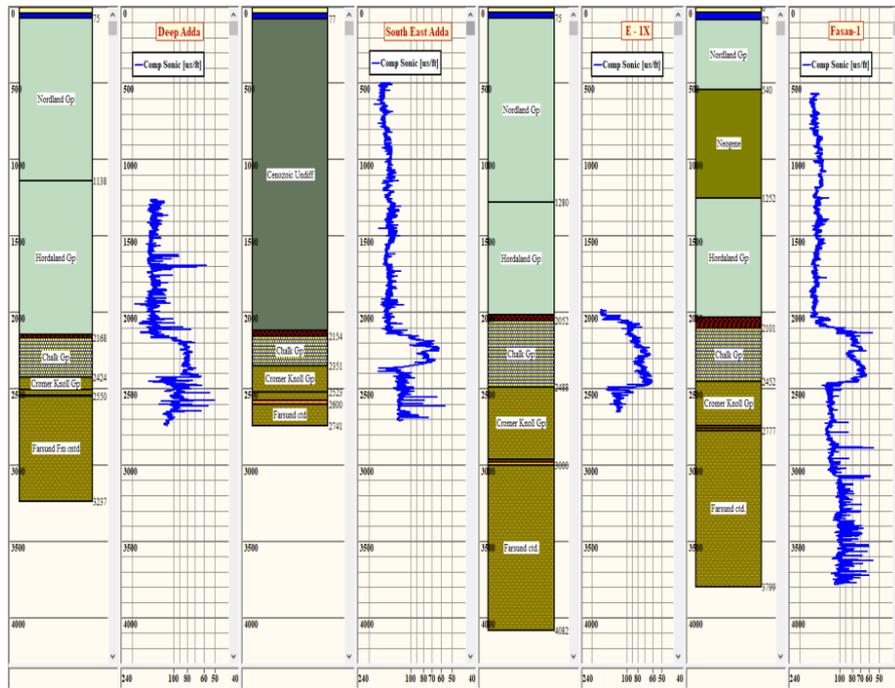
Data (petrophysical logs and well reports) for the following wells: Fasan-1, Deep Adda, South East Adda and E1-X (Figure 1), were provided by DHRTC. The wells are located in the Tyra field, Danish sector of the Central Graben, approximately 200 km west of Esbjerg.



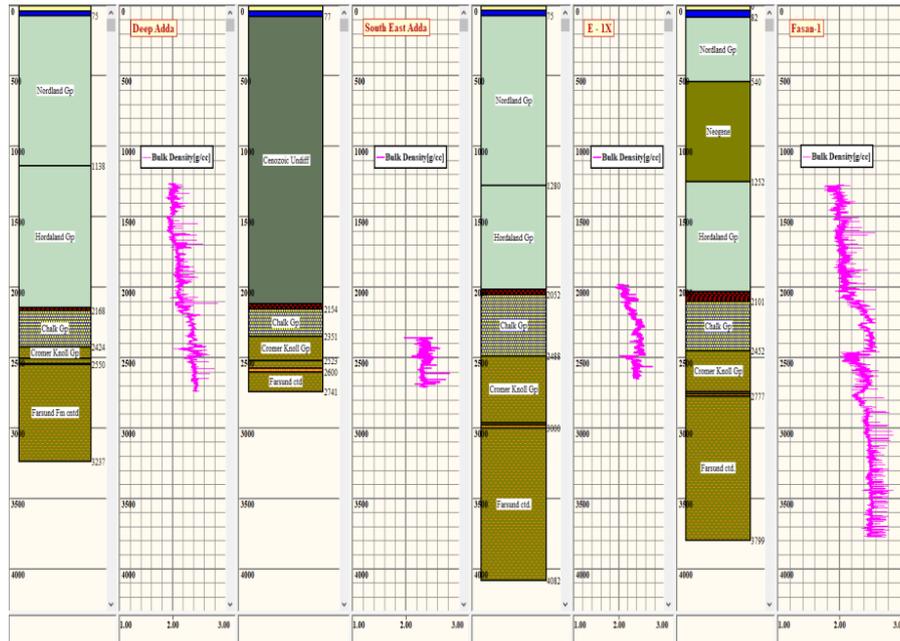
**Figure 1.** Location of the wells used in the analysis.

Figure 2 and 3 show the lithology columns and, respectively, the depth coverage of the sonic and density logs acquired in the four above mentioned wells. The figures illustrate very well the challenges related to the availability of log data in the overburden. In all four wells, no sonic (or density) data were acquired in the shallow section from sea bed to approximately 500 m depth. In the well E-1X, both sonic and density logs were acquired only in the reservoir (chalk) section, while in South East Adda, density data are available only below the Chalk group, in the deepest section of the well. This probably is due to the fact that the purpose of the South East Adda well was to investigate the hydrocarbon potential of the Crome Knoll Group.

Since the overburden provides the reservoir seal, it is important to have sufficient data of good quality in order to analyze its integrity and/or strength. One very important parameter is the fracture pressure of the seal – if the pressures at the top of the reservoir exceeds the fracture pressure of the seal a breach will occur and the reservoir fluid (CO<sub>2</sub> or hydrocarbons) will escape.



**Figure 2.** Overview of the sonic compressional logs (us/ft), acquired in the four wells. From left to right: Deep Adda, South East Adda, E-1X and Fasan-1. The depth reference is Rotary Table (RT, i.e. the rig floor).



**Figure 3.** Overview of the density logs (g/cc), acquired in the four wells. From left to right: Deep Adda, South East Adda, E1X and Fasan-1. The depth reference is Rotary Table (RT, i.e. the rig floor).

### Conceptual modeling - Tyra: Future evolution of the overburden

As mentioned above, the overburden provides the reservoir seal, and its integrity is a crucial parameter to consider when evaluating the risk of reservoir fluid (CO<sub>2</sub> or hydrocarbons) escape. The present-day stress and strain state of the Post-Chalk overburden in the North Sea is the result of the sedimentary deposition and geological evolution which took place in the last 56 MY. During production, the stress and strain changes due to reservoir depletion affect the overburden as well, among the examples are the subsidence experienced in the Tyra and the Ekofisk fields.

The overburden of the Chalk reservoirs in the Danish sector of the Central Graben consist of a thick (up to 2000-2200 m) pack of Cenozoic sediments: the Nordland, the Hordaland and the Rogaland Groups (Figure 2 and 3). The well reports of the four wells (Jensen, 2004; Rong et al., 1985; Kleist et al., 1977), used in this study, hold useful information about the lithological composition of the overburden. The upper (most shallow 0-500 m depth) part of the Nordland Group consist of predominantly Quaternary sand sand/clay mixtures, while the lower part consist predominantly of claystone with occasional thin limestone layers (Jensen, 2004). The Hordaland Group consists predominantly of clay-rich (shale) formations, in some interval interbedded with thin limestone layers (Jensen, 2004). The Rogaland Group is situated at the top of the Chalk Group and thus represents the seal for the uppermost Chalk reservoirs (Danian age, Lower Paleocene). The thickness of Rogaland varies across the field, as illustrated by the well lithology columns in Figure 2 and 3. The upper part of the Rogaland Group (the Balder formation) is characterized by tuffaceous claystone, while the lower part (Sele, Lista, Vaale) is predominantly claystone with stringers of marl (Jensen, 2004).

The evolution in geological time is modelled in the terms of a finite element method, using the software Elfen (ELFEN, Rockfield Software Ltd.). The framework and the theory behind the software are given in details in Crook et al. (2003), Peric and Crook (2004), (Crook et al. (2006a, 2006b), Thornton and Crook (2014).

The medium composing the overburden is represented as a fully saturated poro-elasto-plastic material. The mechanical field (the solid part) is solved explicitly, while the seepage field is solved by implicit time integration schemes and the two fields are coupled at given time intervals (Thornton and Crook, 2014). The mechanical properties of the water-saturated medium, is expressed as:

$$\text{div}(\boldsymbol{\sigma}') + [(1 - \varphi)\rho_s + \varphi\rho_f](\mathbf{g} - \mathbf{a}_s) = 0. \quad (1)$$

The fluid (water only) flow over geological time is represented with a transient equilibrium equation:

$$\text{div}\left[\frac{k(\varphi)}{\mu} \nabla P_f - \rho_f(\mathbf{g} - \mathbf{a}_s)\right] = [\varphi/K_f + (\alpha - \varphi)/K_s] \frac{\partial P_f}{\partial t} + \alpha \frac{\partial \varepsilon_v}{\partial t} \quad (2)$$

where:

$\boldsymbol{\sigma}' = \boldsymbol{\sigma} - \alpha P$  is the effective stress;  
 $P_f$  is the fluid pressure;  
 $\rho_s$  and  $\rho_f$  are the solid and the fluid density, respectively;  
 $\varphi$  is the porosity;  
 $k(\varphi)$  is the porosity-dependent permeability;  
 $\mathbf{g}$  is the Earth's gravitational acceleration;  
 $\mathbf{a}_s$  is the acceleration of the solid phase;  
 $K_f$  is the fluid bulk modulus;  
 $K_r$  is the frame bulk modulus;  
 $\alpha$  is the Biot's coefficient and  
 $\varepsilon_v$  is the volumetric strain.

The bulk modulus is expressed as a function of the mean effective stress  $P'$  (Thornton and Crook, 2014):

$$K = K_0 + \frac{(1-A)P_{co}}{\kappa} \exp\left[\frac{\varphi_0 - \varphi}{\lambda(1-\varphi_0)(1-\varphi)}\right] + \frac{A*\sigma'}{\kappa(1-\varphi)} \quad (3)$$

where

$\varphi_0$  is the initial porosity,  
 $K_0$  is the initial bulk modulus,  
 $P_{co}$  is the initial pre-consolidation pressure,  
 $A$  is a weighting factor,  $\kappa$  and  $\lambda$  are material constants.

The mean effective stress is derived by the vertical and horizontal stresses and the pore pressure. The vertical stress resulting by the weight of the overlying sediments is calculated as:

$$S_v = g * \sum \rho_i D_i \quad (4)$$

where  $\rho_i$  and  $D_i$  are the density and the thickness of the  $i$ -th sediment layer. In the present study, we assume that the horizontal stresses are isotropic, i.e. the minimum and the maximum horizontal stresses are equal, and that the vertical stress and horizontal stresses are related through:

$$S_h = K_{eff} * S_v \quad (5)$$

where  $K_{eff}$  is the so-called effective stress ratio (Matthews and Kelly, 1967). Finally, the mean effective stress is obtained as

$$\sigma' = \left[ \frac{S_v + 2 * S_h}{3} \right] - \alpha P_p = \sigma - \alpha P_p \quad (6)$$

where  $\alpha$  is the Biot's coefficient. Density and porosity are derived by well data (sonic and density logs).

In the cases where only a compressional sonic log from well data is available, the bulk density can be estimated by the empirical relationship, proposed by Gardner et al. (1974):

$$\rho_{bulk} = 0.23 * V_p^{0.25}, \quad (7)$$

where  $V_p$  is the p-wave velocity. The porosity can be calculated from the bulk density and grain density as follows:

$$\varphi = \frac{\rho_{gr} - \rho_{bulk}}{\rho_{gr} - \rho_{water}} \quad (8)$$

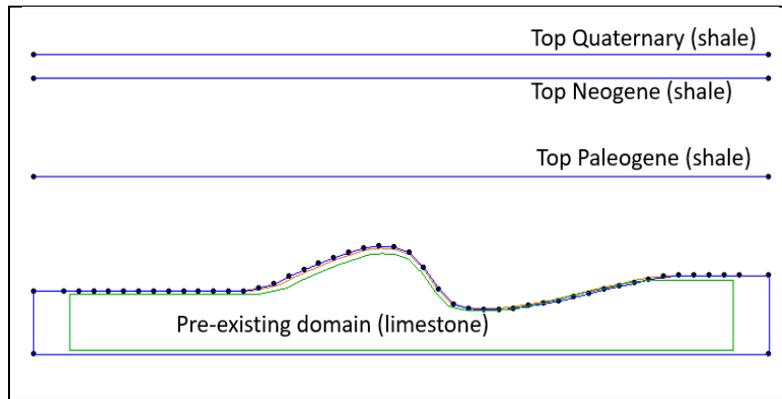
where  $\rho_{gr}$  (=2700 kg/m<sup>3</sup>) and  $\rho_{water}$  (=1035 kg/m<sup>3</sup>) are the grain density and the water density, respectively.

The evolution of these properties in time is driven by the rate of deposition and subsequent burial of fully saturated material, as described by Orozova-Bekkevold et al. (2021).

### Conceptual Model 1 - Tyra: Past evolution of the deposition of the Cenozoic overburden

The post-Chalk Cenozoic deposition in the Tyra field is represented by three main stages: Paleogene (Eocene-Oligocene), Neogene and Quaternary with the approximate duration of 33, 21 and 1.6 MY, respectively.

The geometry of the 2D model (Fig 4) is simplified as follows: a pre-existing domain with a top surface of generic anticline shape (might be imagined as a Top Chalk reservoir) upon which the three main units of the overburden were deposited. For simplicity, it is assumed that all three overburden units are composed by shale, while the pre-existing domain is composed by limestone. Since the studied area was covered by sea at all times, a water depth of appr. 100 m is assumed at each stage. At the end of the deposition, the domain is allowed to settle for 1.5 MY under gravity with no additional sedimentation.



**Figure 4.** Sketch of the model geometry

The sedimentation process is modelled as deposition of discrete layers, each with approximately 50 m thickness, over the time duration of the respective stage. The duration in time and the present thickness of the respective stages are summarized in Table 1. A simplified mean sedimentation rate is derived by dividing the present-day thickness by the stage duration.

Stage	Duration [MY]	Thickness [m]	Mean sedimentation rate [m/1MY]	Nr of layers
Paleogene (Eocene-Oligocene)	33.0	880 (above crest)	27	18
Neogene	21.0	1250	60	25
Quaternary	1.6	300	187	6
Post-depo settling	0.5	N/A	N/A	N/A

**Table 1.** Main deposition stages with duration (in Million Years, MY) and present day thickness.

It is assumed that the material, composing the different units, is fully saturated with water, isotropic and homogeneous and at any stage the sediments are at their maximum burial depth. Clay diagenesis and other chemical and temperature effects are not modelled at the current stage.

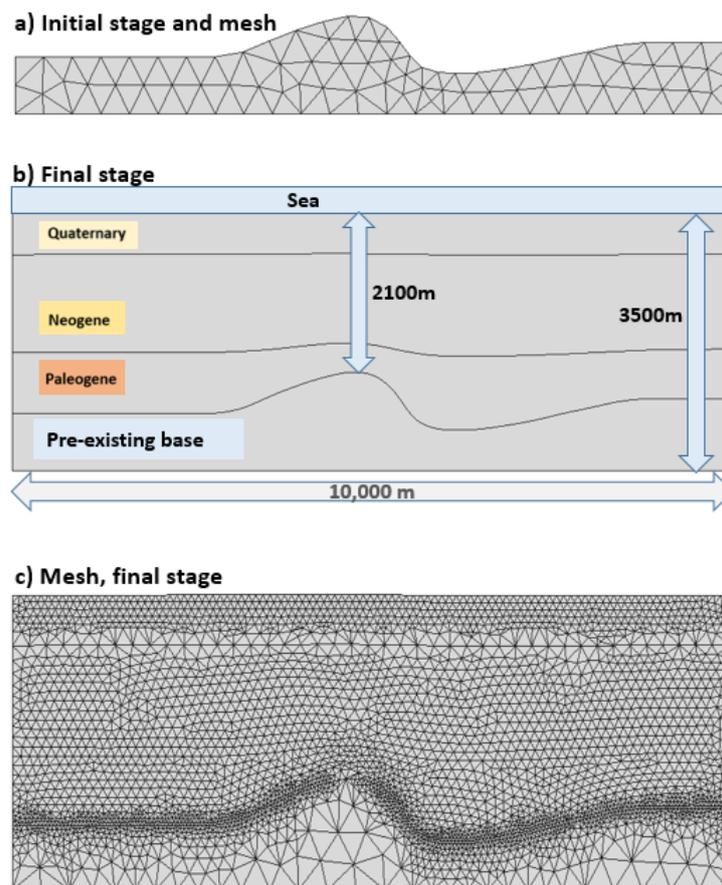
**Finite elements model:** The forward numerical modelling of the build-up of the overburden in time is performed with the finite element software *Elfen*. The equations used in the finite elements modelling are given in the text above.

**Initial and boundary conditions:** The main force acting upon the domain is the gravity. The gravitational force originates at the top of the sediments and acts downwards. This setup is considered representative for the Cenozoic Period in the North Sea basin, where no major tectonic events (uplift,

erosion, collision, subduction etc.) occurred, and thus the maximum stress is caused only by the weight of the deposited material. The pre-existing domain is not allowed to deform at the bottom and across the sides, while the material can deform inside and along the top of the domain. Uniaxial compaction i.e. plain strain conditions is assumed. As mentioned above, the deposited material is modelled as a fully water saturated porous medium. The formation water can flow both vertically and horizontally within the domain, but there is no fluid flow from outside sources. The water does not flow through the bottom and no capillary and temperature effect are taken into account at this stage.

**Meshing:** The finite element mesh is generated by an advancing front algorithm, adding new elements as the geometry expands, following the deposition of new layers of material. The elements are triangular, with initial size of 400. The size of the elements is rescaled depending on the estimated plastic strain at a given step: plastic strain exceeding 2 results in diminished element size. At the beginning of the simulation (time 0), before start of the deposition, the finite elements mesh consist of 138 triangular elements, all with size 400. At the end of the simulation, the mesh consists of 6760 elements with size ranging from 50 to 400.

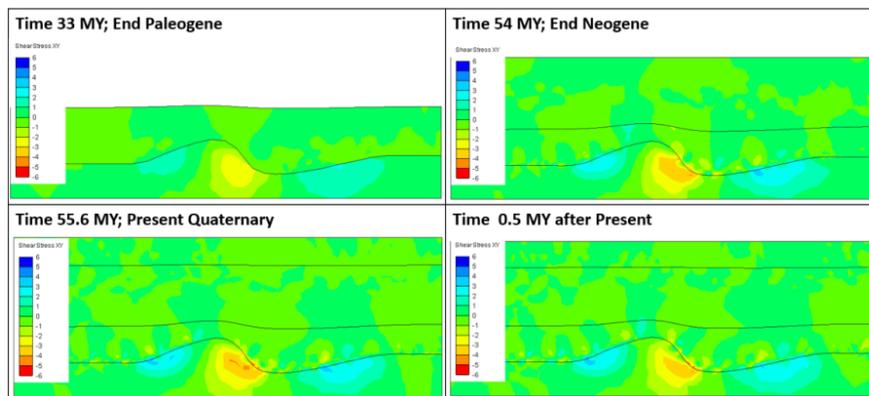
The finite element mesh at the beginning and the end of the simulation is summarized in Figure 5.



**Figure 5.** Model Geometry in time, and the finite elements mesh. From top to bottom: a) Pre-existing domain (reservoir) at time 0 (56 Ma) before the onset of the Cenozoic deposition; b) The final stage at present time after appr. 56 MY of deposition; c) final elements mesh at the final stage.

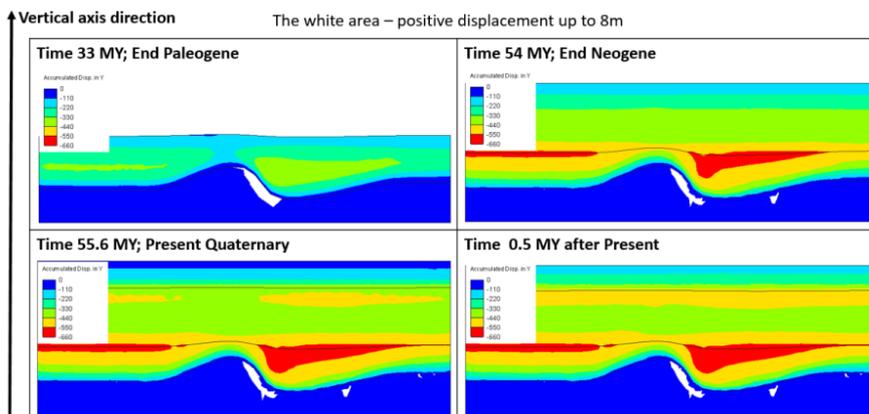
## Results: Conceptual Model 1: stress state at the pre-production stage

If the shear (differential) stress exceeds the strength of the material, it might result in fracturing of the formation. Figure 6 shows the evolution of the shear stress during the deposition of the overburden. As the figure shows, the largest shear stress is found in the reservoir, along the flanks of the anticline structure, close to its top. These zones might be considered being at higher risk of fracturing. The largest shear stress magnitude are estimated to appear at present time where the overburden build-up is completed. If no further deposition is going to occur in the next 0.5 MY, the medium relaxes and the magnitude of the shear stress could decrease.



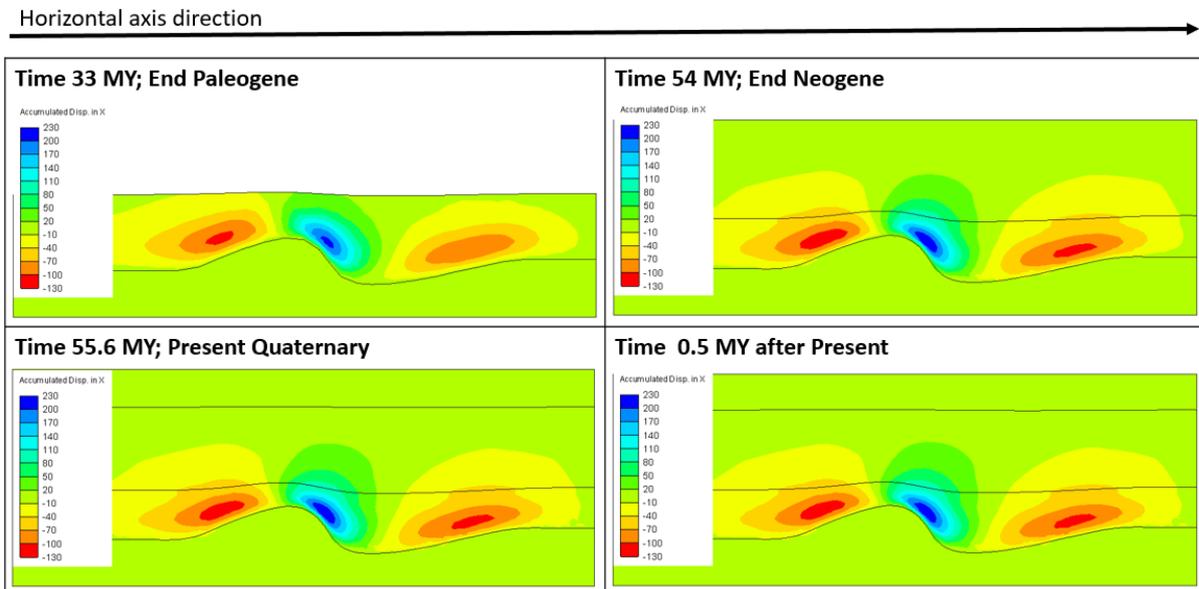
**Figure 6.** Evolution of shear stress during the deposition of the overburden

Figure 7 shows the evolution in the consolidation and compaction (accumulated displacement in the vertical plane) of freshly deposited clay-rich sediments during the burial process. This could also be interpreted as a kind of deformation accumulated during the Cenozoic sedimentation. The maximum concentration is in the overburden, along the Paleogene-Neogene boundary, roughly above the flanks of the anticline structure. Negative values denote displacement downwards. The white areas in the pre-existing domain (the reservoir) on the right side of the anticline, denote an very small (max. 8 m) upwards displacement.



**Figure 7.** Accumulated vertical displacement [m]: Negative values denote downwards movement (opposite to the axis direction). The white areas denote very small (max. 8 m) upwards displacement.

The evolution of horizontal displacement during the deposition of the overburden is displayed in Figure 8.



**Figure 8.** Accumulated horizontal displacement [m]: Negative values, given in red, denote displacement against the axis direction; positive values (blue) denote displacement along the axis direction.

Negative values, given in red, denote displacement against the axis direction; positive values (blue) denote displacement along the axis direction. As it can be seen on the Figure, horizontal displacement is concentrated in the immediate overburden (the cap rock) on the sides of the anticline. It developed already during the 1<sup>st</sup> sedimentation stage, Paleogene, and the pattern was preserved in all following stages. The displacement on the sides of the anticline could be interpreted as material sliding away from the top of the structure along the slope.

Comparing Figure 7 and Figure 8, one can see that the blue area on Figure 8 corresponds roughly to the white area in Figure 7. One possible interpretation could be that the horizontal displacement along the steepest side of the anticline is “dragging” material towards the depression to the right, thus “alleviating” the load on the anticline shoulder.

### Conceptual Model 2 - Tyra: Future subsurface evolution after abandonment

The geological evolution of the entire subsurface (both overburden and reservoir) continues also after abandonment. The development and exploitation of the hydrocarbon reservoir(s) induced man-made changes in the local stress regime, which resulted in disturbing the natural stress balance in the area. After abandonment, the entire subsurface medium, both reservoir and overburden, will continue to react to these changes until a new state of stress equilibrium is reached.

Modelling of the subsurface reaction in time after abandonment of a hydrocarbon field is especially interesting in the view of the possibility to use such reservoirs as a possible CO<sub>2</sub> storage location. One of

the crucial issues in CO<sub>2</sub> storage is the integrity of the cap rock. If the seal is compromised, the CO<sub>2</sub> might escape from the reservoir and migrate to surface/sea bed.

Here, we apply again forward finite elements modelling to investigate the risk for fault re-activation as consequence of reservoir compaction in an abandoned hydrocarbon field. The model is based on data from the Tyra field. The field is not yet abandoned, but it experienced significant reservoir compaction, which lead to sea bed subsidence.

It must be emphasized that the geometry of the 2D model presented here is only a very simple sketch of the actual subsurface system in the Tyra field.

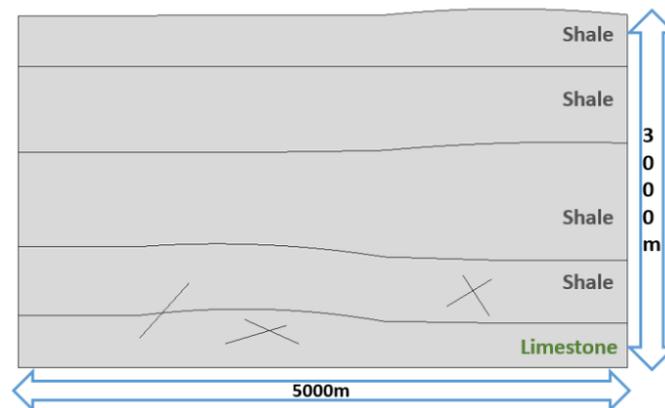
The geometry of the 2D model is represented in Figure 9. The length of the domain is 5000 m, the thickness – approximately 3000m. The bottom section represents a limestone reservoir with assumed thickness from 370 to 485m, overlaid by 4 overburden shale units. The initial thicknesses incorporate expected consolidation and compaction of the material under gravity.

In order to investigate the risk of seal failure, represented by sub-seismic fractures (i.e. fractures that cannot be seen in the seismic data), a couple of arbitrary pre-existing faults are incorporated: one crossing from the reservoir into the seal, two only in the reservoir and two only in the cap rock. Note that the fractures might not be representative for the actual fault system in the Tyra field.

The evolution in time is modelled as two stages:

Stage 1: settling under gravity with the duration 0.5 MY;

Stage 2: reservoir compaction over 0.1 MY modeled as vertical displacement of Top Reservoir reaching 150 m at the end of the stage. The values of stage duration and vertical displacement are arbitrary.



**Figure 9.** Model geometry: A limestone (reservoir) unit, overlaid by thick clay-rich overburden. The geometry and location of the fault lines are arbitrary.

**Finite elements model:** The forward numerical modelling is performed with the same finite element software Elfen. In this simulation, we investigate only the mechanical field (the solid part of the porous medium) and do not model fluid flow.

**Initial and boundary conditions:** As before, the main force acting upon the domain is the gravity force; it originates at the top of the sediments and acts downwards. Since gravity is a global force, it is active in all stages, also during compaction.

Deformation is not allowed at the bottom and across the sides, while the material can deform inside and along the top boundary. Uniaxial compaction i.e. plain strain conditions are assumed.

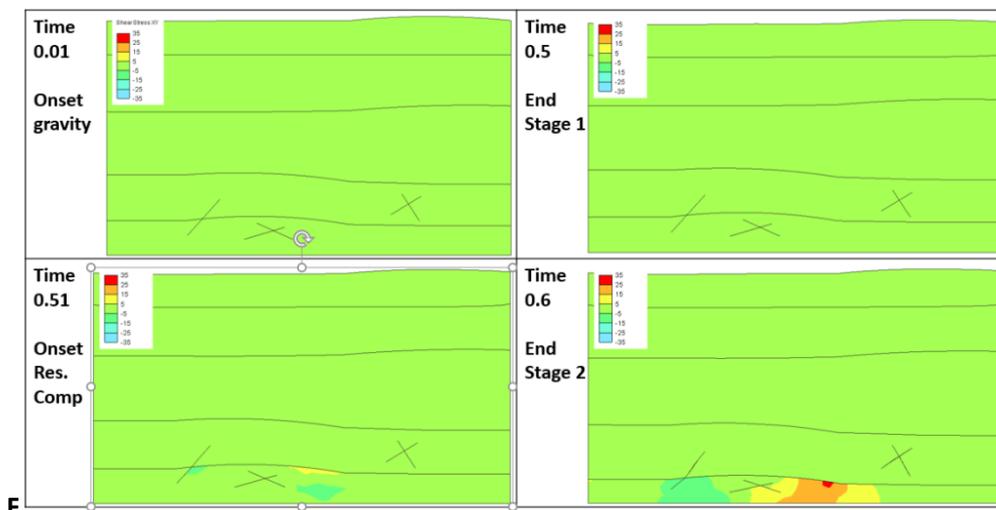
**Meshing:** The finite element mesh is composed by 3230 triangular elements with size 100.

### Results: Conceptual Model 2: Reservoir compaction – slip on fault

This section reports the results of the conceptual simulation of reservoir compaction after the abandonment of a depleted hydrocarbon field.

The time evolution is modeled in two stages. Stage 1: Gravitational settlement of the system (roughly corresponding to the pre-production state) with duration of 0.5 model time units (for example, million years) and Stage 2: A 150 m vertical downwards displacement of the top of the reservoir (compaction) occurring over 0.1 model time units.

Figure 10 shows the evolution of the shear stress in time. Negative values denote compressive stress, positive – tensile. As it can be seen on the figure, shear stress starts to build up in the reservoir itself at the onset of compaction (time 0.51), reaching maximum values at the end of compaction (time 0.6). There is no significant shear stress in the cap rock and the rest of the overburden.

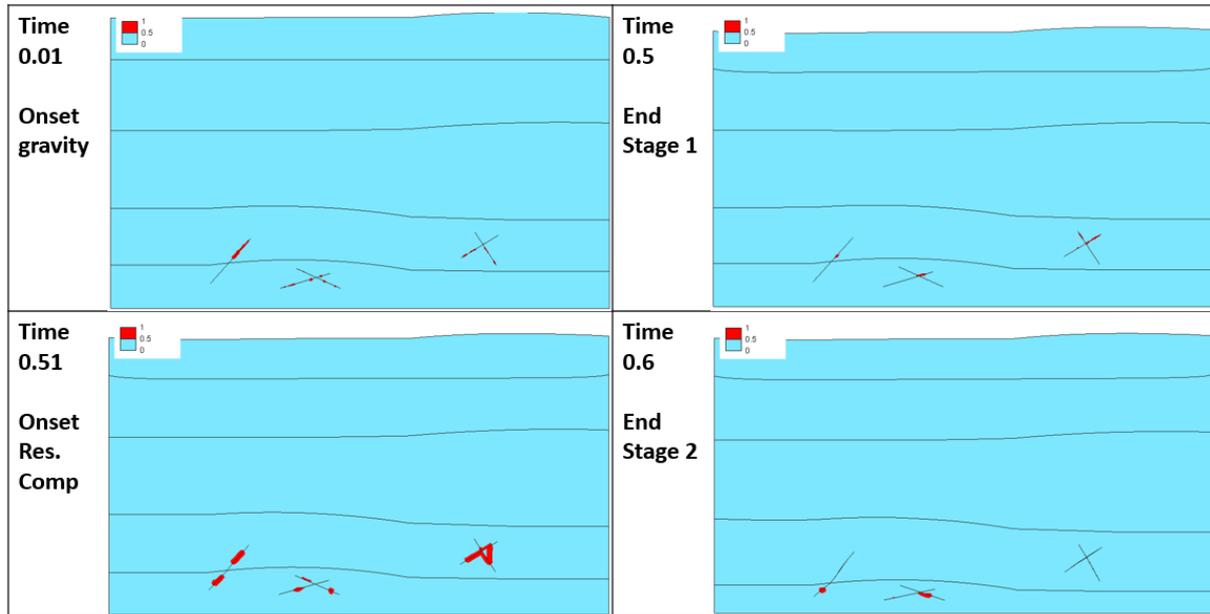


**Figure 10.** Shear stress [Mpa]: - compressive; + tensile. The lines denote pre-existing faults.

Figure 11 shows the possibility of fault re-activation, expressed in terms of slip-on-fault. A value of 1 indicates that slip may occur; a value of 0 denotes no slip.

As it can be seen on the figure, some slip on all faults is possible during the gravitational settling of the system (time 0.01 to 0.5). At the onset of compaction (time 0.51) the possibility for fault slip is very high,

especially in the faults in the overburden. At the end of compaction (time 0.6) there is still possibility for slip on the faults in the reservoir, while the faults in the overburden are quiet.



**Figure 11.** Slip on fault: 1 – yes; 0 – no. Time 0.5 (End of Stage 1) correspond to pre-production state. Time 0.51 corresponds on the 1<sup>st</sup> onset of compaction. Time 0.6 corresponds to the end of compaction.

## Seismic Study of the Fault Detection Limit

A NE-SW striking 2D seismic profile (selected from a 3D seismic volume) from the Tyra field, connecting the wells E-1X and Deep Adda-1, was selected for this pilot study. The geographical location of the selected profile is shown in Figure 12. The subset of the seismic profile used in the numerical study is shown in Figure 13.

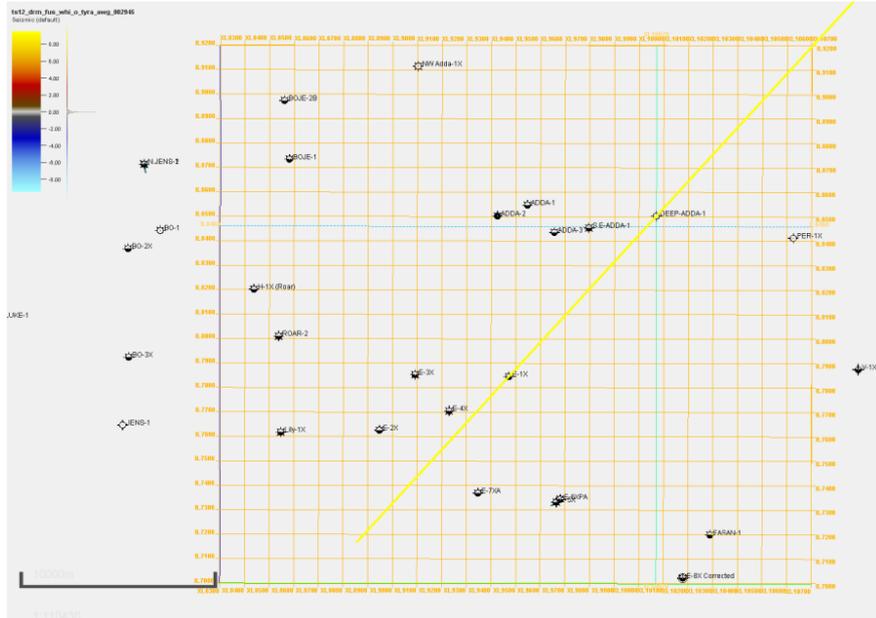


Figure 12. Location of the 2D data set used in this study (yellow line)

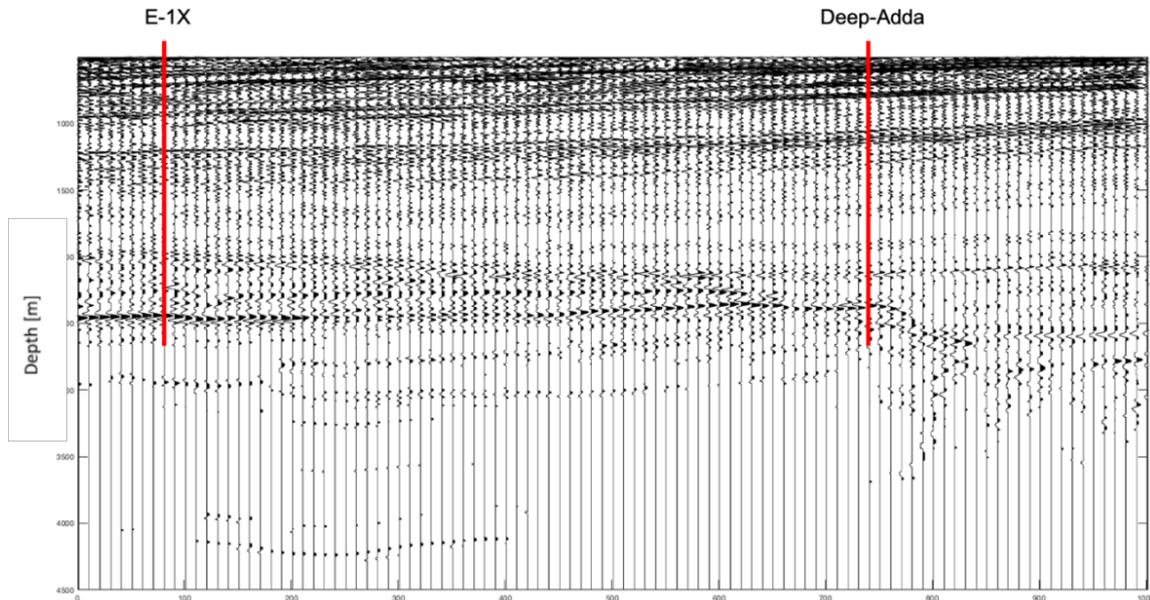


Figure 13. The 2D seismic data used in our analysis (depth section)

Assuming that the reflectivity is approximately a white noise series, a zero phase statistical wavelet was estimated from data from the reservoir overburden around the Deep-Adda wellsite (Figure 14).

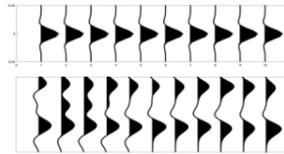


Figure 14. Top: Estimated wavelet at the Deep-Adda wellsite.  
Bottom (for comparison): Data from the Rogaland Group, close to the wellsite

The effective noise variance of the data was estimated as the variance of the data residual (the difference between the observed data and the synthetic data computed from the wavelet) at the well site. Synthetic data were computed from the reservoir model by first generating zero offset seismic data using the exploding reflector model (Loewenthal et al., 1976), followed by a depth migration using a rough background velocity model derived from well data and a simple interpretation of data (Figure 15). The rock density  $\rho$  is everywhere calculated from the approximate formula  $\rho [g/cm^3] = 260 v [ft/\mu s]$ .

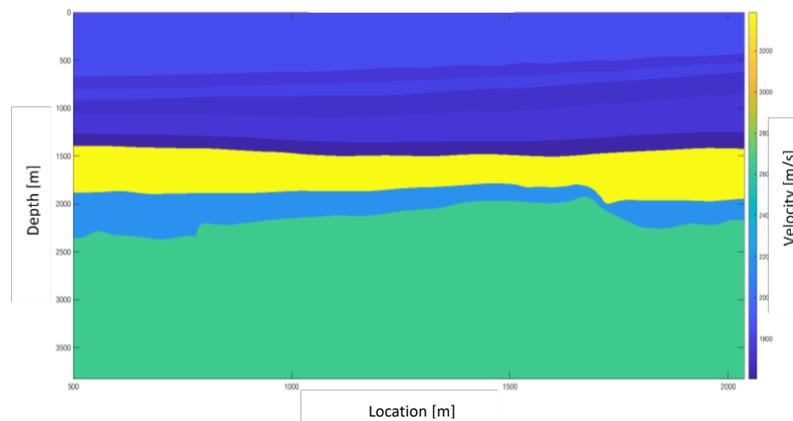


Figure 15. Seismic Background Velocity Model interpolated between wells, guided by interpretation.

Our synthetic data simulate in this way the processing (stacking, deconvolution and migration) of the data used in this study (Figure 13), including the horizontal smoothing due to imperfections in the processing, especially the data migration.

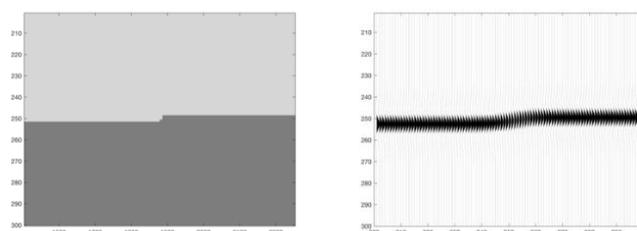


Figure 16. Left: A minor fault with a vertical displacement of 3m.  
Right: A simulated seismic profile across the fault.

The objective of the seismic study was to carry out a simplified inversion study aimed at estimating the detection limit of small-scale faults. A simple illustration of the problem is given in Figure 16 showing a

minor fault with a near-vertical displacement of 3 m, and its seismic response on a stacked and migrated seismic section. In this example, the fault is so small that it is indistinguishable from a small flexure in the layering. Only faults above a certain size (depending on the noise level and the impedance of the surrounding rocks) can be detected by the seismic data. Our aim was to detect possible locations of such faults, and to calculate the likelihood of sub-resolution faults and fractures everywhere in the reservoir overburden.

Above the seismic resolution, the seismic data provide most of the information needed to estimate fault/fracture probability. The limiting factor here is the noise level in the data. Below the seismic resolution, fault/fracture probability is derived from our study of the mechanical evolution of the reservoir. Our final output is a cross-section of fault/fracture probability in the overburden. Only if the reservoir is separated from the sea bottom with a connected zone of near-zero fault/fracture probability, the overburden is deemed uncompromised.

## Seismic inversion

---

Our basic method is a Monte Carlo procedure to randomly perturb layer velocities and layer boundaries in such a way that the resulting synthetic data fit the observed data within the error bars, and the resulting reservoir model is consistent with the prior information. Our layer boundary perturbations include generation of faults.

The general solution to inverse problems in a probabilistic formulation was given by Tarantola and Valette 1982 as the posterior probability density:

$$\sigma(\mathbf{m}) = L(\mathbf{m})\rho(\mathbf{m})$$

where  $\rho(\mathbf{m})$  and  $L(\mathbf{m})$  is the prior probability density and the likelihood function, respectively. The prior carries information that is independent of the observed data (in our case information from the initial study of the mechanical evolution of the reservoir, including the current shear stresses), and the likelihood function measures the degree of fit between the calculated and the measured data (in our case the seismic data). The objective of our calculations is to characterize  $\sigma(\mathbf{m})$  by a sample of models  $\mathbf{m}$ . This sample will, if sufficiently large, contain all information, in a probabilistic form, about the solution to the problem.

We use the algorithm proposed by Mosegaard and Tarantola (1995) to sample  $\sigma(\mathbf{m})$ , given a numerical procedure for generating faults with the prior probability (see above), and a routine for performing the forward modeling (producing stacked and migrated profile from a given reservoir model).

Our algorithm performs *importance sampling* where the density of sampled models in the model parameter space is proportional to the posterior  $\sigma(\mathbf{m})$ . This allows us to calculate any kind of probabilistic information related to the solution.

## Results: Seismic Inversion

---

Fundamental to our study is the prior information about the subsurface stress field provided by our numerical study of the mechanical evolution of the reservoir. From this information, we can, in principle, derive the faulting/fracturing probability that goes into our calculations. The seismic data is able to

constrain information about major faults, and hence also (indirectly) minor faulting and fracturing. This is illustrated in Figure 17 from Choi et al. (2016) showing the modern picture of fault zone architecture. A real fault plane is, an extended zone (the 'core' shown in the figure), surrounded by a damage zone whose width is related to the fault displacement. The larger the fault displacement, the broader the damage zone (see Figure 17 Right). When we calculate 'fault density' in our study, it includes the entire damage zone of each fault.

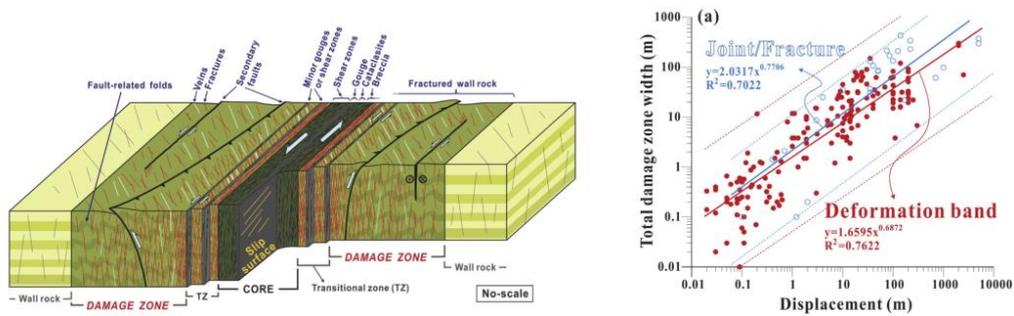


Figure 17. Left: Fault Zone Architecture. Right: Field data observations of the relation between the total damage zone width and fault displacement. (From Choi et al., 2016)

We assume a simple linear relationship between fault/fracture probability and shear stress in overburden (see Figure 7):

$$P_{frac} = 0.1 + 0.08 \cdot \sigma_{shear}$$

assigning larger probabilities to areas with high shear stress. The above relation is reasonable, but arbitrary, and further study is needed to establish a reliable relationship. The prior model generator in our Monte Carlo algorithm pseudo-randomly proposes faults in the overburden according to the above rule, with fault orientations between given angles, derived from regional studies of the area. The results shown are based on fault angles between  $24^\circ$  and  $36^\circ$  from vertical.

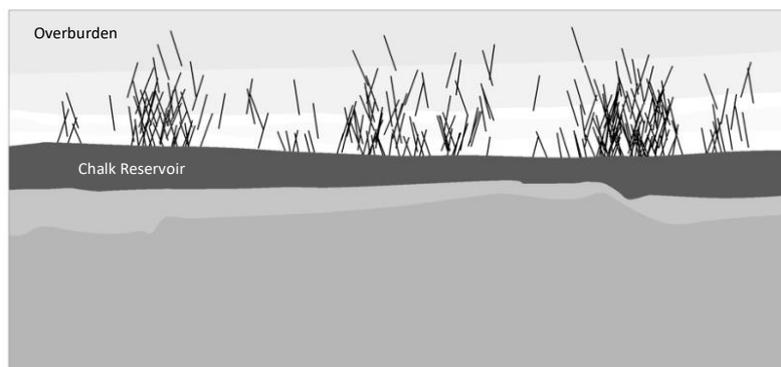


Figure 18. Sample output of the reservoir with fault locations after 3000 iterations. The correlation between fault/fracture density and the slopes of the deep layer boundaries is seen.

Figure 18 shows a sample output reservoir/fault location model produced after 3000 iterations of the algorithm. The black, oblique lines in the overburden are not fault planes, but lines centered at the

location of maximum fault displacement, indicating the fault direction. Using the last 2000 models (after an initial equilibration period of 1000 iterations) allows us to estimate the fault/fracture probability in the overburden (see Figure 19).

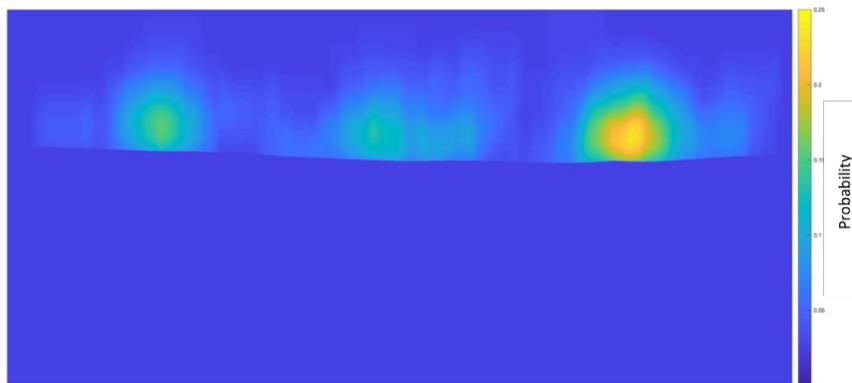


Figure 19. Posterior fault/fracture probability inferred from estimated shear stress fields and seismic data

The results indicate that the reservoir is separated from the sea bottom by a connected, low fault/fracture probability zone, meaning that the likelihood of having a tight reservoir is high.

## Conclusions and Discussions

---

Two forward conceptual models were developed to model the evolution in time of the subsurface. The first model investigated the evolution of the stress and strain fields as result of the deposition of the 2000-2500m thick clay-rich Post-Chalk overburden in the last 56 Million years. The second model simulated the effect of possible post-abandonment reservoir compaction. The main purpose of both models was to study possible risks, in terms of probability of fracturing or fault re-activation, related to seal integrity.

Since the resolution of even the best quality seismic data does not allow to detect small fractures and faults, the forward conceptual models allowed us to identify areas of possible fracturing and probable fault re-activation due to reservoir compaction, especially in the cap rock (immediate overburden).

Conceptual model 1 showed that the deposition of thick clay-rich Cenozoic Overburden in the last 56 MY on the top of an anticline limestone (reservoir) structure could have produced significant shear stress and vertical and horizontal displacement along the shoulders of the anticline. There areas could be at higher risk for fracturing.

Conceptual model 2 showed that the reservoir compaction of depleted fields can result in fault slip after abandonment, both on faults inside the reservoir and in the cap rock. The slip is most likely to occur at the onset of compaction and its likelihood decreases in time.

It should be kept in mind that the forward models of the evolution of the subsurface presented here are conceptual and were used to illustrate two important processes in the evolution of the subsurface. They can be improved by including further details.

A simplified probabilistic analysis of a seismic 2D profile across the Tyra reservoir, intersecting the E1-X and Deep Adda wellsites were carried out. The focus was on the detection limit of the data, and the result was a map of faulting and fracturing probabilities across the area. The result is a combination of seismic information with information about stress fields, derived from the subsurface evolution model.

Based on this study, we currently cannot make any conclusions about the actual conditions in the Tyra overburden. The subsurface evolution model was not specifically calibrated to the Tyra structure, and the seismic analysis was oversimplified and did not take advantage of all available data and all state-of-the-art methods available. A notable weakness in the seismic analysis is the lack of accurate data modeling (the wavelet estimation and the wave simulation), and we have also not included an obvious source of information, namely anisotropy of rock properties, which may be derived from full waveform data or AVO/AVA data. An investigation of anisotropy may contain valuable information about rock fracturing, and should be taken up in future studies.

The outcome of the study is, however, encouraging. We have proposed a new avenue in the analysis of caprock integrity of reservoirs for CO<sub>2</sub> storage. Our subsurface evolution model potentially allows us to predict current and future changes in a reservoir, our probabilistic approach to seismic data analysis allows our numerical model of the stress field to be incorporated in the final result, and our fault/fracture probability maps are ready to be included in a quantitative risk analysis.

## Acknowledgements

---

The authors would like to thank DHRTC for providing the data the financial support for this study.

The Technical University of Denmark and Ivanka Orozova-Bekkevold acknowledge the support of this work by Halliburton Software and Services, a Halliburton Company by allowing the authorized use of the "Drillworks/Predict" software.

## References

---

Bachu, S. 2008: CO<sub>2</sub> storage in geological media: Role, means, status and barriers to deployment. *Progress in Energy and Combustion Science* 34, 254–273. <https://doi.org/10.1016/j.pecs.2007.10.001>.

Bergmann, P., Diersch, M., Götz, J., Ivandic, M., Ivanova, A., Juhlin, C., Kummerow, J., Liebscher, A., Lüth, S. & Meeke, S. 2016: Review on geophysical monitoring of CO<sub>2</sub> injection at Ketzin, Germany. *Journal of Petroleum Science and Engineering* 139, 112–136. <https://doi.org/10.1016/j.petrol.2015.12.007>.

Crook, A.J.L., S.M. Willson, J.G. Yu and D.R.J. Owen (2006a) Predictive modelling of structure evolution in sandbox experiments. *Journal of Structural Geology*, v. 28, 729–744.

Crook, A.J.L., Owen, D.R.J., Willson, S., and J. Yu (2006b) Benchmarks for the evolution of shear localization with large relative sliding in frictional materials. *Comput. Methods Appl. Mech. Eng.*, v. 195, 4991-5010.

Crook, T., Willson, S., Yu, J.G. and R. Owen (2003) Computational modeling of the localized deformation associated with borehole breakout in quasi-brittle materials. *J. Petrol. Sci. Eng.*, v. 38, 177-186.

Daley, T., Myer, L., Peterson, J., Majer, E. & Hoversten, G. 2008: Time-lapse crosswell seismic and VSP monitoring of injected CO<sub>2</sub> in a brine aquifer. *Environmental Geology* 54, 1657–1665.  
<https://doi.org/10.1007/s00254-007-0943-z>.

Doughty, C., Freifeld, B. & Trautz, R. 2008: Site characterization for CO<sub>2</sub> geologic storage and vice versa: the Frio brine pilot, Texas, USA as a case study. *Environmental Geology* 54, 1635–1656.  
<https://doi.org/10.1007/s00254-007-0942-0>.

Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T.A. & Høier, L. 2011: Lessons learned from 14 years of CCS operations: Sleipner, In Salah and Snøhvit. *Energy Procedia* 4, 5541–5548.  
<https://doi.org/10.1016/j.egypro.2011.02.541>.

Elfen, Finite Element and Discrete Element software tool written at, and by Rockfield Software Ltd., Rockfield Software, Swansea, United Kingdom.

Gardner, G.H., Gardner L.W. and A.R. Gregory (1974) Formation velocity and density – the diagnostic basics for stratigraphic traps. *Geophysics*, v. 39.6, 770-780.

Gozalpour, F., Ren, S. & Tohidi, B. 2005: CO<sub>2</sub> EOR and storage in oil reservoir. *Oil & gas science and technology* 60, 537–546. <https://doi.org/10.2516/ogst:2005036>.

Jensen, J.R., 2004. FASAN-1, Final Well Report, DONG E&P, 2004

Kempka, T. & Kühn, M. 2013: Numerical simulations of CO<sub>2</sub> arrival times and reservoir pressure coincide with observations from the Ketzin pilot site, Germany. *Environmental Earth Sciences* 70, 3675– 3685.  
<https://doi.org/10.1007/s12665-013-2614-6>.

Kleist, J., Newhouser J., Lee, L., Booth, S., Ducote C. and Williams, W, 1977. ADDA-1, Well Completion Report. Chevron Petroleum Denmark, 1977.

Maersk Olie & Gas, 1992, South East ADDA-1, Final Well Report. Maersk Olie & Gas, 1992.

Matthews, W.R. and J. Kelly (1967) How to Predict Formation Pressure and Fracture Gradient. *Oil and Gas Journal*, v. 65, 92-1066.

Michael, K., Golab, A., Shulakova, V., Ennis-King, J., Allinson, G., Sharma, S. & Aiken, T. 2010: Geological storage of CO<sub>2</sub> in saline aquifers—A review of the experience from existing storage operations.

International Journal of Greenhouse Gas Control 4, 659– 667.

<https://doi.org/10.1016/j.ijggc.2009.12.011>.

Ogata, K., Senger, K., Braathen, A. & Tveranger, J. 2014: Fracture corridors as seal-bypass systems in siliciclastic reservoir-cap rock successions: Field-based insights from the Jurassic Entrada Formation (SE Utah, USA). *Journal of Structural Geology* 66, 162– 187. <https://doi.org/10.1016/j.jsg.2014.05.005>.

Ogaya, X., Ledo, J., Queralt, P., Marcuello, Á. & Quintà, A. 2013: First geoelectrical image of the subsurface of the Hontomín site (Spain) for CO<sub>2</sub> geological storage: a magnetotelluric 2D characterization. *International Journal of Greenhouse Gas Control* 13, 168–179.

<https://doi.org/10.1016/j.ijggc.2012.12.023>.

Orozova-Bekkevold, I. and T.G. Petersen, 2021. Numerical Forward Modeling of the Overpressure Build-up in the Cenozoic Section of the Central Graben in the North Sea Submitted to (last update Oct 6<sup>th</sup>, 2020): *Petroleum Exploration and Production Technology Manuscript* number: PEPT-S-20-00569, Under revision.

Peric, D. and A.J.L. Crook (2004) Computational strategies for predictive geology with reference to salt tectonics. *Comput. Methods Appl. Mech. Eng.*, v. 193, 5195-5222.

<https://doi.org/10.1016/j.cma.2004.01037>

Ringrose, P., Thorsen, R., Zweigel, P., Nazarian, B., Furre, A.-K., Paasch, B., Thompson, N. & Karstad, P. 2017: Ranking and Risking Alternative CO<sub>2</sub> Storage Sites Offshore Norway. Fourth Sustainable Earth Sciences Conference 3–7 September, Malmö, Sweden.

Rong O., Baker, C., Cowie I., and Wagner, E.R. 1985. DEEP ADDA-1, Well Completion Report. Chevron Petroleum Denmark, 1985.

Thornton, D.A., and A.J.L. Crook (2014) Predictive modeling of the evolution of fault structure: 3-D modeling and coupled geomechanical/ flow simulation. *Rock Mech. Eng.*, v. 47, 1533-1549.

Tarantola A and Valette B 1982 Inverse problems = Quest for information *J. Geophys.* 50 159–70

Vilamajó, E., Queralt, P., Ledo, J. & Marcuello, A. 2013: Feasibility of Monitoring the Hontomín (Burgos, Spain) CO<sub>2</sub> Storage Site Using a Deep EM Source. *Surveys in Geophysics* 34, 441–461.

<https://doi.org/10.1007/s10712-013-9238-y>.

Zunino, A., Mosegaard, K., Lange, K. Melnikova, Y., Hansen, T.M., 2015. Monte Carlo reservoir analysis combining seismic reflection data and informed priors. *Geophysics* 80 (1), R31-R41.

***E.5 Coupled multi-scale well and surface seismic data  
inversion using wavelet decomposition***

*Sarouyeh Khoshkholgh, Andrea Zunino, Thomas Mejer Hansen, Klaus Mosegaard*

# **Coupled multi-scale well and surface seismic data inversion using wavelet decomposition**

Sarouyeh Khoshkholgh, Andrea Zunino, Thomas Mejer Hansen, Klaus Mosegaard

The resolution of well data is usually much higher than that achievable from inversion of seismic data, and in conventional methods, the lower and higher frequencies from wells are treated independently from the intermediate frequencies obtained from seismic data.

By a multi-resolution analysis method, such as wavelet transform, signals can be analyzed at different scales. Through wavelet transform, a signal can be decomposed into coefficients representing different frequency bands: one low frequency “approximation” and other sub-bands controlling the details at different frequencies. Wavelet coefficients can then be back transformed to the original domain without any loss of information. Thus, through wavelet decomposition the earth can be represented at different scales by means of appropriate coefficients.

In this preliminary study, we perform wavelet decomposition of an image from a chalk cliff. The decomposition will provide us an approximation array which is a smooth version of the original image and some directional detail contained in other arrays of coefficients. Each decomposition level produces one approximation and three directional sub-bands and the approximation can be further decomposed to another approximation and some other sub-bands. The total number of wavelet coefficients is the same as the number of elements in the original image.

The wavelet coefficients in the approximation matrix can be constrained by geophysical data. For instance, data from well logs can be used to constrain the wavelet coefficients at different levels of resolution and be used for conditioning the details provided by the sub-bands. By considering the coefficients of a wavelet transform as model parameters, a multi-scale inversion of seismic data can be set up, where information at different resolutions and with different directionality is taken into account. In this framework, prior information from well logs and other sources, containing both short and long wavelengths, are naturally coupled, allowing us to handle data sets with different resolutions in an integrated approach.

